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

Актуальность

Переполнение буфера остаётся одной из наиболее распространённых ошибок в программах на языке Си, и в то же время считается одной из критичных ошибок, так как создаёт потенциальные уязвимости в программе. Для выявления ошибок переполнения буфера и других видов ошибок на стадии разработки программного обеспечения широко применяется статический анализ исходного кода. К современным промышленным статическим анализаторам предъявляются жёсткие требования, включая масштабируемость на программные системы до миллионов строк кода, не менее 50% истинных срабатываний, возможность обнаружения ошибок на отдельных путях выполнения программы. В то же время промышленные статические анализаторы, удовлетворяющие этим требованиям, реализованы в рамках закрытых инструментов и применяющиеся в них подходы известны лишь в общих чертах. Таким образом, актуальной является задача построения и реализации алгоритма поиска переполнения буфера, удовлетворяющего требованиям к промышленному анализатору.

Структура работы

Диссертационная работа состоит из введения, семи глав, заключения, списка литературы из 76 наименований и одного приложения. Объём диссертации составляет 145 страниц.

Во введении обосновывается актуальность темы исследования, формулируются цель работы и поставленные для её достижения задачи, определяется научная новизна, теоретическая и практическая значимость, приводятся основные положения, выносимые на защиту.

В первой главе приводится обзор существующих методов обнаружения переполнения буфера и определяются требования к детектору этого вида ошибок.

Во второй главе рассматривается разработанное определение ошибки доступа к буферу, которое может быть использовано для построения детектора переполнения буфера, чувствительного к путям выполнения. Наличие ошибки в функции по данному определению следует из свойств графа потока управления этой функции, что позволяет избежать ложных срабатываний из-за неизвестных предусловий анализируемой функции.

Третья глава посвящена описанию общего алгоритма внутрипроцедурного анализа для поиска переполнения массивов константного размера, основанного на символьном выполнении с объединением состояний. Изложение методов анализа приводится для программ на модельном языке. В первую очередь определяются понятия конкретного состояния программы и абстрактного состояния анализа, для последнего вводятся определения корректности и точности. Приводится описание отображения, с помощью которого строятся достаточные условия ошибки переполнения буфера. Для всех инструкций языка, кроме инструкции вызова, строятся передаточные функции, описывается процедура слияния абстрактных состояний, изменение семантики инструкций для последней итерации развёртки цикла. Доказываются теоремы о корректности и точности анализа при некоторых ограничениях на анализируемую функцию, а также теорема о корректности построенных достаточных условий для внутрипроцедурного случая.

В четвёртой главе дано описание межпроцедурного алгоритма поиска переполнения, основанного на методе резюме. Для поиска переполнения буфера при обращении по индексу, вычисленному из параметров функции, введённое в третьей главе отображение расширяется на формальные параметры функции. Также приводится описание алгоритма миграции абстрактного состояния в контекст вызова при применении резюме. Для обнаружения ошибки доступа к буферу, который передан как параметр в данную функцию, факт доступа к этому буферу, снабжённый условиями, помещается в резюме функции, при трансляции в контекст вызова проверяется на корректность и, при необходимости, помещается в резюме уже вызывающей функции. Данный подход применим и при анализе вызовов библиотечных функций, если в анализаторе имеются спецификации для них.

Доказаны теорема о корректности достаточных условий ошибки для межпроцедурного случая и теорема о сходимости анализа.

В пятой главе описаны расширения алгоритма анализа. К таковым относятся поиск переполнения при работе со строками языка Си, при использовании данных из недоверенного источника для вычисления индекса и при обращении к буферу в цикле.

Шестая глава посвящена алгоритмам поиска переполнения буфера произвольного размера. Рассмотрены два подхода: с использованием построенного отображения для анализа размера буфера и путём построения напрямую по определению ошибки достаточного условия, имеющего вид формулы с кванторами всеобщности.

В седьмой главе приводится описание реализации предложенных алгоритмов в промышленном статическом анализаторе Svace, разрабатываемом в ИСП РАН. Проведённые экспериментальные оценки показывают, что количество ложных срабатываний на проектах Android и Tizen по семи разработанным детекторам в среднем не превышает 35%, то есть находится в заданных в работе ограничениях. Покрытие на тестах из набора Juliet Test Suite составляет 47,6%, что существенно превышает тот же показатель открытого аналога – инструмента Infer, разрабатываемого и используемого в компании Facebook.

В заключении формулируются основные результаты диссертационной работы и возможные направления дальнейших исследований.

Научная новизна и практическая значимость

В работе получены следующие новые научные результаты:

1. Формальное определение переполнения буфера, которое может быть использовано при построении чувствительного к путям детектора.
2. Алгоритм межпроцедурного чувствительного к путям и контексту анализа, позволяющий построить достаточные условия наличия ошибки переполнения буфера константного размера, корректность которых доказана для программ, удовлетворяющих определённым ограничениям.
3. Алгоритм обнаружения переполнения буфера произвольного размера.

4. Расширение предложенных алгоритмов для поиска переполнений в циклах, при использовании данных, полученных их недоверенного источника, и переполнений, возникающих при обработке строк.

Разработанные методы реализованы в инструменте статического анализа Svace, внедрённого в компании Самсунг и в некоторых российских компаниях.

Достоверность и обоснованность научных положений и выводов работы

Результаты исследования опубликованы в 8 печатных работах (в том числе 6 в журналах из перечня ВАК, среди них 2 индексированы системой Web of Science), докладывались на 7 российских и международных конференциях.

Замечания

По содержанию диссертационной работы имеются следующие замечания:

1. В разделе 3.2.3, где описывается построение достаточных условий переполнения для результата арифметических операций, недостаточно подробно рассмотрены операции умножения и деления.

2. Для одного из предложенных подходов к обнаружению ошибок переполнения буфера произвольного размера, описанного в разделе 6.2, не приводится экспериментальная оценка результатов реализации.

3. В разделе 1.2.2 рассматривается выборка уязвимостей из реальных проектов, причиной которых явилась ошибка переполнения буфера, но в разделе о реализации не приводится оценка срабатываний разработанных алгоритмов на этой выборке.

4. В описании метода резюме (раздел 4.1) не рассмотрена обработка вызовов рекурсивных функций.

5. В главе 7 желательно было бы представить результаты сравнения разработанных автором средств не только с анализатором Infer, но и с другими аналогичными инструментами.

Приведенные замечания не влияют на общую положительную оценку

работы.

Заключение

Диссертация Дудиной Ирины Александровны «Поиск ошибок переполнения буфера в исходном коде программ с помощью символьного выполнения» является завершённой работой, в которой автору удалось решить поставленные перед диссертационным исследованием задачи. Основные результаты диссертации были полностью и своевременно опубликованы и докладывались на российских и международных конференциях. Автореферат диссертации верно отражает её основное содержание.

Диссертация отвечает требованиям Положения ВАК РФ о порядке присуждения учёных степеней, а её автор, Дудина Ирина Александровна, заслуживает присуждения ей учёной степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Доктор физико-математических наук, заведующий
сектором Федерального государственного
учреждения «Федеральный научный центр Научно-
исследовательский институт системных
исследований Российской академии наук»

В.А. Галатенко

«04» апреля 2019 г.