Отзыв

официального оппонента Галатенко Владимира Антоновича на диссертационную работу Каушана Вадима Владимировича «Поиск ошибок выхода за границы буфера в бинарном коде программ», представленную к защите на соискание учёной степени кандидата технических наук по специальности 05.13.11 — «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность

Диссертационная работа В.В. Каушана посвящена проблеме поиска ошибок выхода за границы буфера в исполняемом коде программ. Такие ошибки распространены в программах, написанных на языках программирования, которые не предоставляют возможность автоматической проверки границ массивов, таких как Си и Си++. Наличие в программе ошибок выхода за границы буфера может стать причиной уязвимостей, эксплуатация которых приводит к самым серьёзным последствиям: выполнению произвольного кода, что означает полную компрометацию системы. Уязвимости, порожденные ошибками работы с буферами памяти, относятся к категории наиболее опасных. Используемые на практике методы поиска таких ошибок и соответствующие программные инструменты, хоть и закрепились в процессах промышленной разработки ПО, но имеют множество ограничений и недостатков. Большая часть инструментов предполагает доступность исходного кода программ; символьное выполнение затрудняется экспоненциальным ростом числа рассматриваемых состояний, а фаззинг не способен обнаружить ошибку, если она не вызвала аварийное завершение. Таким образом, разработка новых методов, способных находить ошибки выхода за границы буфера в бинарном коде, является важной и **актуальной** задачей. В диссертационной работе предлагается новый подход к поиску ошибок, не требующий доступности исходных кодов, и применимый как к прикладному, так и к системному коду.

Структура работы

Диссертационная работа содержит 92 страницы и состоит из введения, четырёх глав, заключения, списка литературы из 55 наименований, списка рисунков, списка таблиц и одного приложения.

Во введении обосновывается актуальность работы, формулируются цель и основные задачи работы, обосновываются её научная новизна и практическая значимость, приводятся данные об апробации.

Первая глава посвящена обзору существующих подходов к задаче поиска ошибок выхода за границы буфера. Оцениваются преимущества и недостатки подходов, на основе которых составляется список требований, которым должен удовлетворять разрабатываемый подход к поиску ошибок. Среди основных недостатков существующих подходов отмечается относительно высокий уровень ложных срабатываний, что присуще методам, основанным на статическом анализе кода. Это приводит к значительным трудозатратам на этапе проверки ошибок для больших проектов. Поэтому в работе делается попытка минимизировать число ложных срабатываний за счёт автоматизированной проверки проявления найденных ошибок. К недостаткам большинства методов динамического анализа относят то, что они способны находить только те ошибки, которые проявились на этапе выполнения анализируемой программы. Одним из требований к разрабатываемому методу является возможность поиска ошибок, которые не проявлялись во время анализа.

Во второй главе описывается предлагаемый автором метод поиска ошибок выхода за границы буфера. Метод основан на символьной интерпретации трассы с дополнительным анализом длин обрабатываемых буферов. Интерпретация трассы происходит в несколько этапов: сначала машинные инструкции трассы транслируются в промежуточное представление, позволяющее проводить машинно-независимый анализ бинарного кода, а затем код в промежуточном представлении преобразуется в набор символьных уравнений для решателя в соответствии с предлагаемыми правилами интерпретации. При составлении уравнений используется логика QF_ABV, которая позволяет описывать уравнения над машинными словами и массивами слов без использования кванторов. В этой же главе описываются три вспомогательных метода. *Метод восстановления границ буферов* позволяет автоматически определять границы на основе статического представления программы, а также данных о функциях работы с динамической памятью. *Метод предварительного расширения покрытия кода* использует динамическое символьное выполнения для перебора путей в программе и получения наборов входных данных, при обработке которых достигается большее покрытие кода, чем на начальном наборе входных данных. В свою очередь, *метод восстановления функций работы со*

строками даёт возможность специальным образом анализировать не только библиотечные строковые функции, которые определены явно, но и те строковые функции, которые подверглись встраиванию в процессе компиляции.

В третьей главе описывается программная реализация предложенных методов. Для реализации была использована модульная среда анализа бинарного кода, разрабатываемая в ИСП РАН. Для получения трасс используется эмулятор QEMU с функциями детерминированного воспроизведения и трассировки. Метод предварительного расширения покрытия кода реализован на основе открытого программного инструмента S^2E.

В четвёртой главе описываются результаты применения предложенных методов на наборе программ для ОС Windows и Linux, а также на встроенном программном обеспечении для сетевого маршрутизатора. Детально разобраны примеры, на которых демонстрируется обнаружение различных ошибочных ситуаций.

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

1. Разработан метод поиска ошибок выхода за границы буфера с использованием символьной интерпретации трасс выполнения. Метод позволяет находить ошибки, которые не проявлялись в трассе и не требует наличия исходных кодов и отладочной информации.

2. Разработан метод предварительного расширения покрытия кода, а также метод выявления функций работы со строками, которые подверглись встраиванию в процессе компиляции. Методы позволяют повысить точность поиска ошибок выхода за границы буфера.

3. Предложенные методы реализованы в рамках программного инструмента.

В заключении также представлены возможные направления дальнейших исследований.

Научная новизна и практическая значимость

Научной новизной обладает предложенный метод символьной интерпретации трасс с использованием абстрагирования от длины переменных-массивов в программе. Метод позволяет проводить анализ в отсутствии исходных кодов или отладочной информации. Кроме того,

научной новизной обладает метод поиска ошибок выхода за границы буфера, основанного на символьной интерпретации трасс. Данный метод позволяет находить даже такие ошибки, которые не проявились в анализируемой трассе.

Предложенные методы поиска ошибок выхода за границы буфера реализованы в рамках среды динамического анализа бинарного кода.

Разработанный инструмент используется в образовательном процессе для обучения студентов ФУПМ МФТИ и ВМК МГУ. Полученные научные результаты могут быть применены для развития инструментов поиска ошибок, используемых в промышленной разработке, а также для сертификации программного обеспечения.

Достоверность и обоснованность научных положений и выводов работы

Достоверность исследований и результатов работы подтверждается апробацией её результатов на трех научных конференциях.

По теме диссертационной работы опубликовано 6 печатных работ, из которых 4 работы опубликованы в журналах, рекомендованных ВАК, 1 работа индексирована в Scopus и Web of Science.

Замечания

К содержанию и оформлению диссертационной работы имеются следующие замечания.

1. В работе отсутствует сравнение инструмента, реализованного автором, с другими анализаторами бинарного кода, а также со статическими анализаторами исходных текстов на языках более высокого уровня. Подобное сравнение позволило бы лучше оценить практическую значимость полученных результатов.
2. Многие статьи, на которые автор ссылается в обзорной части работы, относятся к 2003-2008 гг.; желательно было бы использовать более свежие публикации.
3. Имеются погрешности в оформлении списка литературы (см., например, элементы 14, 20, 43).

Перечисленные замечания не снижают качества выполненного научно-технического исследования и не влияют на общую положительную оценку работы.

Заключение

Диссертация В.В. Каушана представляет собой завершённое исследование, проведенное на высоком научно-техническом уровне. Основное содержание диссертации отражено в опубликованных диссертантом статьях, доложено на научных конференциях. Полученные результаты работы соответствуют поставленным задачам. Автореферат отражает основные положения диссертации.

Диссертационная работа В.В. Каушана соответствует требованиям ВАК РФ, предъявляемых к диссертациям на соискание учёной степени кандидата технических наук, а её автор, Каушан Вадим Владимирович, заслуживает присуждения ему учёной степени кандидата технических 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

«25» января 2018 г.