УТВЕРЖДАЮ
Директор Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН),
академик

И.А. Соколов
«28» 01 2018 г.

ОТЗЫВ
ведущей организации на диссертацию Каушана Вадима Владимировича «Поиск ошибок выхода за границы буфера в бинарном коде программ», представленную на соискание учёной степени кандидата технических наук по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Диссертационная работа Каушана В.В. посвящена исследованию и разработке методов автоматизации поиска ошибок выхода за границы буфера в бинарном коде программ по трассам выполнения. Разработанные методы не должны зависеть от целевой процессорной архитектуры, а также от операционной системы, используемой для запуска анализируемой программы.

Актуальность. Информационные системы и автоматизированные системы управления применяются в большинстве сфер человеческой деятельности, включая критическую инфраструктуру. Ошибки в программном обеспечении (ПО) несут ощутимый экономический ущерб, который в случае злонамеренной эксплуатации этих ошибок способен

значительно вырасти. Уязвимости ПО, в основе которых лежат ошибки выхода за границы буфера, являются одними из наиболее опасных, эксплуатация которых может привести к выполнению произвольного кода на уязвимой системе. Несмотря на развитие технологий программирования и появление новых «безопасных» языков, выявление ошибок выхода за границы буфера по-прежнему актуально. Такие языки как Си/Си++ продолжают использоваться в промышленной разработке, их популярность хоть и снизилась за последние десятилетия, но остается очень высокой, уступая только языку Java. Современный цикл разработки предполагает согласованное использование инструментов статического и динамического анализа, однако они не гарантируют избавления от всех ошибок. Когда доступен лишь исполняемый код программы, автоматизированный поиск ошибок ограничивается фаззингом. Такого рода ситуации возникают, если разработчики включают в свое ПО проприетарные библиотеки, а также при проведении тематических исследований ПО и аппаратно-программных средств по требованиям безопасности информации. Для дополнения существующих инструментов анализа и улучшения возможностей находить ошибки выхода за границы буфера при отсутствии исходных кодов требуется разработка новых методов поиска ошибок данного типа. В связи с этим, тема диссертации Каушана В.В, посвященной разработке такого метода, является актуальной.

Общая характеристика диссертационной работы. Диссертация имеет общий объём 92 страницы и состоит из введения, четырёх глав, заключения, списков литературы, рисунков, таблиц, сокращений и одного приложения. Список литературы включает 55 наименований.

В работе предложен новый метод обнаружения ошибок выхода за границы буфера в исполняемом (бинарном) коде программ. Метод автоматизируем, применим как к пользовательскому, так и к системному

коду, расширяем на различные процессорные архитектуры. К **основным результатам** работы можно отнести следующее.

1. Разработан метод поиска ошибок выхода за границы буфера на основе символьной интерпретации трассы с использованием абстрактной длины переменных-массивов, полученных по результатам обратной инженерии бинарного кода. Метод не требует наличия исходных кодов и отладочной информации и позволяет находить ошибки, даже если они не проявлялись в анализируемой трассе.

2. Разработаны методы, улучшающие точность поиска ошибок выхода за границы буфера за счёт выявления функций работы со строками, которые подверглись встраиванию в процессе компиляции и предварительного расширения покрытия кода при сборе трассы выполнения.

3. На основе предложенных методов разработано и реализовано программное средство для поиска ошибок выхода за границы буфера. Реализованные методы являются машинно-независимыми, а также абстрагированы от операционной системы, используемой для запуска анализируемой программы.

Во введении описывается актуальность задачи, формулируется цель и задачи диссертационной работы, раскрывается научная новизна и практическая значимость, а также приводятся основные положения, выносимые на защиту.

В первой главе приводится обзор и сравнение существующих подходов к задаче поиска ошибок выхода за границы буфера. Рассматриваются подходы на основе статического анализа, а также на основе инструментации и динамического символьного выполнения. Оцениваются преимущества и недостатки подходов, на основе которых составляется список требований, которым должен удовлетворять разрабатываемый подход к поиску ошибок. Среди основных

недостатков методов, основанных на статическом анализе, можно выделить относительно высокий уровень ложных срабатываний, который приводит к значительным трудозатратам на этапе проверки ошибок для больших проектов. В работе делается попытка минимизировать число ложных срабатываний за счёт автоматизированной проверки проявления найденных ошибок. К недостаткам многих 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.

Вторая глава посвящена описанию предложенного автором метода к поиску 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. Также в этой 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 восстановления функций работы со строками даёт возможность специальным образом анализировать строковые функции, которые подверглись встраиванию в процессе компиляции: такой анализ позволяет абстрагироваться от длины буферов не только в библиотечных функциях, но и в строковых функциях, представленных в

виде циклов. В свою очередь, метод предварительного расширения покрытия кода выполняет перебор путей в программе с помощью динамического символьного выполнения для получения наборов входных данных, при обработке которых достигается большее покрытие кода, чем на начальном наборе входных данных.

В третьей главе рассматривается программная реализация предложенных методов. Для реализации была использована среда анализа бинарного кода, разрабатываемая в ИСП РАН. Данная среда позволяет анализировать трассы выполнения, полученные с помощью полносистемного эмулятора QEMU. Метод предварительного расширения покрытия кода реализован на базе открытого программного инструмента S2E.

Четвёртая глава посвящена результатам применения описанных в работе методов. Для апробации были выбраны приложения для операционных систем Windows и Linux, а также встроенное программное обеспечение для сетевого маршрутизатора. Глава включает детальный разбор четырёх примеров, на которых демонстрируется обнаружение различных ошибочных ситуаций.

В заключении перечисляются основные результаты работы и предлагаются дальнейшие направления исследований.

Практическая ценность предложенного метода состоит в возможности его использования для анализа различных классов программ: пользовательских приложений ОС Windows и Linux, а также ядер ОС, работающих на настольных компьютерах, серверах и сетевых маршрутизаторах. Представленная диссертантом программная реализация поддерживает анализ кода широко распространенной архитектуры Intel64. Благодаря применению промежуточного представления для анализа бинарного кода имеется возможность

быстрого расширения перечня поддерживаемых процессорных архитектур, что недоступно большинству аналогичных инструментов.

Достоверность научных результатов обусловлена тем, что автором применен формализованный подход, задающий правила интерпретации трасс промежуточного представления. Математическую основу проведенного исследования составляет теория множеств, теория алгоритмов и математическая логика. Полученные экспериментальные данные соответствуют теоретическим результатам.

Основные результаты диссертационной работы опубликованы в открытой печати: 4 статьи в изданиях, включенных в список ВАК, из них одна работа индексирована Scopus и WoS, 2 публикации в сборниках трудов научных конференций. Результаты докладывались на трех конференциях. Автореферат в полной мере раскрывает содержание работы.

В диссертации могут быть отмечены следующие недостатки.

1. В работе заявлена возможность абстрагирования от процессорной архитектуры, однако результаты применения метода поиска ошибок приводятся только для программ, выполняющихся на процессорах с архитектурой Intel64.

2. Не проводились исследования границ применимости предложенного метода, как по объему требуемой памяти, так и по времени работы основных этапов.

3. Объем проведенных экспериментов не велик; не приводятся оценки длительности всего тракта анализа, который состоит не только из интерпретации трассы, но и построения тестовых данных и снятия трассы машинных команд.

Отмеченные недостатки не снижают общей положительной оценки диссертационной работы.

Данная работа соответствует требованиям утвержденного Постановлением Правительства Российской Федерации «О порядке присуждения ученых степеней» от 24.09.2013 г. №842, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук. В целом диссертационная работа соответствует паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» поскольку относится к области исследований «Модели, методы и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования».

Таким образом, диссертация Каушана Вадима Владимировича является законченной научно-квалифицированной работой и удовлетворяет всем требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, 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.

Отзыв обсужден и утвержден на заседании секции Ученого совета ИПИ РАН Федерального исследовательского центра «Информатика и управление» Российской академии наук, протокол № 1 от 25 января 2018 года

Главный научный сотрудник ФИЦ ИУ РАН,
Заслуженный деятель науки РФ,
доктор технических наук, профессор

_____ И.Н. Синицин

«25» 01 2018 г.