ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

Шабанова Бориса Михайловича

на диссертацию Саргсяна Севака Сениковича

«Методы оптимизации алгоритмов статического и динамического анализа программ»,

представленную на соискание ученой степени доктора технических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Актуальность темы

Актуальность диссертационной работы Саргсяна Севака Сениковича обусловлена необходимостью улучшения методов статического и динамического анализа программного обеспечения (ПО), что вызвано ростом сложности и объема современных программных систем. С увеличением объема и сложности ПО растет и количество уязвимостей, что делает выявление ошибок сложной задачей на всех этапах проектирования и реализации. Несмотря на развитие средств анализа и внедрение безопасных практик разработки, существующие инструменты анализа обладают значительными ограничениями. Важным аспектом является также недостаточная интеграция различных методов анализа: современные платформы не предоставляют удобных средств для их комбинирования, что ограничивает комплексное выявление уязвимостей. Помимо этого, использование открытого ПО сопряжено с рисками массового распространения уязвимостей, что подчеркивает критическую важность разработки более эффективных и универсальных инструментов анализа. Таким образом, обоснованной и актуальной задачей становится создание

комплексной платформы, способной интегрировать различные методы анализа и эффективно выявлять уязвимости, что особенно важно в условиях современных требований к безопасности и растущих киберугроз.

Теоретическая и практическая значимость полученных результатов, их научная новизна

Теоретическая значимость данной работы заключается в разработке новой концепции платформы анализа ПО, способной интегрировать методы статического и динамического анализа для выявления сложных уязвимостей и ошибок. В работе предложены методы и алгоритмы, которые на этапе экспериментального тестирования продемонстрировали высокую эффективность и показали преимущество по сравнению с существующими подходами. Практическая значимость результатов подтверждается созданием платформы "GENES ISP", реализующей предложенные методы анализа и уже внедренной в циклы разработки ПО в нескольких компаниях и учреждениях. Созданная платформа поддерживает процессы безопасной разработки в соответствии с требованиями ГОСТ Р 56939-2016 и "Методики выявления уязвимостей и недекларированных возможностей в программном обеспечении" ФСТЭК РФ. Отдельно разработанные методы в рамках диссертационной работы также нашла применение в инструментах Svace и ISP-Fuzzer, которые сегодня являются индустриальными стандартами в разработке безопасного ПО.

Научная новизна исследования заключается в создании архитектуры анализа, объединяющей сбор артефактов для масштабного открытого ПО и методов анализа кода, включая новый масштабируемый метод поиска клонов и метод сопоставления исходного и бинарного кода, а также уникальный подход к обнаружению утечек памяти для языков C/C++. Было также

предложено множество новых методов повышения эффективности фаззинга для различных сценариев, включая генерацию сложно структурированных данных и комбинацию статического анализа и символьного выполнения с фаззингом.

Достоверность и обоснованность научных положений и выводов

В рамках диссертационной работы автором проведен глубокий анализ существующих методов анализа кода, в результате которого выявлены основные недостатки. На основе этого исследования предложены пути преодоления существующих ограничений и недостатков. Эффективность предложенных методов подтверждена экспериментальной проверкой, продемонстрировавшей высокую результативность в обнаружении реальных ошибок в открытом ПО. Предложенные автором методы включают статический и динамический анализ программ, а также методологию их универсальной комбинации для выявления сложных дефектов. Для достижения цели автором использованы теория графов, современные методы символьного выполнения и фаззинга, а также теория решеток и компиляции.

Апробация результатов работы

Основные результаты диссертационной работы обсуждались на 12 международных конференциях. По теме диссертации автором опубликовано 24 статьи в журналах, включенных в списки ВАК, Web of Science и Scopus. Кроме того, по результатам диссертационной работы получены 7 свидетельств о государственной регистрации программ для ЭВМ.

Структура и содержание диссертации

Диссертационная работа состоит из введения, семи глав, заключения и списка литературы из 271 наименования. Общий объем работы составляет 268 страниц с 56 рисунками и 35 таблицами.

Во **введении** формулируются цель и задачи работы, проводится обоснование актуальности исследования раскрывается научную новизну и практическую значимость полученных результатов, а также приводятся ключевые положения, выносимые на защиту.

В **главе 1** рассматриваются текущие методы анализа кода и их ограничения. На основе этого выделяются ключевые направления исследования, включая создание платформы для интеграции инструментов анализа программ, разработку средств поиска клонов и копий известных уязвимостей, сопоставление исходного и бинарного кода, выявление ошибок, связанных с форматными строками и динамической памятью, а также оптимизацию методов фаззинга.

В **главе 2** представлен обзор методов анализа программ, соответствующих тематике диссертационной работы. Рассматриваются методы статического и динамического анализа, включая поиск клонов кода, сравнение исполняемых файлов, анализ изменений между версиями ПО, а также фаззинг для различных сценариев.

В **главе 3** представлено описание разработанных методов поиска клонов кода в исходных и бинарных файлах, основанных на анализе графов зависимостей по управлению и данным. Также рассматриваются другие инструменты, разработанные на базе технологии поиска клонов кода, включая выявление неисправленных ошибок и сопоставление исходного и бинарного кода.

Глава 4 посвящена методам обнаружения некорректного использования динамической памяти и способам анализа помеченных данных. Описан механизм создания аннотаций для описания основных функциональностей отдельных функций с целью избегания повторного анализа кода. Кроме того, представлено описание направленного символьного выполнения для верификации обнаруженных ошибок.

В **главе 5** представлено описание разработанных методов фаззинга для различных сценариев применения. Рассматриваются комбинации фаззинга с символьным выполнением и статическим анализом. Также описан разработанный инструмент генерации сложно структурированных данных, который интегрирован в инструмент фаззинга для улучшения покрытия кода при фаззинге различных компиляторов и интерпретаторов.

В **главе 6** представлены детали реализации интеграционной платформы, для объединения разработанных методов анализа программ. Приведены примеры использования доступного программного интерфейса.

Глава 7 посвящена обобщению обнаруженных ошибок в открытом программном обеспечении и коммерческих проектах. Подчеркивается критичность найденных ошибок.

В **заключении** представлены основные результаты работы, а также предложены актуальные направления для дальнейших исследований.

Автореферат правильно отражает содержание диссертации и позволяет достаточно точно оценить основные полученные результаты, степень их обоснованности и достоверности.

Замечания.

1. В разделах 3.1 и 3.2 представлено описание алгоритмов поиска клонов исходного и бинарного кода, в рамках которых графы зависимостей

программ для каждой функции разделяются на единицы сравнения. Затем для каждой пары единиц сравнения выполняется поиск клонов кода как поиск схожих подграфов максимального размера. Однако не описано, каким образом фрагменты клонов объединяются в один крупный блок, поскольку в функции могут быть найдены многочисленные последовательные фрагменты кода, совпадающие с аналогичной последовательностью в другой функции.

2. В разделе 4.3 описан двухэтапный метод поиска утечек памяти, где на втором этапе выполняется направленное символьное выполнение для верификации обнаруженных ошибок. В процессе направленного символьного выполнения значения некоторых переменных автоматически символизируются, что может привести к ложной верификации невыполнимых путей. Автором не проведен анализ таких случаев в реальных условиях.

Указанные замечания не являются критическими и не снижают научную и практическую ценность работы и проведенных исследований.

Заключение. Диссертация Саргсяна Севака Сениковича «Методы оптимизации алгоритмов статического и динамического анализа программ» представляет собой завершенное научное исследование, основное содержание диссертации отражено в опубликованных статьях и обсуждено на научных конференциях. В работе содержатся значительные научные и практические результаты, касающиеся оптимизации алгоритмов анализа программ и их универсальных комбинаций для поиска сложно находимых ошибок.

Диссертационная работа отвечает всем требованиям ВАК РФ, предъявляемым к работам на соискание ученой степени доктора технических

наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей, а ее автор, Саргсян Севак Сеникович заслуживает присуждения ученой степени доктора технических наук по вышеуказанной специальности.

Официальный оппонент
член-корреспондент РАН, доктор технических наук
Заместитель директора по
исследованиям в области
информационных технологий
и вычислительных систем
Федеральное государственное бюджетное учреждение
«Национальный исследовательский центр
«Курчатовский институт»

Шабанов Борис Михайлович

«28» ноября 2024 г.