**УТВЕРЖДАЮ**

Проректор Московского
государственного университета
имени М. В. Ломоносова,
доктор физико-математических наук,
профессор
___ Федянин А. А.
«14» марта 2025 г.

# ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

**Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М. В. Ломоносова», факультета вычислительной математики и кибернетики, кафедры информационной безопасности**

на диссертационную работу Арутюнян Мариам Сероповны «Статический анализ исходного и исполняемого кода на основе поиска клонов кода», представленную к защите на соискание ученой степени кандидата технических наук по специальности 2.3.5 — "Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей".

***Актуальность***. Одним из наиболее распространенных явлений при разработке ПО является клонирование фрагментов кода — копирование и адаптация существующих частей программ. Это может ускорять процесс разработки, но в то же время приводит к ряду проблем: увеличению объема кода, усложнению его сопровождения и распространению уязвимостей.

Одним из критически важных аспектов является распространение уязвимостей через клонированные фрагменты кода и сторонние библиотеки. Если код, содержащий уязвимость, копируется в разные части проекта или даже в другие программные продукты, это увеличивает область потенциальной атаки. Более того, в одном месте уязвимость или ошибка может быть исправлена, но в ее клонах остаться незамеченной, что создает дополнительные риски. Аналогичная проблема возникает при использовании уязвимых библиотек: если в них обнаружена ошибка безопасности, все зависимые программы автоматически подвергаются риску. Поэтому важно не только анализировать собственный код, но и проводить статический анализ подключаемых библиотек, чтобы своевременно выявлять и устранять угрозы безопасности.

Существующие инструменты поиска клонов кода сталкиваются с рядом ограничений. Многие методы анализируют код только на уровне текста или синтаксиса, что приводит к низкой точности обнаружения клонов. Также, большинство инструментов плохо масштабируются и неэффективны при анализе больших программных проектов, содержащих миллионы строк кода.

Исследование, проведенное в диссертации, посвящено решению вышеперечисленных задач и представляет собой важный вклад в область статического анализа программного обеспечения. В связи с этим, разработка эффективных методов обнаружения клонов кода с учетом их семантики, нахождения известных уязвимостей, идентификация используемых библиотек и сравнения различных версий программного обеспечения представляет собой задачу первостепенной важности для обеспечения качества и безопасности современных программных систем.

***Общая характеристика работы.*** Диссертация имеет четкую структуру, включает введение, пять глав и заключение. Полный объем

диссертации составляет 133 страницы, включая список литературы из 202 наименования, 27 рисунков и 24 таблицы.

Во введении обосновывается актуальность работы, формулируются цель и задачи работы, приводятся выносимые на защиту результаты.

Первая глава содержит обзор методов поиска клонов исходного и исполняемого кода, сравнения программ и идентификации статически связанных библиотек. Проведен анализ их преимуществ и недостатков, а также дана оценка применимости в различных сценариях.

Вторая глава посвящена разработке унифицированного метода поиска клонов кода, основанного на графах зависимостей управления и данных. Описаны алгоритмы построения графов, методы их анализа и критерии выявления клонов. Приведены оценки работы метода на разных тестовых системах и реальных программ. Также приведены результаты сравнения с существующими инструментами.

Третья глава описывает метод оптимизации программ путем выявления фрагментов, реализующих вычисление циклической проверки избыточности (ЦПИ). Рассмотрены различные реализации ЦПИ-вычислений, предложен подход к их автоматическому обнаружению и замене на более эффективные реализации. Для оценки эффективности предложенного метода проведены тесты нахождения реализаций ЦПИ в проектах. Также проведены тесты, сравнивающие производительность оригинальных и оптимизированных версий кода.

Четвертая глава рассматривает двухэтапный метод выявления изменений между версиями программ, который сочетает метрический подход и анализ графов зависимостей программы. Рассматривается процесс генерации графов вызовов функций, использование локально-чувствительного хеширования для предварительного сопоставления функций, а разработанный

метод нахождения клонов кода используется для точного анализа изменений. Для повышения точности сопоставления функций применялись методы машинного обучения: на этапе предварительного сопоставления использовалась модель, обученная на парах функций с разным процентом сходства. Веса различных хеш-метрик определялись автоматически, что позволило учесть их разную значимость. Представлены результаты тестирования разработанного инструмента (BCC) на различных версиях Coreutils, включая сравнение с BinDiff и Diaphora, где BCC продемонстрировал более высокую точность и полноту.

Пятая глава описывает методы идентификации статически связанных библиотек и обнаружения уязвимостей с помощью поиска клонов кода. Предложенный подход анализирует исполняемые файлы и библиотеки, выявляя статически связанные версии с высокой полнотой (98–100%) и точностью (89–91%). Метод поиска уязвимостей сравнивает уязвимые фрагменты кода с целевыми проектами, что позволило обнаружить проблемы в Debian 12.6 и ряде GitHub-репозиториев. Из 17 выявленных уязвимостей 9 подтверждены, а 2 отклонены. В целом, результаты подтверждают эффективность предложенных методов.

В заключении приводятся основные результаты работы.

Результаты, полученные в диссертационной работе, соответствуют поставленной цели и сформулированным задачам. Содержание диссертации соответствует требованиям специальности 2.3.5. Текст диссертации и автореферата оформлены в соответствии с требованиями, предъявляемым к диссертационным работам. Автореферат объективно отражает содержание диссертационной работы.

***Основные результаты диссертационной работы, обладающие научной новизной.*** Научная новизна работы заключается в следующем:

- Разработан унифицированный метод поиска клонов кода, основанный на графах зависимостей управления и данных, что позволяет анализировать и сопоставлять фрагменты кода разного уровня представления (исходного и исполняемого) и разной степени отличия. Который обладает высокой точностью, полнотой и производительностью для анализа десятков миллионов строк исходного и соответствующего исполняемого кода.
- Предложен и реализован метод оптимизации программ за счет автоматического выявления и замены неэффективных реализаций алгоритмов ЦПИ.
- Разработан и реализован двухэтапный метод выявления изменений между версиями программ с использованием метрического подхода и разработанным методом поиска клонов кода на основе графов, для сопоставления функций в формате «многие ко многим» и отображения измененных инструкций.
- Разработаны и реализованы методы идентификации статически связанных библиотек и поиска копий известных уязвимостей, основанные на обнаружении клонов кода.

***Достоверность и обоснованность научных положений и выводов.*** Достоверность и обоснованность результатов исследования подтверждается проведенным тестированием, внедрением разработанных программных продуктов в различные системы анализа кода, публикациями в рецензируемых научных изданиях, докладами в международных конференциях, а также прошедшими экспертизами работы, при получении грантов.

Непосредственно по теме диссертации автором опубликовано 12 работ, из них 4 — в журналах из списка ВАК, 4 — в журналах, индексируемых в Scopus,

и 2 — в журналах, входящих в первый квартиль по SJR. Получено 2 свидетельства о регистрации программ для ЭВМ.

***Теоретическая и практическая значимость.*** Теоретическая значимость работы заключается в разработке новых алгоритмов анализа кода, обеспечивающих высокую точность при выявлении клонов. Работа вносит значимый вклад в развитие статического анализа программного обеспечения. Полученные результаты способствуют расширению инструментальных средств для обеспечения надежности и безопасности программных систем, комплексов и компьютерных сетей.

Практическая ценность работы подтверждается интеграцией разработанного метода оптимизации ЦПИ в основную ветку компилятора GCC, что свидетельствует о его высокой применимости и востребованности в компиляторных технологиях. Кроме того, разработанные методы внедрены в системы GenesISP и BinSide. Применение этих методов способствует повышению безопасности программного обеспечения и снижению рисков, связанных с уязвимостями.

Разработанные инструменты могут применяться в жизненном цикле разработки безопасного ПО, а также в сертификационных центрах и компаниях, использующих этот процесс, что соответствует требованиям ГОСТ Р 56939-2024 «Защита информации. Разработка безопасного программного обеспечения. Общие требования» и «Методики выявления уязвимостей и недекларированных возможностей в программном обеспечении» ФСТЭК Российской Федерации.

***Замечания.*** Несмотря на высокий уровень выполнения работы, можно отметить следующие недостатки:

1. В разделе 2.3. рассматривается возможность генерации графов зависимостей программы тремя разными способами для исходного кода,

и одним способом для исполняемого кода в зависимости от задачи, но не производится оценка замедления инструмента в каждом случае.

2. Не приведены результаты сравнения предлагаемой в 5.2. разделе инструмента нахождения известных уязвимостей с существующими аналогами.
3. В работе упоминается применимость основных результатов диссертации для использования в процессах разработки программ в соответствии ГОСТ Р 56939-2024 «Разработка безопасного программного обеспечения. Общие требования», но не приведены конкретные методические рекомендации для такого использования.

***Заключение.*** Диссертационная работа Арутюнян Мариам Сероповны является законченной научно-квалификационной работой, содержащей новые научные результаты, имеющие теоретическую и практическую значимость. Положения, выносимые на защиту, апробированы и в достаточной мере освещены в научной печати, в том числе, в изданиях из перечня рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени кандидата наук, обсуждены на международных научных конференциях. Автореферат отражает основные научные положения и выводы, сделанные в диссертации. Работа полностью соответствует требованиям, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5 — "Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей".

Диссертационная работа и отзыв рассмотрены и утверждены на заседании кафедры информационной безопасности факультета вычислительной

математики и кибернетики Московского государственного университета имени М. В. Ломоносова (протокол №1 от 19 марта 2025 года).

Ученый секретарь кафедры
информационной безопасности
факультета ВМК
МГУ имени М. В. Ломоносова
К.ф.-м.н., доцент И. В. Чижов

Заведующий кафедрой
информационной безопасности
факультета ВМК
МГУ имени М. В. Ломоносова
академик И. А. Соколов

Заместитель декана по научной работе
факультета ВМК
МГУ имени М. В. Ломоносова
д.ф.-м.н., профессор В. В. Фомичев

Сведения об организации:
Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова», факультет вычислительной математики и кибернетики
Адрес: 119991, г. Москва, Ленинские горы 1с52
Телефон: +7 (495) 939-30-10
E-mail: cmc@cs.msu.ru
Веб-сайт: https://cs.msu.ru/