ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Шимчика Никиты Владимировича
по теме **«Исследование и разработка методов поиска уязвимостей в программах на C и C++ на основе статического анализа помеченных данных»**,
представленную к защите на соискание учёной степени кандидата технических наук по специальности 2.3.5 – "Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей"

Актуальность темы

Рост размера и сложности разрабатываемых программ приводит к появлению большого количества ошибок и дефектов в исходном коде – некоторые из этих дефектов являются уязвимостями. Уязвимости позволяют злоумышленнику вмешиваться в работу программы, приводя к утечкам данных, отказу в обслуживании или другим негативным последствиям для пользователей. Из-за больших объёмов кода, полная ручная проверка является невозможной, потому задача разработки инструментов, автоматизирующих поиск уязвимостей актуальной является актуальной.

Одним из способов обобщения различных типов уязвимостей является задача анализа помеченных данных, которая исследует зависимости по данным между выделенными функциями и может решаться средствами как статического, так и динамического анализа. Особенностью большинства методов статического анализа является более высокая полнота за счёт проверки всех путей выполнения программы, а не только тех, для которых удалось найти соответствующие входные данные, что позволяет находить больше уязвимостей.

Для обнаружения таких опасных уязвимостей как Heartbleed на практике недостаточно существующих методов из-за количества и длины исследуемых путей выполнения программы. Одним из подходов к решению этой проблемы является её сведение к специальной задаче о достижимости на межпроцедурном графе – так называемой задаче IFDS. Такой подход позволяет эффективно исследовать пути распространения помеченных данных от истоков к стокам, отслеживая только сами помеченные данные.

Основными препятствиями к практическому использованию этого подхода является низкая точность, одной из причин которой является отсутствие у анализа чувствительности к путям. Также возникает проблема масштабирования, которая заключается во времени анализа и объёме оперативной памяти, необходимых для обработки больших проектов, и общая для всех методов статического анализа проблема полноты, вызванная недостатком информации об анализируемой программе и проявляющаяся в необходимости моделировать поведение библиотечных функций, косвенных и виртуальных вызовов и не только.

Разработка методов, повышающих точность, масштабируемость и полноту статического анализа помеченных данных является актуальной задачей, поскольку это улучшает практическую применимость инструментов для поиска уязвимостей в исходном коде программ и повышает безопасность использования анализируемых программ.

Теоретическая и практическая значимость полученных результатов, их научная новизна

Основными результатами, выносимыми на защиту, являются:

1. Алгоритм статического анализа косвенных вызовов на основе решения задачи IFDS, добавляющий в качестве кандидатов для вызова те функции, адреса которых могли достичь точки вызова. Применение этого алгоритма позволяет повысить полноту анализа помеченных данных без появления ложных кандидатов.
2. Алгоритм снятия помеченности с целочисленных переменных, значение которых ограничено константой. Данный алгоритм не зависит от конкретного пути выполнения программы и подходит для использования в нечувствительном к путям анализе, каким является алгоритм решения задачи IFDS. Применение этого алгоритма позволяет повысить точность анализа.
3. Алгоритм направленного распространения помеченности через глобальные переменные, позволяющий ускорить анализ проектов, в которых помеченные значения записываются в глобальные переменные. Применение этого алгоритма позволяет повысить масштабируемость анализа.
4. Метод проверки консистентности путей на основе обхода графа распространения помеченности. Применение этого метода позволяет убедиться, что из возможных путей от истока к стоку помеченности будет выбран тот, который не содержит неконсистентного выбора кандидатов для вызова, что повышает точность анализа.

Теоретическая значимость работы состоит в разработке новых методов и алгоритмов, повышающих точность, масштабируемость и полноту статического анализа помеченных данных.

Практическая значимость полученных результатов подтверждается наличием 5 зарегистрированных программ для ЭВМ. В рамках данной работы, на основе предложенных методов и алгоритмов был разработан инструмент Irbis, осуществляющий поиск уязвимостей при помощи статического анализа помеченных данных, который поставляется заказчикам в составе набора инструментов статического анализа ИСП РАН.

Структура диссертации

Диссертация состоит из введения, пяти глав, заключения и списка литературы из 90 наименований. Общий объём диссертации составляет 108 страниц с 5 рисунками и 9 таблицами.

Во **введении** формулируются цели и задачи диссертационной работы, обосновывается актуальность темы исследования, теоретическая и практическая значимость работы, обозначается научная новизна и основные положения, выносимые на защиту.

В **главе 1** приводится обзор предметной области по теме диссертации, рассматриваются и сравниваются существующие методы и инструменты поиска

уязвимостей, в частности подходы на основе статического символьного выполнения и анализа потоков данных.

Глава 2 посвящена методам и алгоритмам повышения точности анализа помеченных данных. В ней приводится общая схема работы анализатора, даётся описание подхода на основе двухэтапного анализа, позволяющего проверять выполнимость находимых основным инструментом путей распространения помеченных данных при помощи статического символьного выполнения. Предлагается метод проверки консистентности косвенных и виртуальных вызовов вдоль путей в графе распространения помеченности. Предлагается алгоритм снятия помеченности с целочисленных переменных, позволяющий уменьшить избыточную помеченность и устранить часть ложных срабатываний. Для всех предложенных методов и алгоритмов в главе приведены результаты экспериментальной оценки.

Глава 3 посвящена методам и алгоритмам повышения полноты анализа. В ней предлагается алгоритм анализа косвенных вызовов на основе решения задачи IFDS, а также эвристика, позволяющая обнаруживать новые функции, осуществляющие чтение или освобождение данных. Для всех предложенных методов и алгоритмов в главе приведены результаты экспериментальной оценки.

Глава 4 посвящена повышению масштабируемости анализа. В ней даётся теоретическое обоснование большего влияния помеченных глобальных переменных на время анализа, по сравнению с помеченными локальными значениями, при решении задачи IFDS. Предлагается алгоритм направленного распространения помеченности через глобальные переменные. приводится теоретическое обоснование корректности предложенного алгоритма и экспериментальная оценка его влияния на время анализа.

Глава 5 посвящена программной реализации описанных методов в анализаторе Irbis и содержит общую схему работы этого инструмента, список детекторов и тегов предупреждений, описывает некоторые особенности реализации и проводит общую оценку его работы на тестовом наборе Juliet Test Suite и реальных проектах с открытым исходным кодом. Помимо прочего, в данной главе проводится сравнение работы инструмента с другими анализаторами помеченных данных в инструментах Svace, Clang Static Analyzer и Infer Static Analyzer.

В **заключении** приведены основные результаты работы, а именно: разработаны методы повышения точности, полноты и масштабируемости статического анализа помеченных данных на основе IFDS, включая алгоритм анализа косвенных вызовов, алгоритм направленного распространения помеченности, алгоритм снятия помеченности с целочисленных переменных и метод проверки консистентности путей.

В целом диссертация Н.В. Шимчика является законченным исследованием, представляет решение актуальных задач, объединённых общим подходом, обеспечивающим возможность практического применения статического анализа для поиска потенциальных уязвимостей в реальных проектах.

Апробация результатов работы

Результаты, полученные в ходе диссертационной работы, представлены в восьми публикациях, 4 из которых опубликованы в журналах, рекомендованных ВАК, 1 из

которых индексируется Scopus. Ещё 4 опубликованы в сборниках статей. Получены 5 свидетельств о государственной регистрации программ для ЭВМ.

Все статьи подготовлены в соавторстве с научным руководителем и другими авторами, в шести из них соискателю принадлежит определяющее участие. Результаты по разработке статического анализатора помеченных данных на основе символьного выполнения в инструменте SharpChecker, а также основные результаты, относящиеся к разработке инструмента статического анализа Irbis, получены лично автором.

Основные результаты диссертационного исследования прошли апробацию на шести конференциях:

- Открытая конференция ИСП РАН имени В.П. Иванникова 2021, 2022 и 2023 годов.
- Международная конференция Spring/Summer Young Researchers' Colloquium on Software Engineering 2019 года.
- "Ломоносовские чтения", секция "Вычислительная математика и кибернетика", 2018 и 2020 годов

Замечания

За исключением главы 5, в большинстве таблиц приводится оценка результатов работы алгоритмов только для проектов Juliet Test Suite или OpenSSL. Хотелось бы видеть оценки для всех четырех наборах.

Недостаточно формализована запись алгоритмов. В алгоритмах 2 и 3 формальная запись с использованием псевдокода перемежается с шагами, описанные текстом на русском языке, причём эти шаги могут быть нетривиальными и вносить существенный вклад в общее время работы алгоритма. Алгоритм 1 изначально написан в виде текста.

В работе не приведено описание используемой классификациии уязвимостей CWE (Common Weakness Enumeration).

В тексте работы представляется недостаточно обоснованным выбор поддерживаемых типов уязвимостей, в частности, перечисленного в разделе 5.4 списка из 9 ошибок классификации CWE, на которых проводилась оценка результатов работы инструмента.

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

Заключение

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

Диссертационная работа Шимчика Никиты Владимировича по теме «Исследование и разработка методов поиска уязвимостей в программах на C и C++ на основе статического анализа помеченных данных» является законченным научным исследованием, основное содержание диссертации отражено в опубликованных статьях и обсуждено на научных конференциях.

Диссертационная работа отвечает требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, её содержание соответствует паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», а её автор, Шимчик Никита Владимирович, заслуживает присуждения учёной степени кандидата технических наук по вышеуказанной специальности.

Официальный оппонент
член-корреспондент РАН, доктор технических наук, доцент,
директор Межведомственного Суперкомпьютерного Центра РАН,
заместитель директора по научной работе
Федерального государственного учреждения
«Федеральный научный центр Научно-исследовательский
институт системных исследований
Российской академии наук»

Шабанов Борис Михайлович

«27» мая 2024 г.
119334 Москва, Ленинский проспект, 32а,