УТВЕРЖДАЮ
Заместитель директора по научной работе
ИПМ им. М.В. Келдыша РАН,
член-корреспондент РАН, д.ф.-м.н., профессор

М.В. Якобовский
«24» мая 2024 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук» на диссертацию Шимчика Никиты Владимировича на тему «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 диссертационной работы

Диссертационная работа Шимчика Н.В. посвящена разработке методов и алгоритмов, осуществляющих поиск уязвимостей на основе статического анализа помеченных данных с низким процентом пропущенных и ложных срабатываний, а также масштабируемостью на проекты размером в миллионы строк кода. Методы автоматизации поиска уязвимостей в программах являются важной темой в современной индустрии разработки программ из-за роста объёма программного обеспечения и потенциального ущерба, который может быть причинён злоумышленником при помощи уязвимостей, которые не были своевременно обнаружены и исправлены. Ярким примером является Heartbleed – уязвимость в библиотеке OpenSSL, которая была найдена только спустя 3 года после появления и позволяла злоумышленнику получать данные из памяти сервера при помощи отправки некорректно сформированного сообщения.

Один из распространённых подходов к поиску потенциальных уязвимостей – это сведение их к задаче анализа помеченных данных. В данной работе используется подход на основе решения задачи IFDS, которую можно решить за полиномиальное время, обеспечивающий высокую полноту анализа. Среди недостатков этого подхода можно назвать отсутствие чувствительности к путям, что приводит к снижению точности анализа, а также высокое потребление памяти на больших проектах, что приводит к ограничению масштабируемости. Для решения указанных проблем в диссертационной работе предложены методы и алгоритмы, повышающие полноту, точность и масштабируемость анализа помеченных данных за счёт анализа косвенных вызовов, автоматического поиска источников помеченных данных, проверки консистентности найденных путей, алгоритма снятия помеченности с целочисленных

переменных, направленного распространения помеченности через глобальные переменные и некоторых других способов.

Существующие инструменты статического анализа помеченных данных либо предназначены для высокоуровневых языков, либо испытывают проблемы с нахождением уязвимостей на практике, если они проявляются на сложных межпроцедурных путях в реальных программах. В связи с этим в рамках данной работы был разработан и реализован инструмент статического анализа помеченных данных Irbis, предназначенный для поиска более чем десятка классов потенциальных уязвимостей и демонстрирующий практическую способность обнаруживать уязвимость Heartbleed. Таким образом, тема диссертационной работы Шимчика Н.В. является актуальной.

В диссертации представлены результаты экспериментального исследования, демонстрирующие как результаты применения отдельных предлагаемых алгоритмов и методов, так и оценивающие результаты работы инструмента в целом на тестовом наборе Juliet Test Suite и реальных проектах. Также в работе приведены результаты сравнения с тремя другими инструментами статического анализа помеченных данных.

Общая характеристика работы

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** сформулирована цель работы и постановка задачи, продемонстрирована актуальность и научная новизна темы работы, раскрыта научная и практическая значимость, сформулированы основные положения, выносимые на защиту, описана структура диссертации.

В **первой** главе проведён обзор существующих методов поиска уязвимостей, приведены их достоинства и недостатки, а также вводятся основные термины, связанные со статическим анализом помеченных данных на основе задачи IFDS.

Во второй главе приводится общая схема работы анализатора, а также описываются предлагаемые методы и алгоритмы, предназначенные для повышения точности анализа. Для устранения ложных срабатываний, вызванных отсутствием чувствительности к путям у алгоритма решения задачи IFDS, предложены подходы на основе двухэтапного анализа. В одном варианте обнаруженные пути распространения помеченных данных передаются на вход инструментам, осуществляющим статический или динамический анализ на основе символьного выполнения, что позволяет проверить выполнимость условий на интересующих путях выполнения программы. Другой вариант предполагает отдельный проход по построенному графу распространения помеченности от истока к стоку, который позволяет проверить консистентность путей с точки зрения конкретных критериев: согласованность выбора кандидатов для косвенных или виртуальных вызовов вдоль пути выполнения программы. Для уменьшения количества ложных срабатываний, вызванных избыточной помеченностью, предлагается алгоритм снятия помеченности с целочисленных переменных. Также в данной главе анализируются распространённые типы ложных срабатываний, встречающиеся при применении анализа помеченных данных к поиску уязвимостей, вызванных ошибками работы с памятью. Для

предложенных в главе методов и алгоритмов приводится экспериментальная оценка, демонстрирующая повышение точности на тестовом наборе Juliet и реальных проектах.

Третья глава посвящена методам и алгоритмам, предназначенным для повышения полноты анализа. Рассматриваемыми в данной работе проблемами, приводящими к снижению полноты анализа, являются косвенные вызовы и отсутствие информации о реализациях библиотечных функций, вызываемых в программе. Для решения первой проблемы предлагается алгоритм анализа косвенных вызовов на основе IFDS, который позволяет добавлять только тех кандидатов для вызова, адреса которых действительно могут использоваться в точке вызова, в отличие от подхода на основе типов параметров функций. Для решения проблемы библиотечных функций приводится формат спецификаций, позволяющий разработчику и пользователям описывать поведение функций с точки зрения анализа помеченных данных, использующий регулярные выражения, учитывающий возможность перегрузки имён в языке C++ и позволяющий абстрагироваться от конкретной реализации классов, хранящих данные. Помимо ручного создания спецификаций, в данной главе предлагается эвристика на основе названия и типов параметров функций, которая позволяет обнаруживать новые истоки, осуществляющие чтение или освобождение данных. Для всех методов и алгоритмов приведена экспериментальная оценка, демонстрирующая повышение полноты анализа и оценивающая их влияние на масштабируемость.

Четвёртая глава посвящена повышению масштабируемости анализа. В ней приводится теоретическое обоснование того, что при решении задачи IFDS помеченные глобальные переменные вносят больший вклад в сложность алгоритма, чем локальные переменные и параметры функций. Также в данной главе предлагается алгоритм направленного распространения помеченности через глобальные переменные, приводится обоснование его корректности и экспериментальная оценка, демонстрирующая двукратное ускорение анализа на тестовом наборе и незначительное ускорение анализа на реальном проекте. Помимо этого в данной главе описываются некоторые изменения в использовании решателя задачи IFDS, которые потребовались для практической применимости данного метода для анализа проектов в миллионы строк кода.

В **пятой** главе описываются подробности реализации инструмента анализа помеченных данных Irbis, приводится схема его работы, перечисляются типы реализованных детекторов и особенности реализации, а также приводится общая оценка результатов его работы на тестовом наборе Juliet Test Suite и четырёх реальных проектах с открытым исходным кодом. Помимо общей оценки полноты, точности и масштабируемости инструмента, приводятся результаты его сравнения с тремя другими анализаторами, реализующими анализ помеченных данных.

В **заключении** сформулированы основные результаты диссертационной работы.

Достоверность полученных результатов

Достоверность полученных результатов подтверждается экспериментальной и теоретической проверкой работоспособности предложенного подхода, а также апробацией на семинарах, конференциях различного уровня и 8 научными статьями, 4

из которых изданы в журналах, рекомендованных ВАК, 1 из которых входит в индекс Scopus, а 4 опубликованы в сборниках статей. Кроме этого, получены 5 свидетельств о государственной регистрации программ для ЭВМ.

Основные научные результаты и их значимость для науки и практики

В рамках данной работы получены следующие основные результаты, выносимые на защиту:

1. Алгоритм анализа косвенных вызовов на основе решения задачи IFDS.
2. Алгоритм снятия помеченности с целочисленных переменных.
3. Алгоритм направленного распространения помеченности через глобальные переменные.
4. Метод проверки консистентности путей на основе обхода расширенного суперграфа.

На основе указанных алгоритмов и методов был разработан инструмент статического анализа Irbis, осуществляющий поиск уязвимостей путём анализа помеченных данных согласно решению задачи IFDS. Теоретическая значимость результатов заключается в разработанных методах и алгоритмах, которые дополняют существующие методы статического анализа и решают отдельные проблемы анализа помеченных данных на основе решения задачи IFDS. Практическая значимость полученных результатов заключается в реализации инструмента, пригодного для поиска потенциальных уязвимостей с процентом истинных срабатываний не менее 50%, масштабируемостью на проекты в миллионы строк кода и способностью находить такие уязвимости как Heartbleed. Этот инструмент интегрирован в промышленный статический анализатор Svace и внедрён.

Замечания

1) Для алгоритмов анализа косвенных вызовов и снятия помеченности с целочисленных переменных не приведены обоснование корректности и оценка вычислительной сложности.
2) В работе приведены таблицы, демонстрирующие вклад каждого из предложенных алгоритмов по отдельности, но не приведены сравнительные результаты их совместной работы.

Заключение

Указанные недостатки не влияют на общую положительную оценку диссертации и не ставят под сомнение полученные в ней результаты. Диссертационная работа «Исследование и разработка методов поиска уязвимостей в программах на C и C++ на основе статического анализа помеченных данных» является законченным научным исследованием по актуальной тематике, выполненным на высоком уровне. Название диссертации соответствует основному содержанию диссертации. Автореферат достаточно полно отражает содержание работы.

Работа удовлетворяет требованиям ВАК РФ, предъявляемым к работам на соискание степени кандидата технических наук по 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 указанной специальности.

Отзыв на диссертацию подготовлен на основании заключения структурного подразделения «Отдел программного обеспечения высокопроизводительных вычислительных систем и сетей» по результатам проведённого обсуждения диссертации и заслушивания доклада Н.В. Шимчика в ИПМ им. М.В. Келдыша РАН на заседании семинара «Программирование» им. М.Р. Шура-Бура 25 апреля 2024 г.

Заведующий Информационно-издательским отделом
д.ф.-м.н.

Горбунов-Посадов М. М.

Сведения об организации:
Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»
Адрес: 125047, город Москва, Миусская пл., д. 4
Телефон: +7 499 978-13-14
E-mail: office@keldysh.ru
Веб-сайт: https://www.keldysh.ru