ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Михайлова Андрея Анатольевича «Методы декомпиляции объектного кода Delphi», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.11 — Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Обоснование актуальности исследования

Использование готовых программных компонентов для создания прикладного программного обеспечения, поставляемых в виде библиотек скомпилированных программных модулей, является общепринятой и широко распространенной практикой разработки сложных программных систем. Такой подход, используемый в большинстве современных высокоуровневых средств разработки, позволяет сократить сроки и стоимость разработки, но существенно усложняет верификацию и валидацию программ, проверку на отсутствие недокументированных функций и уязвимостей, а также создает трудности модификации и сопровождения унаследованного программного обеспечения из-за невозможности доступа прикладного программиста к исходному коду используемых скомпилированных модулей.

Для верификации и валидации программных систем в отсутствие исходных текстов программ на практике используют средства обратной разработки – дизассемблеры, позволяющие восстановить последовательность команд процессора из исполняемого (скомпилированного) модуля. Полученный в результате восстановления код программы на языке ассемблера характеризуется очень большим объемом, который для современных сложных программных систем может составлять десятки мегабайт исходного кода. Поскольку каждому оператору языка программирования высокого уровня соответствует серия команд на языке ассемблера, восстановленный (декомпилированный) код обладает сложной и запутанной структурой, так как компиляторы всех языков программирования высокого уровня (C, C++, C#, ObjectPascal/Delphi, .NET и др.) при оптимизации программы зачастую используют команды безусловных переходов и вызовы стандартных функций по адресу в ячейке памяти, что повышает эффективность выполнения программы, но многократно усложняет восстановленный текст на ассемблере.

Большой объем получаемого низкоуровневого кода на языке ассемблера и сложность его структурирования существенно затрудняют анализ восстановленной программы и выдвигают очень высокие требования к квалификации специалиста. Программные средства декомпиляции, позволяющие восстановить исходный (или близкий к нему) текст программы на

языке программирования высокого уровня, в настоящее время практически отсутствуют, а существующие средства обладают целым рядом серьезных ограничений по языку программирования исходной программы, версии используемого компилятора, версии библиотек и средств оптимизации для их практического применения.

Сложившееся противоречие между практической необходимостью восстановления исходных текстов программ на языке высокого уровня для верификации и валидации программ и существующими средствами низкоуровневого анализа программ на языке ассемблера обуславливает актуальность научной задачи разработки новых методов и средств декомпиляции исполняемых и объектных кодов для восстановления исходного текста на языке программирования высокого уровня.

Таким образом, **тема диссертационного исследования**, посвященная разработке методов декомпиляции и анализа объектного кода Delphi, и **научная задача** создания методов восстановления исходного текста программы на языке высокого уровня Delphi для скомпилированного под платформу .NET объектного кода, решенная в диссертационной работе Михайлова А.А., являются актуальными и имеют важное практическое значение для верификации и валидации программных систем при отсутствии исходных текстов программ.

Научная новизна основных научных результатов диссертации

В диссертации получены следующие новые научные результаты.

Соискателем разработан оригинальный метод декомпиляции объектного кода Delphi для платформы .NET, отличающийся использованием синтаксического анализа объектного кода Delphi для построения промежуточного представления программы и управляющего графа, в результате анализа потока управления которого осуществляются структурирование кода подпрограмм и анализ их потоков данных для генерации кода на языке программирования высокого уровня.

На основе предложенного метода анализа потоков управления диссертантом разработан новый метод визуализации управляющего графа на плоскости, отличающийся использованием соглашений для отображения блок-схем для эффективного структурного отображения графового представления высокоуровневых операторов восстанавливаемой программы.

Теоретическая значимость результатов работы

Работа автора и результаты проведенных исследований дополнили теорию программирования в части методов проектирования, анализа и верификации программ и программных систем новыми методами декомпиляции объектного

кода Delphi для платформы .NET, позволяющими осуществлять обратные эквивалентные преобразования программ – восстанавливать исходный текст на языке высокого уровня Delphi.

Практическая значимость результатов работы

Разработанные во второй главе диссертации методы декомпиляции объектного кода Delphi доведены до практической реализации декомпилятора объектных модулей Delphi и модуля структурной раскладки графов потоков управления, что подтверждается соответствующими свидетельствами о государственной регистрации программ для ЭВМ. С помощью разработанного декомпилятора была разобрана стандартная библиотека визуальных компонентов Delphi 8, при этом в 98,7% случаев удалось получить структурное высокоуровневое представление программы без операторов непосредственного перехода.

Разработанная автором программная реализация декомпилятора объектных файлов Delphi для платформы .NET обеспечивает восстановление представленной на низкоуровневом языке CIL программы в текст программы на языке Delphi и будет интересна специалистам в области разработки и анализа программного обеспечения.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

В работе обоснована теоретическая и практическая значимость исследования, раскрыто содержание научной проблемы. Полученные в диссертации теоретические результаты, выводы и рекомендации в достаточной степени обоснованы с помощью корректного применения основ теории компиляции, теории графов и методов объектно-ориентированного программирования. Полученные теоретические результаты, выводы и рекомендации подтверждены проведенными автором экспериментами и результатами практического использования.

Высокий уровень обоснованности также подтверждается 16 печатными работами, 3 из которых являются публикациями в журналах, рекомендованных ВАК РФ, 1 статья из списка Web of Science, а также апробацией полученных научных результатов на 8 международных и всероссийских конференциях. По теме диссертации получено два свидетельства о государственной регистрации программ для ЭВМ.

Достоверность полученных в диссертации результатов

К наиболее существенным новым научным результатам диссертационного исследования можно отнести:

- методы декомпиляции объектного кода Delphi для платформы .NET;
- метод визуализации управляющего графа на плоскости с учетом правил отображения блок-схем.

Достоверность результатов не вызывает сомнений и подтверждается непротиворечивостью математических выкладок, собственными вычислительными экспериментами соискателя, внедрением результатов работы и сравнительным анализом результатов российских и зарубежных ученых по тематике диссертационного исследования.

Внедрение результатов работы

Материалы диссертационной работы использованы в рамках проекта СО РАН IV.38.2.3 «Новые методы, технологии и сервисы обработки пространственных и тематических данных, основанные на декларативных спецификациях и знаниях», проекте РФФИ №15-37-20042-мол и в учебном процессе кафедры «Информационных технологий» ИМЭИ ИГУ.

Апробация и публикации материалов работы

Основные результаты работы докладывались и получили положительные отзывы на 8 международных и всероссийских конференциях: XIII Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям (г. Новосибирск, 2012 г.); II Российско-Монгольской конференции молодых ученых (п. Ханх, Монголия, 2013 г.); XVIII Байкальской Всероссийской конференции «Информационные и математические технологии в науке и управлении» (г. Иркутск, 2013 г.); Малых Винеровских (г. Иркутск, 2013) и Ляпуновских (г. Иркутск, 2012, 2013, 2014 гг.) чтениях; III Российско-монгольской конференции молодых ученых по математическому моделированию, вычислительно-информационным технологиям и управлению (п. Ханх, Монголия, 2015 г.); 5th InternationalWorkshop on Computer Science and Engineering (г. Москва, 2015 г.); 39th International ICT Convention – MIPRO (г. Опатия, Хорватия, 2016 г.).

Объем работы и публикации

Диссертация состоит из введения, четырёх глав, заключения и библиографии. Общий объём диссертации 155 страниц, из них 108 страниц текста, включая 19 рисунков. Библиография включает 98 наименований на 10 страницах.

В первой главе представлен обзор существующих методов и средств автоматического восстановления программ на языках высокого уровня из программ на языках низкого уровня, поставлена проблема декомпиляции в общем виде, рассмотрены известные типы декомпиляторов. Отмечено, что существующие декомпиляторы Delphi обладают рядом недостатков, ограничивающих их возможности. Рассмотрены формат и специфика объектного кода Delphi для платформы .NET.

Во второй главе для декомпиляции объектного кода Delphi предлагаются: методы лексического анализа и промежуточного представления подпрограмм, методы генерации управляющего графа и анализа потоков управления, методы структурного анализа и анализа потоков данных программ. Для разработанных методов приведены: алгоритмы разбора блоков кода процедур, алгоритм генерации базовых блоков управляющего графа, итерационный алгоритм поиска доминаторов, алгоритмы структурирования, структурного анализа и анализа потоков данных.

В третьей главе рассматривается практическая реализация предложенных во второй главе методов декомпиляции в инструментальном программном средстве анализа объектного кода Delphi – декомпиляторе DCUIL2PAS. Представлено описание архитектуры декомпилятора и реализованных модулей и процедур загрузки объектных файлов, дизассемблирования, генерации выражений и управляющего графа программы, восстановления высокоуровневых операторов, процедур и функций с помощью графа регионов. Предложена оптимизация восстанавливаемого исходного кода для сокращенного вычисления логических выражений при генерации кода программы на языке высокого уровня. Проведено исследование эффективности декомпиляции на тестовой задаче восстановления исходного текста программы сжатия. Приведенные результаты экспериментов в соответствии с введенной мерой качества декомпиляции показывают примерно двукратное преимущество разработанного автором декомпилятора по качеству восстановления исходного кода при оперативно-приемлемом времени восстановления.

В четвертой главе рассматривается применение разработанных методов для визуализации управляющего графа, создаваемого разработанным декомпилятором при восстановлении текста программы на языке высокого уровня из объектного кода для платформы .NET. Визуальное представление потока управления программы необходимо для структурирования программы с учетом особенностей восстанавливаемых графов для улучшения его восприятия при последующем структурном анализе и анализе потоков данных. Предложены критерии качества визуализации для блоков действия, логических блоков, границ цикла и блоков начала и конца процесса. Описан метод поуровневого

изображения графов, состоящий из распределения вершин управляющего графа по уровням, определения порядка и координат вершин на уровнях и проведения дуг. Описан предложенный автором алгоритм структурирования регионов, выделенных в соответствии с описанными во второй главе методами, и их раскладки для последовательного отображения. Приведены результаты экспериментальной проверки разработанного структурного раскладчика на тестах SPEC CPU2000.

По теме диссертации опубликовано 16 печатных работ: из них 5 статей, 3 из которых - в изданиях, рекомендованных ВАК для публикации основных результатов и положений диссертации, 1 статья - в издании, включенном в наукометрическую базу данных Web of Science. В публикациях отражено основное содержание результатов диссертационного исследования.

Замечания

1) В тексте диссертации (главы 2 и 4) при описании методов автор зачастую не разделяет анализ известных решений и разработку новых научных результатов.
2) В формулировках результатов, выдвигаемых для защиты, автор не удосужился привести отличительные признаки полученных научных результатов.
3) Не обоснован выбор процедур алгоритма LZW и библиотеки VCL в качестве объектов тестирования разработанного метода декомпиляции и реализующего его программного средства.
4) Не обоснован выбор тестов 197.parser и 252.eon из всей библиотеки SPEC CPU2000 для тестирования структурного раскладчика атрибутивных графов потоков управления.
5) В тексте диссертации часто используется иллюстративный материал без ссылок (например, таблица 1.1 на с.21, листинг 1.2 на с. 27, таблица 2.2 на с.46, рис. 2.8 на с.59).

Заключение по работе

Диссертационная работа Михайлова Андрея Анатольевича на тему «Методы декомпиляции объектного кода Delphi» представляет собой самостоятельную, целостную, завершенную научно-квалификационную работу, содержащую решение актуальной научной задачи, имеющей научную и практическую ценность.

Содержание диссертации соответствует п. 1 («Модели, методы и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования») областей

исследования паспорта специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Содержание диссертации Михайлова Андрея Анатольевича подробно раскрывает решение поставленной научной задачи. Представленные в диссертационной работе результаты апробации позволяют сделать вывод о достаточной степени ее проработки, наличии в работе новых научных результатов, их достоверности и обоснованности.

Высказанные замечания несколько снижают ценность выполненного научного исследования, но не умаляют теоретической и практической значимости работы и научной новизны, заявленных автором.

Диссертация отвечает требованиям пп. 9-14 «Положения о порядке присуждения ученых степеней» ВАК Минобрнауки РФ, утвержденного Постановлением Правительства РФ от 24.09.2013 N 842 (в редакции от 28.08.2017), предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор, Михайлов Андрей Анатольевич, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент –
кандидат технических наук А.И. Дордопуло
«21» ноября 2017 г.

Сведения об оппоненте: Дордопуло Алексей Игоревич
347900, г. Таганрог, пер. Итальянский, 106. Телефон: (918) 561-16-29
e-mail: dordopulo@superevm.ru
ООО «НИЦ супер-ЭВМ и нейрокомпьютеров», начальник отдела математического и алгоритмического обеспечения

Подпись Дордопуло А.И. заверяю.
Начальник отдела кадров НИЦ СЭ и НК _____А.В. Коваленко
М.П.