Отзыв научного руководителя на диссертацию

Михайлова Андрея Анатольевича

«Методы декомпиляции объектного кода Delphi»

представленную на соискание учёной степени кандидата технических наук по специальности 05.13.11 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Задача декомпиляции программного кода до сих пор не решена в полном объёме. Хотя существуют примеры декомпиляторов, корректно восстанавливающих исходный код для некоторых типов файлов, для исполняемых файлов, содержащих машинный код, возможности всех существующих реализаций декомпиляторов очень ограничены, что не позволяет использовать результаты их работы на практике. Сложность разработки таких декомпиляторов объясняется тем, что для полноценной декомпиляции исполняемых файлов требуется решить такие задачи, как: разделение кода и данных, учёт семантики машинных команд, вывод типов, распознавание системных библиотек. При этом, например, задача разделения кода и данных в общем случае является алгоритмически неразрешимой.

Объектные файлы содержат информацию о программном коде и данных, необходимую для сборки исполняемого файла редактором связей. Эти файлы более структурированы, чем исполняемые: код и данные в значительно большей степени разделены, сохранена информация об именах подпрограмм и глобальных переменных (как определяемых, так и используемых), при этом объектные файлы содержат такой же машинный код, что и соответствующие исполняемые файлы. Таким образом, задача декомпиляции для объектных файлов должна решаться проще и с бóльшими шансами на успех. Однако, задача декомпиляции объектных файлов обычно не рассматривается, поскольку такие файлы в основном воспринимаются программистами, как некоторый кэш компилятора – вспомогательные данные, формируемые компилятором в ходе работы и не представляющие самостоятельной ценности.

Исключение составляет формат DCU объектных файлов Delphi. Помимо образов памяти подпрограмм и глобальных данных, в файлах `.dcu` кодируется вся информация из интерфейсной части модуля, т.е. они совмещают функции `.obj` и `.h` файлов, поэтому очень часто использующие Delphi разработчики распространяют свои модули и целые библиотеки модулей в формате DCU, без предоставления исходных текстов. При этом формат DCU частично изменяется с каждой новой версией продукта, поэтому программисты, зависящие от чужих модулей, должны полагаться на то, что разработчик такого модуля не прекратит свою работу и скомпилирует его для

следующих версий компилятора, когда это потребуется. Время показывает, что эта надежда не оправдывается очень часто. Таким образом, для мира Delphi задача декомпиляции объектных файлов DCU является чрезвычайно актуальной. При этом эта задача ещё не решена, м.б. потому, что формат DCU официально не документирован. Наиболее полно результаты исследования формата DCU представлены в коде программы DCU32INT, которая позволяет частично восстановить исходный текст модуля, но при этом в качестве кода подпрограмм вместо операторов языка Паскаль отображает ассемблерный код.

В диссертации А.А. Михайлова решена задача реализации декомпилятора для одной из разновидностей формата DCU – файлов `.dcuil`, создаваемых компиляторами тех версий Delphi, которые работали для платформы .NET. Разработан метод решения этой задачи, состоящий из ряда этапов: синтаксический анализ кода CIL; формирование графа потока управления; генерация промежуточного представления; структурирование графа потоков управления; анализ потоков данных с учётом семантики команд CIL; улучшение промежуточного представления с учётом особенностей работы компилятора Delphi; генерация кода. С использованием разработанного метода структурирования графа потоков управления также был предложен и реализован оригинальный метод визуализации управляющего графа, который позволяет более наглядно увидеть структуру подпрограммы и, в том числе, для тех видов кода, которые пока не удаётся декомпилировать.

Результат декомпиляции файлов `.dcuil` оказывается более качественным, более понятным для исследователя кода по сравнению с результатами декомпиляции исполняемых файлов .NET, поскольку в нём отображается дополнительная информация, не попадающая в исполняемые файлы, например, имена переменных. Кроме того, учёт таких особенностей компилятора, как сокращённое оценивание логических выражений, на стадии улучшения промежуточного представления позволяет получить более понятный код.

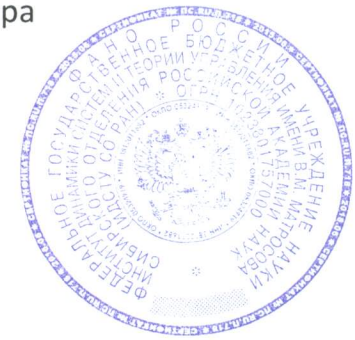
Несмотря на то, что в последних версиях Delphi платформа .NET не поддерживается, разработка метода декомпиляции для одной из разновидностей формата DCU открывает дорогу к разработке аналогичных методов для других разновидностей этого формата. Многие этапы разработанного метода и реализованные для их работы подпрограммы не зависят от особенностей кода для платформы .NET. Основную сложность для распространения метода на другие платформы представляет описание семантики машинных инструкций реальных процессоров, система команд которых сложнее байт-кода виртуальной машины, как по количеству инструкций, так и по их эффектам. Кроме того, декомпиляция файлов DCUIL может быть непосредственно востребована при исследовании унаследованного программного кода, скомпилированного для этой платформы.

Материалы диссертации опубликованы в 16 печатных работах, из них 3 статьи в рецензируемых журналах из перечня, рекомендованного ВАК, 1 статья из списка WOS, 2 авторских свидетельства. Результаты работы прошли апробацию на научных семинарах ИДСТУ СО РАН, ИСИ СО РАН, ИСП РАН, докладывались на отечественных и международных научных конференциях. Все выносимые на защиту результаты получены автором лично.

В процессе работы над диссертацией А.А. Михайлов глубоко изучил алгоритмы и структуры данных, применяемые для анализа программного кода, проанализировал большой объём литературы по теме исследований, научился самостоятельно решать возникающие перед ним научные задачи. Результаты диссертации свидетельствуют о высоком профессиональном уровне её автора.

Диссертационная работа «Методы декомпиляции объектного кода Delphi» выполнена на высоком профессиональном уровне, является законченной научно-квалификационной работой, полностью соответствующей всем требованиям, предъявляемым ВАК к диссертациям на соискание учёной степени кандидата технических наук по специальности 05.13.11. В диссертации получены новые результаты в области разработки декомпиляторов и исследования машинного кода. Автореферат полностью отражает содержание диссертационной работы. Основные результаты диссертации опубликованы согласно требованиям ВАК. Считаю, что А.А. Михайлов заслуживает присуждения ему учёной степени кандидата технических наук по специальности 05.13.11 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Научный руководитель,
первый заместитель директора
по информатизации
ИДСТУ СО РАН,
кандидат технических наук,
доцент



А.Е. Хмельнов

Подпись заверяю
Нач. отдела делопроизводства
и организационного обеспечения
ИДСТУ СО РАН
Г.Б. Кононенко
06.09.2017