Анализ программного кода объектных файлов Delphi с использованием спецификации семантики машинных команд

Михайлов Андрей Анатольевич
Институт динамики систем и теории управления СО РАН (Иркутск), Россия
e-mail: mikhailov@icc.ru

Процесс декомпиляции является важной, а при решении некоторых задач (таких, как анализ программного обеспечения без возможности использования исходного кода) и неотъемлемой, частью анализа. В общем случае (для произвольных файлов) эта задача является очень сложной, например, требуется разделить память программы на код и данные. В объектных файлах Delphi программа оказывается более структурированной, например выделены блоки памяти, соответствующие коду каждой процедуры и т. д. В общем виде формат файла dcu[1] выглядит следующим образом: сначала идет заголовок, в котором содержится общая информация о файле, такая как размер, время компиляции и т. д. После заголовка следуют поток теговой информации, который можно разделить на следующие группы: описания включаемых модулей и объектных файлов; импортируемых из этих модулей определений; описания определений; блок памяти, составленный из блоков кода для процедур и функций, образов констант, и т. д.; информация для редактора связей; отладочная информация. Таким образом, при работе с файлами dcu задача декомпиляции становиться более достижимой.

В программе DCU32INT[2] используется статический дизассемблер, который, например не может определить имя виртуального метода по коду его вызова. Для этого требуется проследить последовательность присвоений регистров и ячеек памяти, и по смещению в таблице виртуальных методов извлечь имя вызываемого метода. Для решения таких задач требуется использовать информацию о семантике машинных команд.

Основным результатом рассматриваемой работы является создание структур данных, предназначенных для описания семантики машинных команд и механизмов их использования в сочетании с дизассемблером DCU32INT. Предложенные структуры данных позволяют точно описать семантику наиболее важных для рассматриваемой задачи инструкций. При разборе инструкций используется результат её анализа полученный с помощью статического дизассемблера DCU32INT, позволяющий определить операцию и её аргументы. В результате при анализе кода программ в dcu появляется возможность анализа потоков данных порождаемых наиболее важными инструкциями языка Delphi. С помощью разработанного механизма описания спецификаций семантики машинных команд решаются задачи извлечения имен вызываемых виртуальных методов, распространения констант, подстановка имен используемых переменных, и т. д., которые позволяют существенно сократить временные ресурсы, требуемые на получение описания алгоритма исследуемой программы.

1. Хмельнов А. Е. Язык "FlexT"для спецификации бинарных форматов данных. Дис. канд. тех. наук. Иркутск. 2000. 118 с.

2. http://hmelnov.icc.ru/DCU/index.ru.html - инструмент DCU32INT (для разбора юнитов Delphi).