

Е. Н. Трошина, А. В. Чернов

Восстановление типов данных в задаче декомпилирования в язык C

Декомпиляция — одна из сложнейших задач обратной инженерии. Одной из подзадач декомпиляции является задача восстановления типов данных. В работе подробно рассматриваются методы восстановления типов данных языка C, как базовых, так и производных.

Создание и разработка сложных программных систем различного назначения часто ведется посредством интеграции отдельных компонентов, выполненных как собственными, так и сторонними разработчиками. Это позволяет значительно сократить стоимость и время разработки программного обеспечения. Внешние модули могут поставляться без исходного кода.

Наличие таких модулей в системе уменьшает уровень надежности разрабатываемого приложения с точки зрения информационной безопасности. В частности, сторонние модули могут содержать закладки или уязвимости, способствующие утечке информации и успешным атакам на информационную систему.

Кроме того, программные модули от внешних разработчиков могут содержать ошибки, исправление которых оказывается затруднительным. Следовательно, весь сторонний код должен подвергаться аудиту с точки зрения безопасности его внедрения и использования.

Программные компоненты, представленные в виде исполняемых файлов или на языке ассемблера, сложны для анализа специалистами в области информационной безопасности. Для более качественного и продуктивного анализа их лучше предоставлять специалистам на более высоком уровне представления, например, на языке высокого уровня, в частности на языке программирования C. Ассемблерный код и, тем более, исполняемые файлы не позволяют с приемлемыми трудозатратами оценить взаимосвязь элементов про-

граммы, а также идентифицировать в программе различные алгоритмические конструкции, в то время как наличие восстановленной программы на языке высокого уровня дает возможность преодолеть указанные выше трудности. В качестве одного из средств для повышения уровня абстракции представления программы может использоваться декомпиляция.

Декомпиляция — это процесс автоматического восстановления программы на языке высокого уровня из программы на языке низкого уровня. Под декомпилятором мы будем понимать инструментальное средство, получающее на вход программу на языке ассемблера или другое аналогичное низкоуровневое представление и выдающее на выход эквивалентную ей программу на некотором языке высокого уровня.

Также декомпиляция может использоваться для обеспечения совместимости программных приложений, а именно для анализа протоколов взаимодействия в случае, когда они описаны недостаточно полно или не описаны вообще. Декомпиляция позволяет упростить восстановление состояний и структур данных протокола взаимодействия.

В настоящее время из широко используемых компилируемых языков программирования высокого уровня распространены языки C и C++, поскольку именно они наиболее часто используются при разработке прикладного и системного программного обеспечения для платформ Windows, MacOS и Unix. Поэтому декомпиляторы с этих языков имеют наибольшую