Декомпиляция программы CGEN.COM из компилятора Hi-Tech v3.09

Введение

На одном из русских форумов по старым компьютерам обсуждались компиляторы языка C для процессора Z80. Автор под именем OrionExt выложил ссылку на листинг CGEN.COM, разобранный с помощью программы IDA. По мнению OrionExt, компилятор Hi-Tech C был написан на языке C и создан с помощью самого себя, но без использования опции оптимизации кода.

Меня заинтересовал данный вопрос и я продолжил сделанную им работу по декомпиляции этой программы. В результате все оказалось немного сложнее. Компилятор написан на Hi-Tech C и компилировался с оптимизацией. Спасибо OrionExt за проделанную работу и начальную информацию.

Файл CGEN.HUF содержит файлы полученные с помощью дизассемблирования бинарного исполняемого файла CGEN.COM (генератора кода компилятора Hi-Tech C v3.09 для CP/M). Исходный код на языке C и ассемблере адаптирован для компиляции исполняемого файла CGEN1.COM, побайтно совместимого с оригинальным файлом при компиляции кода на ассемблере.

Для русско-язычных пользователей в каталоге DOC выложен перевод руководства пользователя компилятора Hi-Tech C и описание декомпиляции на русском языке в файлах Z80DOC3rus.pdf и Readme_ru.pdf.

Компиляция программы

Файлы объединены в пакет с помощью программы enhuff. Для извлечения их в рабочую директория необходимо использовать программу dehuff:

```
dehuff x CGEN.HUF
```

после этого в рабочей директории появятся следующие файлы:

*.c	Исходные коды дизассемблированной программы на
*.asm	языках C и ассемблера;
makefile	Файл для компиляции новой исполняемой программы на языке ассемблера;
make_c	Файл для компиляции новой исполняемой программы на языке C;
lkcgen	Файлы для компоновки объектных файлов (вызываются
linkcgen	из makefile или makefile_c);
Readme_en.txt	Файл с описанием на английском (этот файл);
cgen.h	Включаемый файл с определениями переменных и функций;
cgen.sym	Файл символов оригинальной программы для отладчика, например, ZSID;
cgen.com	Оригинальный исполняемый файл из комплекта поставки;
0.txt	Простой файл для тестирования созданной программы;
STDIO.H	Измененная версия стандартного включаемого файла;
cgen_all_asm	Скрипты для копирования исходного кода на языках C и
cgen_all_c	ассемблера в один файл для более удобного просмотра.

Еще присутствуют следующие файлы:

LIBRARY.HUF	Библиотека в виде отдельных файлов.
SOURCE.HUF	Исходные файлы на языке C, отсутствующие в CGEN.HUF.

Внимание, все исходные файлы в конце строки используют символы CRLF, принимаемые CP/M.

Для компиляции и компоновки нового исполняемого файла нужно выполнить команду:

```
make
```

после ее окончания будут созданы следующие файлы:

cgen1.com	-	новый исполняемый файл;
cgen1.map	-	карта распределения памяти;
cgen1.sym	-	файл символов для отладчика ZSID;
cgen1.sym.sorted	-	файл символов отсортированный в порядке возрастания адресов;

Для проверки работы скомпилированной программы введите команду

```
cgen1 0.txt
```

и на экране отобразится сгенерированный код на ассемблере z80.

Результаты декомпиляции

Программа CGEN.COM была написана на языке C. Часть функций программы и стандартной библиотеки были изменены авторами с целью оптимизации (уменьшения размера программы, увеличения скорости ее работы) и, *естественно*, затруднения ее декомпиляции.

Эти измененные функций стандартной библиотеки находятся в файлах libc1.asm, libc2.asm, libc3.asm, libc4.asm и libc5.asm. Исходный код разбит на файлы исходя из желания получить точную копию исходного исполняемого файла.

При дизассемблировании оригинального исполняемого кода и декомпиляции его в исходный код на языке C стало ясно, что авторы вносили изменения в код на уровне ассемблера.

Для усложнения декомпиляции в часть функций был добавлен некоторый код, не влияющий на логику функции, однако, затрудняющий понимание ее работы.

С той же целью в некоторых функциях преднамеренно были оставлены переменные и код от версии этой программы для MS-DOS, не используемые в версии CP/M.

В некоторых функция для изменения ее размера в код сгенерированный компилятором были внесены правки ассемблерного кода, не изменяющие логику работы функции, но исключающие использование версии на языке C.

В оригинальном исполняемом файле в нескольких местах программы, включая функцию библиотечки, были удалены команды восстановления стека после вызова функций.

Изменено расположение текстовых констант, используемых для информационных сообщений программы генерации кода.

Использована довольно странная (и сложная для понимания) реализация распределения динамически выделяемой памяти при построении таблицы символов.

Для исправления внесенных багов в разных местах добавлен код исправляющий их действие. Причем не явно, а через обращение к массиву (Пока не понял как это работает).

В целом при создании программы была использована довольно сложная схема защиты от декомпиляции. В результате исходный код ассемблера пока не совсем перемещаемый.

Три восстановленные функции `1F4B.c`, `2D09.c` и `54B6.c` оказались большими для оптимизатора и для компоновки используются соответствующие файлы на ассемблере.

При компиляции исходных файлов на языке С выдается несколько предупреждающих сообщений, связанных с недостаточной проработкой структур при сохранении в них значений переменных. Они в виде комментариев включены в исходные файлы.

В исходные коды на языке ассемблера в качестве комментариев добавлен код генерируемый компилятором С, а также помечены практически все отличия.

Дополнительные возможности

Команда

```
make make_c
```

скомпилирует и создаст исполняемый файл `cgen1.com` из исходных кодов на С, который пока работает некорректно, вернее, не работает совсем.

Выполнение команды

```
make clear
```

Удалит из рабочей директории все созданные объектные и исполняемые файлы, а команда

```
make compress
```

создаст файл пакета, включив в него все необходимые файлы (если у вас есть программа `enhuff`).

Содержимое файлов с расширением `.HUF` по сути является резервной копией используемых файлов.

Зачем все это

Не коммерческая цель этой кропотливой работы – популяризация среди потенциальных поклонников 8-битных компьютеров старого компилятора Hi-Tech C V3.09 (HI-TECH Software) и продление срока его службы за пределами среды CP/M (Digital Research, Inc.), для полноценной работы в Unix-подобной операционной системе UZI-180 без использования ее эмулятора CP/M.

Решение проблемы состоит в том, чтобы воссоздать перемещаемый объектный код, заменить системные функции CP/M (ввод-вывод, выделение памяти и т.д.) аналогичными вызовами UZI-180 и скомпилировать исполняемый файл для этой операционной системы. В последующем воссоздать весь пакет этого замечательного компилятора.

Авторские права

Компилятор HI-TECH C V3.09 предоставляется бесплатно для любого использования, частного или коммерческого, строго как есть. Никакая гарантия или поддержка продукта не предлагается и не подразумевается, включая коммерческую ценность, пригодность для определенной цели или ненарушение прав. Ни при каких обстоятельствах HI-TECH Software или ее корпоративные филиалы не несут ответственности за любой прямой или косвенный ущерб.

Вы можете использовать это программное обеспечение для чего угодно, при условии, что вы **ПРИЗНАЕТЕ**, что авторские права на это программное обеспечение остаются за HI-TECH Software и ее правопреемниками.

Все авторские права на используемые алгоритмы, двоичный код, торговые марки и т.д. принадлежат законному владельцу - Microchip Technology Inc. и ее дочерним компаниям. Коммерческое использование и распространение воссозданных исходных кодов без разрешения правообладателя строго **ЗАПРЕЩЕНО**.

Планы на будущее

- создать полностью перемещаемый исходный код программы `CGEN.COM`;
- проделать аналогичную работу над остальными программами;
- написать инструкцию по использованию компилятора Hi-Tech C V3.09, с точки зрения генерации компактного и оптимального кода, основанную на опыте воссоздания данной программы.

Признательность

- HI-TECH Software за написание компилятора и предоставление его в свободное пользование.
- OrionExt за начальную разборку программы `CGEN.COM`
- Всем авторам, неравнодушным к CP/M и написавшим замечательные эмуляторы: `cpm` (Keiji Murakami), `iz-cpm` (Iván Izaguirre), `zxcc` (John Elliott), `aliados` (Julián Albo), `cpm for osx` (Thomas Harte), `tnylpo` (Georg Brein), и др.),
- Tony Nicholson за сохранение информации об этом компиляторе.
- Автору простого эмулятора x86 и DOS для терминала Linux (`emu2`), позволяющего запускать DOS версию Hi-Tech C compiler V4.11 из `makefile` под Linux или OS X.

Андрей Никитин (nikitinprior@gmail.com)