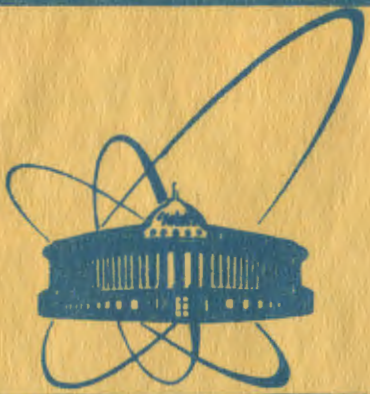


сообщения
объединенного
института
ядерных
исследований
дубна

P11-81-49

Е.Ю.Мазепа, В.А.Саенко

ДЕТРАНСЛЯТОР С ЯЗЫКА ЗАГРУЗКИ НА ЯЗЫК ASS2 ДЛЯ ЭВМ ЕС-1010

1981

Как правило, готовый программный продукт поступает к пользователю в виде двоичного модуля загрузки. Тем не менее зачастую возникает необходимость внесения изменений, дополнений и т.п. Применяемые в таких случаях довольно известные программистам методы "перехватов", "заплат" и т.д. не всегда эффективны.

Известны также многочисленные случаи гибели файлов с исходными текстами либо по вине системы, обслуживающей эти файлы, либо по вине самого программиста.

В связи с этим возникает проблема восстановления исходного текста программы с языка загрузки /2/, т.е. проблема обратной трансляции (детрансляции).

По всей видимости, восстановление исходного текста, написанного на языке высокого уровня, представляется маловозможным. Детрансляция же из двоичного представления программы на язык типа ассемблера возможна, если не всегда полностью, то, по крайней мере, в значительно более удобочитаемую форму.

Детранслятор на ЭВМ ЕС-1010

Двоичный модуль загрузки на ЭВМ ЕС-1010, хранящийся на диске ЕС-5060 /1/ (мини-диске), назовем файлом формата RMI /1/.

Файлы формата RMI бывают двух типов:

а) без ветвей перекрытий (без оверлейной структуры);

б) с ветвями перекрытий (оверлейная структура).

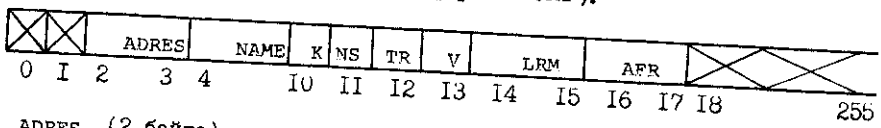
Описываемая версия детранслятора может работать только с файлами типа а .

Для ЭВМ ЕС-1010, где поля команд и данных разделены, удалось написать детранслятор, который воспроизводит текст на языке ASS2 , пригодный для повторной трансляции, причем результат повторной трансляции совпадает с исходным модулем загрузки.

Входными данными для детранслятора на ЭВМ ЕС-1010 является файл формата RMI, расположенный в зоне GIGO /I/ мини-диска.

Опишем структуру файла формата RMI.

0 - сектор (заголовок файла формата RMI):



ADRES (2 байта) - адрес первого свободного сектора, не занятого файлом формата RMI;

NAME (6 байтов) - имя файла формата RMI (код EBCDIC);

K (I байт) - кол-во ветвей перекрытий (когда их нет, K=0);

NS (I байт) - кол-во секторов, занятых файлом;

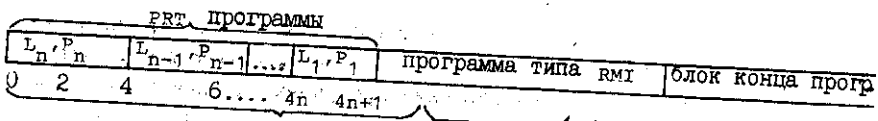
TR (I байт) - если TR=1, то файл формата RMI;

V (I байт) - номер варианта файла с данным именем;

LRM (2 байта) - длина файла формата RMI в байтах;

AFR (2 байта) - адрес сектора, начиная с которого располагается программа типа RMI (в случае зоны GIGO AFR=1).

Последующие секторы содержат PRT /I/ (Program relocation table) программы, саму программу типа RMI, а также блок конца программы и таблицу перекрытий:



где сектор № I сектор № 2 сектор № m

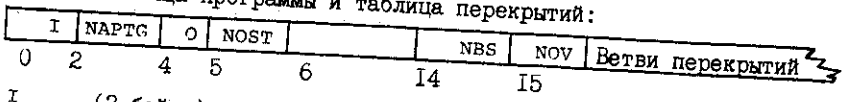
m - кол-во секторов;

n - кол-во программных секций;

L_i (2 байта) - адрес начала локального сегмента данных i -й программной секции относительно начала программы;

P_i (2 байта) - адрес начала программного сегмента i -й программной секции относительно начала программы.

Блок конца программы и таблица перекрытий:



I (2 байта) - если I3 бит = I, то программа будет выполняться в привилегированном режиме;

NAPTG (2 байта) - кол-во адресов PTG (имеет значение, если файл формата RMI - с оверлейной структурой);

NOSI (I байт) – порядковый номер программной секции, на которую передается управление в начале работы программы;

NBS (I байт) – количество секций;

NOV (I байт) – количество ветвей перекрытий.

Логическая схема работы детранслятора

Детранслятор работает по двухпроходной схеме, т.е. файл формата RMI, расположенный в зоне GIGO мини-диска, просматривается детранслятором два раза.

Первый проход. Цель первого прохода состоит в выявлении мест в исходном тексте, в которых должны быть расположены метки. Для этого образуется специальная битовая шкала, число битов которой равно числу байтов в программе типа RMI. Единица в бите этой шкалы означает, что по соответствующему адресу в программе типа RMI должна быть метка. Так как программы на ЕС-1010 имеют строго раздельные сегменты (CDS, LDS, LPS), которые описывают PRT программы, то дальнейшее определение типа метки не вызывает затруднений. Блок-схема первого прохода приведена на рис. I.

Обозначения, принятые в блок-схеме:

XPRT – указатель PRT, первоначально установлен на предпоследнем слове PRT;

AB(XPRT) – элемент таблицы PRT, соответствующий указателю XPRT;

<прав.часть ком.> – байт поля операнда в программе RMI;

<длина программы RMI> – длина файла RMI в байтах, без длины PRT и длины блока конца программы;

BITS(J) – J-й бит битовой шкалы.

Второй проход. На втором проходе производится собственно детрансляция с использованием битовой шкалы, полученной на первом переходе.

Результатами второго прохода могут быть (если это указано):

I) распечатка исходного текста (по виду сходна с листингом трансляций);

2) исходный текст на перфоленте или магнитной ленте либо файл с исходным текстом в зоне DA мини-диска, организованный по правилам стандартной для ЕС-1010 файловой системы FMS-M.

Блок-схема второго прохода приведена на рис.2.

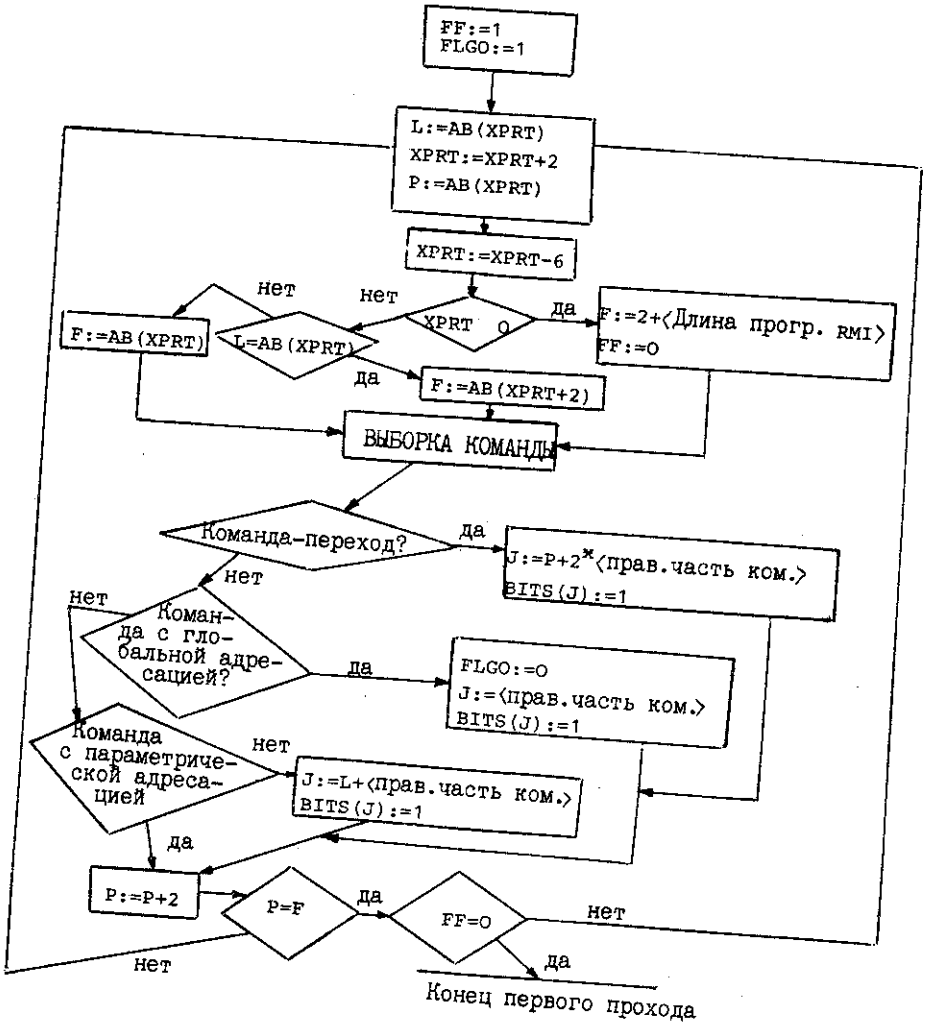


Рис.I. Схема первого прохода детрансляции.

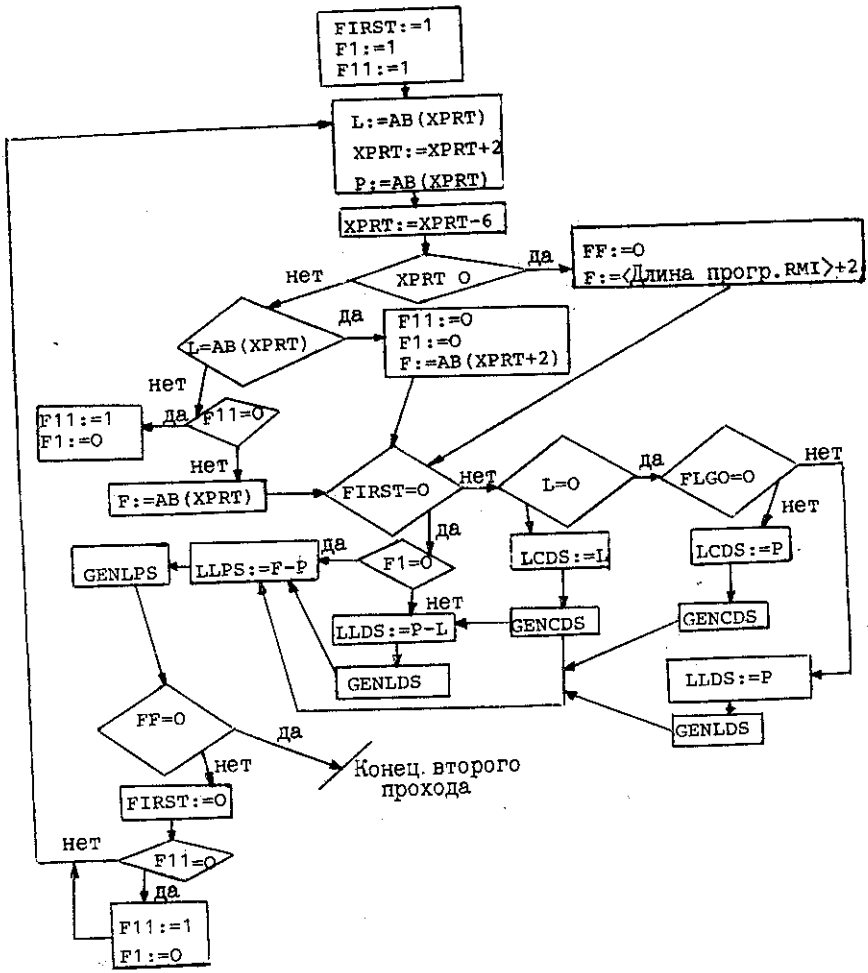


Рис.2. Схема второго прохода детрансляции.

Пояснения к блок-схеме. Обозначения к рис. 2 такие же, как и к рис.I. Кроме этого:

GENCDS - подпрограмма, генерирующая CDS;
GENLDS - подпрограмма генерации конкретного LDS;
GENLPS - подпрограмма генерации конкретного LPS;
LLDS - длина генерируемого сегмента LDS;
LCDS - длина генерируемого сегмента CDS;
LLPS - длина генерируемого сегмента LPS.

В настоящий момент среди пользователей ЭВМ ЕС-1010 распространена программа ANTIAS (разработка фирмы VIDEOTON), производящая обратную трансляцию из загруженной в память машины программы на язык ASS2 . Эта программа имеет сходные характеристики с описываемым детранслятором. Однако:

I) нет описания к этой программе;
2) некоторые программные конструкции детранслируются неверно;
3) исходный текст, получаемый в результате детрансляции, можно выдать только на перфоленту, причем каждая инструкция исходного текста занимает 80 байтов;
4) учитывая I,2,3, практически невозможно развитие этой программы в условиях ОИЯИ.

С другой стороны, в программе ANTIAS есть некоторые сервисные возможности, которые пока отсутствуют в описываемом детрансляторе. Предпринята, например, попытка комментировать сегменты данных. Авторы надеются, что эти программы будут дополнять друг друга.

В заключение хотелось бы выразить искреннюю благодарность Е.Д.Федюнькину за полезные советы и обсуждения.

Литература

I. Руководство по ЕС-1010 . VIDEOTON, 201.095.11.02-SW.
2. Ломидзе О.Н. ОИЯИ, 11-7356, Дубна, 1973.

Рукопись поступила в издательский отдел
23 января 1981 года.