УДК 004.2

Г. В. Копытов

АНАЛИЗ ЧАСТОТЫ ИСПОЛЬЗОВАНИЯ КОМАНД И МЕТОДОВ АДРЕСАЦИИ В ПРОЦЕССОРАХ ИНТЕЛ

Исследуется частота использования различных команд и методов адресации в программном обеспечении процессоров Интел.

The frequency of usage of different instructions and address modes in software for Intel architecture is examined.

Ключевые слова: микропроцессор, архитектура ЭВМ, способы адресации.

Key words: microprocessor, CPU architecture, addressing modes.

Введение

Рассмотрим вопрос о «нужности» или «ненужности» различных расширений архитектуры Интел 8086, а также вопрос относительно частоты использования различных команд и методов адресации. Для

Вестник Балтийского федерального университета им. И. Канта. 2011. Вып. 10. С. 168 – 171.

анализа был взят программный код, поставляемый с операционной системой Red Hat Enterprise Linux 4.4, которая широко используется в коммерческих и учебных заведениях. Код самой операционной системы не анализировался.

Анализ программного кода производился статически по кодовым сегментам программ. Динамический анализ статистики использования различных команд в процессе выполнения программы не является целью данной работы.

1. Структура команды в архитектуре Интел

Все команды процессора Интел имеют формат, представленный на рисунке. TTT — код операции (1 или 2 байта), modTTTr/m специфицирует способ адресации операнда, ss index base, называемый также s-i-b-байтом, был введен в процессоре 80386, он обусловливает способ адресации, связанный с масштабированием индексного регистра. Масштаб определяется следующим образом: если ss = 00, то значение индексного регистра index умножается на 1, если ss = 01 — умножается на 2, если ss = 10 — на 4, если ss = 11 — на 8. Байт s-i-b является необязательным и его присутствие в команде зависит от значения предыдущего байта — modTTTr/m. Вслед за s-i-b-байтом может идти смещение, которое имеет либо 32 разряда (d32), либо 16 (d16), либо 8 (d8), либо оно вообще отсутствует.

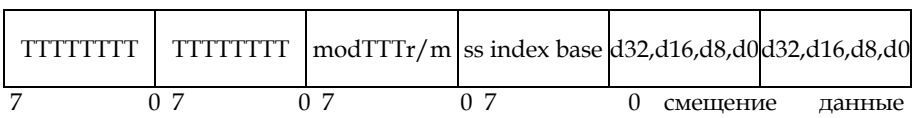

Рис. Формат команд процессора Интел

Разрядность смещения (и его присутствие) определяется кодом операции и режимом работы процессора — 16-разрядный (8086, 80186 или 80286) или 32-разрядный (80386+). Непосредственные данные также могут присутствовать в команде, и их разрядность вычисляется аналогично разрядности смещения. Команде может предшествовать один или несколько однобайтовых префиксов, которые влияют на выполнение команды. В частности, существует префикс смены разрядности данных, который меняет разрядность непосредственных данных с 16 на 32 или наоборот. Есть также префикс смены разрядности адреса, но, как мы увидим далее, он никогда не используется.

2. Реализация и сбор статистики

Был написан декодировщик команд процессора Интел, который охватывает архитектуры с 8086 по Pentium+, включая расширения SSE. Все команды были сгруппированы в соответствии с классификацией Интел. Статистика собиралась по следующим группам: команды пересылки данных, включая стековые операции, арифметические команды,

команды переходов, вызов и возврат из процедур, логические команды, команды управления процессором. Отдельная статистика велась по редко используемым командам, таким, как арифметика упакованных чисел, строковые и побитные операции. Привилегированные операции также были вынесены в отдельную строку, но особого интереса не представляют, так как анализировался только прикладной код, в котором они не могли быть использованы. Также отдельно рассматривалась статистика по командам SETxx, введенных в архитектуре 80386. Статистика по методам адресации оформлялась в виде таблицы, индексами в которой служили поля mod и r/m, которые в совокупности определяют все возможные методы адресации. Отдельно учитывалась статистика по адресации с масштабированием (поле s-i-b). Эти методы характерны для процессоров 80386 и выше.

3. Результаты

Обработанная статистика показывает, что наиболее распространены группы команд пересылки данных (56 %) и команд переходов (13 %). При этом доля команд перехода, использующих полный 32-разрядный адрес, составила более 30 % от общего числа команд переходов. Следующей по численности стала группа логических команд AND, OR, XOR, NOT и TEST. Они составляют 10 % кода.

Команды целочисленной арифметики немного отстают — 9 % кода. Команды арифметики с плавающей точностью тоже присутствуют, но совсем незначительно — около 0,1 %. Это объясняется тем, что анализируемое программное обеспечение имеет общее назначение. В коде вычислительных программ доля такой арифметики значительно выше.

Следующая сравнительно большая группа команд — вызовы и возвраты из подпрограмм — 7 %. Это вполне объяснимо, однако вызывает удивление факт, что команды Enter и Leave, которые появились в процессоре 80186, используются достаточно редко — примерно в 0,5 % случаев, хотя они были введены специально для поддержки языков высокого уровня, но почему-то не получили признания у разработчиков компиляторов.

Доля остальных групп команд незначительна. Выделим команды обработки упакованного десятичного формата, которые не встретились ни разу в исследуемом коде. Это, очевидно, рудимент архитектуры Интел. Отметим еще несколько неиспользуемых возможностей. Префикс подмены сегмента немного встречается в коде, а вот префикс подмены сегментов FS и GS, введенных в 80386, не встречается никогда. Поскольку эти сегменты также никакими командами не применяются по умолчанию, то делаем вывод, что введение дополнительных сегментных регистров FS и GS в архитектуру 80386 было неоправданным.

Отдельная статистика собиралась по командам, расширяющим архитектуру 8086. Несмотря на нелюбовь компиляторщиков к командам ENTER и LEAVE, другие команды архитектуры 80186, PUSHA и POPA, то есть сохранение/восстановление в стеке всех регистров, оказались популярными, на них приходится около 2 % всех команд. Расширения

архитектуры 80386, а к ним относятся команды манипулирования битами BITxx, условные установки байта SETxx, а также групповая пересылка памяти MOVS, оказались оправданными и встречаются почти в 6 % кода. А вот новые команды архитектуры 80486, BSWAP и CMPXCHG почти невостребованными — сотые доли процента. То же самое касается расширений SSE, введенных в Pentium III, их использование ограничивается тысячными долями процентов.

Среди способов адресации операндов самым популярным является адресация по базовому регистру или базово-индексная с коротким смещением (0, 8 или 16 бит) — 57 %, а среди нее — адресация по базовому регистру BP. Это неудивительно, так как именно этот регистр используется для адресации локальных переменных в функциях, а также для доступа к параметрам. Вторая по численности — регистровая адресация (без обращения к памяти) — 30 %. Адресация с длинным 32-разрядным смещением используется в 13 % случаев. Прямая адресация памяти (по адресу) встречается редко — чуть более 1 %, что согласуется с общепризнанной практикой программирования избегать глобально-разделяемых данных. Интересно, что индексная и базово-индексная адресация с масштабированием индекса, введенная в процессоре 80386, оказалась популярной — около 20 % случаев. Этот способ хорошо дополнил адресацию базовой архитектуры 8086, сделав ее логически завершенной.

Список литературы

1. *Юров В. И.* Assembler: учебник для вузов. СПб., 2003.
2. *Гук М.* Процессоры Intel: от 8086 до Pentium II. СПб., 1998.

Об авторе

Герман Васильевич Копытов — канд. физ.-мат. наук, доц., Балтийский федеральный университет им. И. Канта, e-mail: gkopytov@rogers.com.

Author

Dr German Kopytov — assistant professor, I. Kant Baltic Federal University, e-mail: gkopytov@rogers.com.