Машинно-зависимые языки программирования, лекция 5

Каф. ИУ7 МГТУ им. Н. Э. Баумана, 2024 г.

32-разрядные процессоры (386+)

Производство x86: 1985 - ~2010
32-разрядные:

- Регистры, кроме сегментных
- Шина данных
- Шина адреса (2^{32} = 4Гб ОЗУ)

Режимы работы

8086 (1978 г.) -> 80186 (1982 г.)
-> 80286 (1982 г.) добавлен защищённый режим
-> 80386 (1985 г.) архитектура стала 32-разрядной
-> 80486 (1989 г.) -> Pentium -> ... -> (современные процессоры)

"Реальный" режим (режим совместимости с 8086)

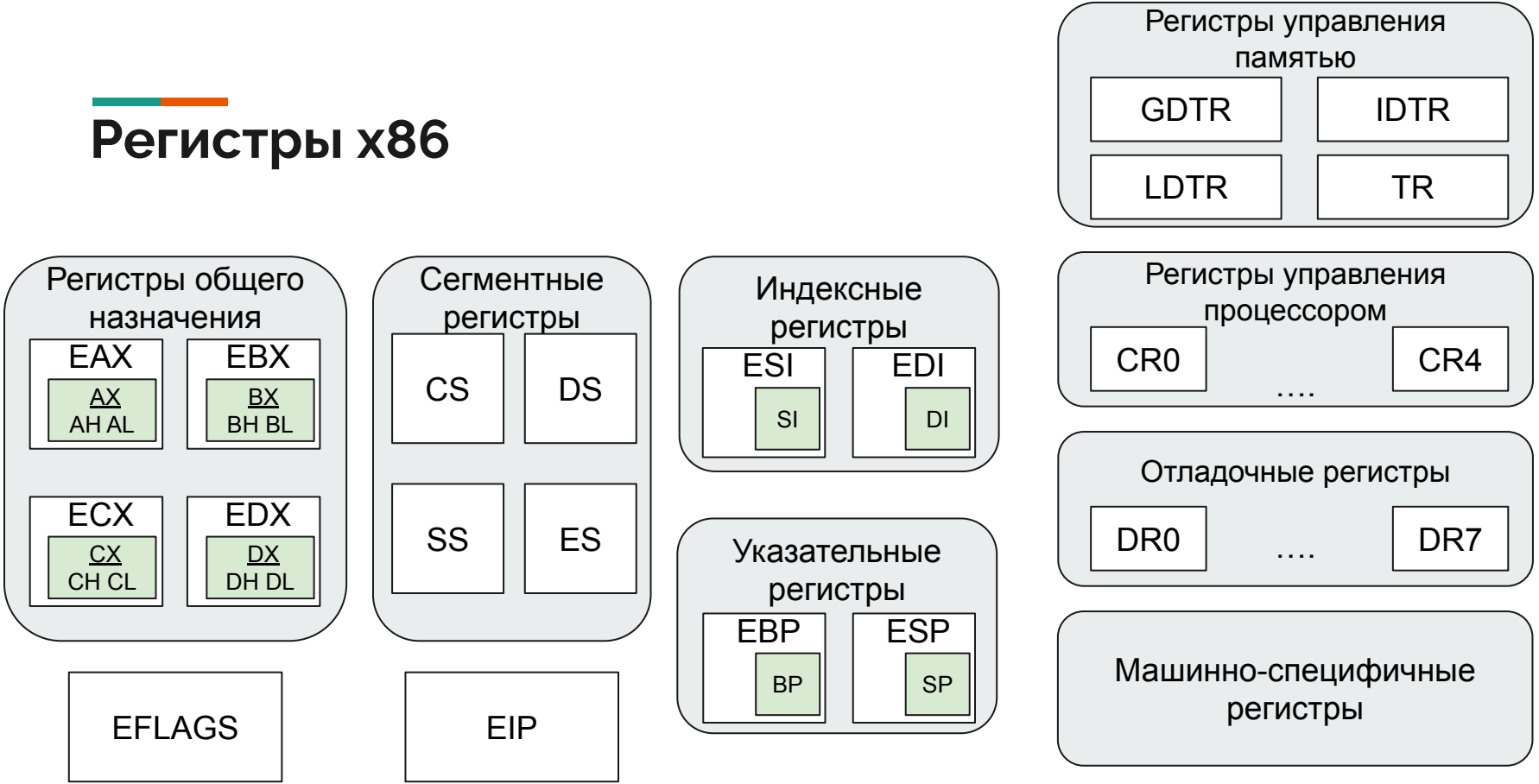
- обращение к оперативной памяти происходит по реальным (действительным) адресам, трансляция адресов не используется;
- набор доступных операций не ограничен;
- защита памяти не используется.

"Защищённый" режим

- обращение к памяти происходит по виртуальным адресам с использованием механизмов защиты памяти;
- набор доступных операций определяется уровнем привилегий (кольца защиты): системный и пользовательский уровни

Режим V86, ...

Регистры x86



Система команд

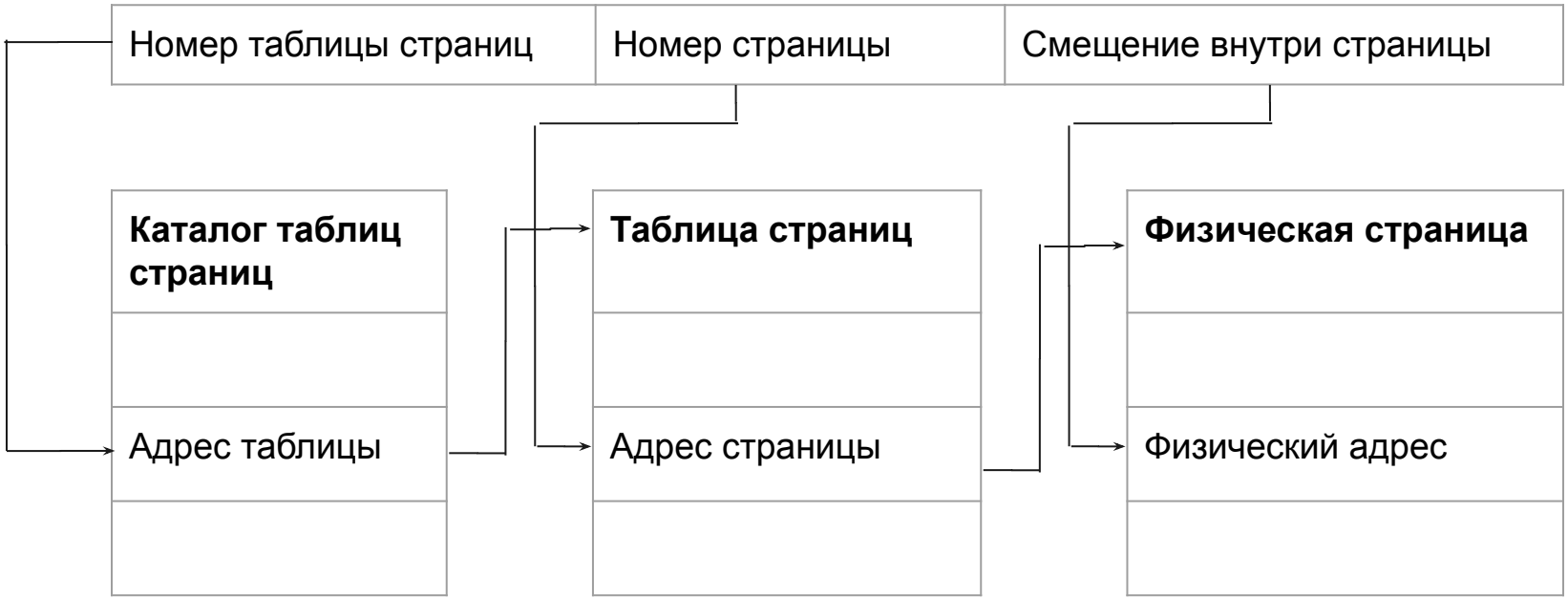
- Аналогична системе команд 16-разрядных процессоров
- Доступны как прежние команды обработки 8- и 16-разрядных аргументов, так и 32-разрядных регистров и переменных
- Пример:

```
mov eax, 12345678h
xor ebx, ebx
mov bx, 1
add eax, ebx    ; eax=12345679h
```

Модели памяти

- Плоская - код и данные используют одно и то же пространство
- Сегментная - сложение сегмента и смещения
- Страничная - *виртуальные* адреса отображаются на физические постранично
  - виртуальная память — метод управления памятью компьютера, позволяющий выполнять программы, требующие больше оперативной памяти, чем имеется в компьютере, путём автоматического перемещения частей программы между основной памятью и вторичным хранилищем (файл, или раздел подкачки)
  - основной режим для большинства современных ОС
  - в x86 минимальный размер страницы - 4096 байт
  - основывается на таблице страниц - структуре данных, используемой системой виртуальной памяти в операционной системе компьютера для хранения сопоставления между виртуальным адресом и физическим адресом. Виртуальные адреса используются выполняющимся процессом, в то время как физические адреса используются аппаратным обеспечением. Таблица страниц является ключевым компонентом преобразования виртуальных адресов, который необходим для доступа к данным в памяти.

Страничная организация памяти



Управление памятью в x86

- В сегментных регистрах - селекторы
    - 13-разрядный номер дескриптора
    - какую таблицу использовать - глобальную или локальную
    - уровень привилегий запроса 0-3
- По селектору определяется запись в одной из таблиц дескрипторов сегментов
- При включённом страничном режиме - по таблице страниц определяется физический адрес страницы либо выявляется, что она выгружена из памяти, срабатывает исключение и операционная система подгружает затребованную страницу из "подкачки" (swap)

Поддержка многозадачности

TSS (Task State Segment — сегмент состояния задачи) — специальная структура в архитектуре x86, содержащая информацию о задаче (процессе). Используется ОС для диспетчеризации задач, в т. ч. переключения на стек ядра при обработке прерываний и исключений

Исключения

- **Исключения** (Exceptions) подразделяются на отказы, ловушки и аварийные завершения.
- **Отказ** (fault) — это исключение, которое обнаруживается и обслуживается до выполнения инструкции, вызывающей ошибку. После обслуживания этого исключения управление возвращается снова на ту же инструкцию (включая все префиксы), которая вызвала отказ. Отказы, использующиеся в системе виртуальной памяти, позволяют, например, подкачать с диска в оперативную память затребованную страницу или сегмент.
- **Ловушка** (trap) — это исключение, которое обнаруживается и обслуживается после выполнения инструкции, его вызывающей. После обслуживания этого исключения управление возвращается на инструкцию, следующую за вызвавшей ловушку. К классу ловушек относятся и программные прерывания.
- **Аварийное завершение** (abort) — это исключение, которое не позволяет точно установить инструкцию, его вызвавшую. Оно используется для сообщения о серьезной ошибке, такой как аппаратная ошибка или повреждение системных таблиц.

Регистр EFLAGS

FLAGS + 5 специфических флагов

Регистры управления памятью

- GDTR: 6-байтный регистр, содержит 32-битный линейный адрес начала таблицы глобальных дескрипторов (GDT) и 16-битный размер (лимит, уменьшенный на 1)
- IDTR: 6-байтный регистр, содержит 32-битный линейный адрес начала таблицы глобальных дескрипторов обработчиков прерываний (IDT) и 16-битный размер (лимит, уменьшенный на 1)
- LDTR: 10-байтный регистр, содержит 16-битный селектор для GDT и весь 8-байтный дескриптор из GDT, описывающий текущую таблицу локальных дескрипторов
- TR: 10-байтный регистр, содержит 16-битный селектор для GDT и весь 8-байтный дескриптор из GDT, описывающий TSS текущей задачи

Регистры управления процессором

- CR0 - флаги управления системой
  - PG - включение режима страничной адресации
  - управление отдельными параметрами кеша
  - WP - запрет записи в страницы "только для чтения"
  - NE - ошибки FPU вызывают исключение, а не IRQ13
  - TS - устанавливается процессором после переключения задачи
  - PE - включение защищённого режима
- CR1 - зарезервирован
- CR2 - регистр адреса ошибки страницы - содержит линейный адрес страницы, при обращении к которой произошло исключение #PF
- CR3 - регистр основной таблицы страниц
  - 20 старших бит физического адреса начала каталога таблиц либо 27 старших бит физического адреса начала таблицы указателей на каталоги страниц, в зависимости от бита PAE в CR4
  - Управление кешированием и сквозной записью страниц
- CR4 - регистр управления новыми возможностями процессоров (с Pentium)

Отладочные регистры

- DR0..DR3 - 32-битные линейные адреса четырёх возможных точек останова по доступу к памяти
- DR4, DR5 - зарезервированы
- DR6 (DSR) - регистр состояния отладки. Содержит причину останова
- DR7 (DCR) - регистр управления отладкой. Управляет четырьмя точками останова

Машинно-специфичные регистры

- Управление кешем
- Дополнительное управление страничной адресацией
- Регистры расширений процессора: MMX и т.д.

Системные и привилегированные команды

- Выполнение ограничено, в основном, нулевым кольцом защиты
- LGDT, SGDT
- LLDT, SLDT
- LTR, STR
- LIDT, SIDT
- MOV CR0..CR4 или DR0..DR7, <источник>
- ...

Страничная адресация - преобразование линейного адреса в физический

- Линейный адрес:
  - биты 31-22 - номер таблицы страниц в каталоге
  - биты 21-12 - номер страницы в выбранной таблице
  - биты 11-0 - смещение от физического адреса начала страницы в памяти
- Каждое обращение к памяти требует двух дополнительных обращений!
- Необходим специальный кеш страниц - TLB
- Каталог таблиц/таблица страниц:
  - биты 31-12 - биты 31-12 физического адреса таблицы страниц либо самой страницы
  - атрибуты управления страницей

Механизм защиты

- Механизм защиты - ограничение доступа к сегментам или страницам в зависимости от уровня привилегий
- К типам сегментов реального режима (код, стек, данные) добавляется TSS - сегмент состояния задачи. В нём сохраняется вся информация о задаче на время приостановки выполнения. Размер - 68h байт.
- Структура:
    - селектор предыдущей задачи
    - Регистры стека 0, 1, 2 уровней привилегий
    - EIP, EFLAGS, EAX, EBX, ECX, EDX, ESP, EBP, ESI, EDI, CS, DS, ES, FS, HS, SS, LDTR
    - флаги задачи
    - битовая карта ввода-вывода (контроль доступа программы к устройствам)

64-разрядные процессоры (x86-64)
AMD - с 2001, Intel - с 2003

- Режимы работы:
  - Legacy mode - совместимость с 32-разрядными процессорами
  - Long mode – 64-разрядный режим с частичной поддержкой 32-разрядных программ. Рудименты V86 и сегментной модели памяти упразднены
- Регистры:
  - целочисленные 64-битных регистры общего назначения - RAX, RBX, RCX, RDX, RSI, RDI, RBP, RSP;
  - новые целочисленные 64-битных регистры общего назначения R8 — R15
  - 64-битный указатель RIP и 64-битный регистр флагов RFLAGS.

Виды трансляторов ассемблера

- MASM
- TASM
- NASM
- FASM
- YASM
- as
- ...

AT&T-синтаксис

Синтаксис стандартного ассемблера для UNIX - **as**

Основные отличия от Intel-синтаксиса:

1. Имена регистров предваряются префиксом %.
2. Обратный порядок операндов: вначале источник, затем приёмник.
3. Размер операнда задается суффиксом, замыкающим инструкцию.
4. Числовые константы записываются в Си-соглашении.
5. Для получения смещения метки используется префикс $.

Создание оконных приложений на ассемблере под x86

Системный вызов — обращение прикладной программы к ядру операционной системы для выполнения какой-либо операции.

Для реализации оконных приложений необходима линковка с соответствующими библиотеками и использование как их функций, так и системных вызовов.

EXE-файлы в Windows

Структура файла формата PE:

1. DOS-секция
2. PE-заголовок
    - NT header: сигнатура, указатели на основной и опциональный заголовки
    - File header: допустимая архитектура, кол-во секций, время создания, указатель на таблицу символов
    - Optional header: точка входа, секция кода, секция данных, базовый адрес, количество каталогов
3. Таблица секций
4. Секции

ELF (Executable and Linking format)

1. Заголовок 52/64 байта, начинающийся с сигнатуры 0x7f ELF:
    - класс файла, метод кодирования, версия заголовка, расширения ABI
    - тип файла
    - архитектура
    - версия формата
    - адрес точки входа
    - смещение таблиц заголовков программы и заголовков секций
    - число заголовков и секций
2. Таблица заголовков
3. Таблица заголовков секций
4. Секции и сегменты

Mach-O

1. Заголовок
    - сигнатура 0xfeedface/0xfeedfacf
    - тип процессора, подтип
    - количество и размер команд загрузки
    - флаги
2. Команды загрузки (указания, как и куда загружать блоки файла)
3. Сегменты, в каждом до 255 секций

Дизассемблирование. Реверс-инжиниринг

Дизассемблер - транслятор, преобразующий машинный код, объектный файл или библиотечные модули в текст программы на языке ассемблера.

Дизассемблирование - процесс получения текста программы на ассемблере из программы в машинных кодах.

Реверс-инжиниринг (обратная разработка) — исследование готовой программы с целью понять принцип работы, поиска недокументированных возможностей или внесения изменений.