

2.11 Кооперативная и вытесняющая многозадачность. Процессы, потоки выполнения. Скалярные, супер-скалярные, векторные, SMT процессоры. Шины доступа к памяти и NUMA.

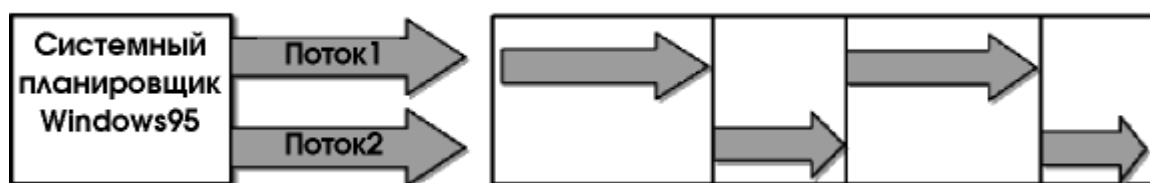
2.11.1 Кооперативная и вытесняющая многозадачность.

Многозадачность - способность операционной системы выполнять более одной программы (задачи) одновременно. Противоположный подход - однозадачность, когда один процесс должен быть завершен прежде, чем сможет начаться другой. MS-DOS - пример однозадачной среды, а Windows 95 и Windows NT - многозадачные среды.

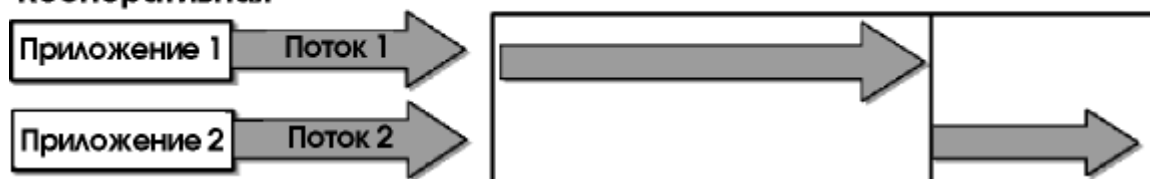
В многозадачной системе программы не выполняются одновременно - процессор переключается между ними. Что позволяет компьютеру эффективно использовать процессорное время, которое например иначе было бы потеряно в ожидании команды пользователя или ответа устройств ввода/вывода.

Существуют два типа многозадачности: **кооперативная** (не вытесняющая) и **вытесняющая** (рис. 1.4). В кооперативной многозадачной среде (например, Windows 3.1) контроль над процессором никогда не отбирается у задачи - приложение должно самостоятельно отказаться от контроля над процессором, чтобы другое приложение заработало. Вытесняющая многозадачность отличается от кооперативной тем, что операционная система может получить контроль над процессором без согласия выполняющегося приложения. Лишение приложения контроля над процессором называется вытеснением. Windows 95 и Windows NT используют вытесняющую многозадачность для MS-DOS и 32-разрядных Windows-приложений.

Вытесняющая



Кооперативная



Разработчики программ, выполняющихся под управлением кооперативной операционной системы, должны учитывать необходимость частого возврата управления процессором операционной системе. Программа, которая недостаточно часто отдает управление, блокирует кооперативную операционную систему.

2.11.2 Процессы, потоки выполнения.

Процесс - это, попросту говоря, выполняемая программа. Ему принадлежат адресное пространство и выделенные ресурсы, а также один или более потоков, выполняющихся в его контексте.

Поток - это основная единица, которой операционная система выделяет процессорное время, и минимальный "квант" кода, который может быть запланирован для выполнения. Кроме того, это часть процесса, выполняющаяся в каждый момент времени. Поток работает в адресном пространстве процесса и использует ресурсы, выделенные процессу.

Любой процесс содержит хотя бы один поток. Каждый процесс 16-разрядного Windows-приложения или программы MS-DOS имеет только один поток, тогда как процессы 32-разрядных Windows-приложений могут включать их несколько.

Примечание. Ресурсами владеют процессы, а не потоки - последние только используют ресурсы, выделенные процессу. Например, если программа запросит порт, им будет управлять процесс. Любой поток процесса может обратиться к порту, но ни один из них не вправе самостоятельно запросить использование порта.

2.11.3 Скалярные, супер-скалярные, векторные, SMT процессоры.

Классификация параллельных архитектур по Флинну (M. Flynn)

- Вычислительная система с одним потоком команд и данных (однопроцессорная ЭВМ — SISD, Single Instruction stream over a Single Data stream).
- Вычислительная система с общим потоком команд (SIMD, Single Instruction, Multiple Data — одиночный поток команд и множественный поток данных).
- Вычислительная система со множественным потоком команд и одиночным потоком данных (MISD, Multiple Instruction Single Data — конвейерная ЭВМ).
- Вычислительная система со множественным потоком команд и данных (MIMD, Multiple Instruction Multiple Data).

Скалярный процессор — это простейший класс микропроцессоров. Скалярный процессор обрабатывает один элемент данных за одну инструкцию (SISD процессор, типичными элементами данных могут быть целые или числа с плавающей запятой).

Суперскалярный процессор — процессор, способный выполнять несколько инструкций процессора за 1 такт. Процессор типа MIMD. Содержит множество операционных устройств. В соответствии с используемым принципом, команды, последовательно формируемые конвейером команд, выполняются свободными операционными блоками, причем результаты операций могут выдаваться с очередностью, отличной от очередности команд в командном конвейере.

Векторный процессор — это процессор, в котором операндами некоторых команд могут выступать упорядоченные массивы данных — векторы (SIMD процессор). Отличается от скалярных процессоров, которые могут работать только с одним операндом в единицу времени. Абсолютное большинство процессоров являются скалярными или близкими к ним. Векторные процессоры были распространены в сфере научных вычислений, где они являлись основой большинства суперкомпьютеров начиная с 1980-х до 1990-х. Но резкое увеличение производительности и активная разработка новых процессоров привели к вытеснению векторных процессоров со сферы повседневных процессоров.

Одновременная многопоточность (SMT — simultaneous multithreading) — следующий шаг в расширении возможностей процессора, ориентированный на приложения, требовательные к производительности, и поддерживающий метод распараллеливания задач на уровне инструкций на несколько каналов обработки процессора.

Одновременная многопоточная обработка на уровне приложений позволяет «видеть» два процессора вместо одного задействованного. В случае когда режим SMT отключен (однопоточная обработка), каждый физический процессор использует один логический.

Когда же режим SMT активирован, каждый процессор, исполняющий в данный момент команды приложения, использует два логических процессора вместо одного. Технология многопоточности реализуется за счет того, что при выполнении разных команд приложения используются разные устройства процессора, что позволяет распределять по ним потоки.

2.11.4 Шины доступа к памяти и NUMA.

Шины микрокомпьютера образует группа линий передачи сигналов с адресной информацией, данных, а также управляющих сигналов. Фактически ее можно разделить на три части: адресную шину, шину данных и шину управляющих сигналов (здесь рассматривается 80286 процессор).

Уровни этих сигналов в данный момент времени определяют состояние системы в этот момент.

На адресную шину, состоящую из 24 линий, микропроцессор выставляет адрес байта или слова, который будет пересылаться по шине данных в процессор или из него. Кроме того, шина адреса используется микропроцессором для указания адресов периферийных портов, с которыми производится обмен данными.

Шина данных состоит из 16 линий. по которым возможна передача как отдельных байтов. так и двухбайтовых слов. При пересылке байтов возможна передача и по старшим 8 линиям, и по младшим. Шина данных двунаправлена, так как передача байтов и слов может производиться как в микропроцессор, так и из него.

Шина управления формируется сигналами, поступающими непосредственно от микропроцессора, сигналами от шинного контроллера, а также сигналами, идущими к микропроцессору от других микросхем и периферийных адаптеров.

Микропроцессор использует шинный контроллер для формирования управляющих сигналов, определяющих перенос данных по шине. Он выставляет три сигнала -SO, -SI, M/-IO, которые определяют тип цикла шины (подтверждение прерывания, чтение порта ввода/вывода, останов, чтение памяти, запись в память). На основании значений этих сигналов шинный контроллер формирует управляющие сигналы, контролирующие динамику данного типа шины.

Для того, чтобы понять динамику работы, разберем, каким образом осуществляется процессором чтение слов из оперативной памяти. Это происходит в течение 4 тактов CLK, или 2 состояний процессора (т.е. каждое состояние процессора длится 2 такта синхросигнала CLK). Во время первого состояния, обозначаемого, как T 4s 0, процессор выставляет на адресную шину значение адреса, по которому будет читаться слово. Кроме того, он формирует на шине совместно с шинным контроллером соответствующие значения управляющих сигналов. Эти сигналы и адрес обрабатываются схемой управления памятью, в результате чего, начиная с середины второго состояния процессора T 4s 0 (т.е. в начале четвертого такта CLK), на шине данных появляется значение содержимого соответствующего слова из оперативной памяти. И наконец, процессор считывает значение этого слова с шины данных. На этом перенос (копирование) значения слова из памяти в процессор заканчивается.

Таким образом, если частота кварцевого генератора, определяющая частоту CLK, равна 20 МГц, то максимальная пропускная способность шины данных равна (20/4) миллионов слов в секунду, или 10 В/сек. Реальная пропускная способность существенно ниже.

NUMA - Non Uniform Memory Access. Применяется в многопроцессорных системах. Суть в том, что для каждого процессора различается три типа памяти: память этого процессора, общая память процессоров и память других процессоров. Соответственно для этих памятей время доступа различается.