

G

# Операционные системы



A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.

# Процессы и потоки

A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.

# Мультипрограммирование

# Мультипрограммирование

Мультипрограммирование, или многозадачность (multitasking) - это способ организации вычислительного процесса, при котором на одном процессоре попеременно выполняются сразу несколько программ.

# Мультипрограммирование

Эти программы совместно используют процессор и другие ресурсы компьютера: оперативную и внешнюю память, устройства ввода-вывода.

# Мультипрограммирование

Мультипрограммирование повышает эффективность использования вычислительных систем.

# Мультипрограммирование

Главными критериями эффективности вычислительных систем являются:

1. пропускная способность - количество задач, выполняемых вычислительной системой в единицу времени;

# Мультипрограммирование

2. удобство работы пользователей:  
например, пользователи имеют возможность интерактивно работать одновременно с несколькими приложениями на одной машине;



# Мультипрограммирование

3. реактивность системы - способность системы выдерживать заранее заданные интервалы времени между запуском программы и получением результата.

# Мультипрограммирование

По способу организации мультипрограммирования ОС делятся на системы:

- пакетной обработки
- разделения времени
- реального времени



# Системы пакетной обработки

# Системы пакетной обработки

Системы пакетной обработки предназначены для решения задач в основном вычислительного характера, не требующих быстрого получения результатов.

# Системы пакетной обработки

Главной целью и критерием эффективности систем пакетной обработки является максимальная пропускная способность, то есть решение максимального числа задач в единицу времени.

# Системы пакетной обработки

При работе ИС задачи часто приостанавливаются по внутренним причинам, связанным, например, с ожиданием ввода данных для обработки.

# Системы пакетной обработки

Данные могут храниться на диске, поступать от человека, работающего за терминалом или от измерительной аппаратуры.

# Системы пакетной обработки

При возникновении таких блокировок выполняемой задачи применяется переключение процессора на выполнение другой задачи, у которой есть данные для обработки.



# Системы пакетной обработки

Эта концепция мультипрограммирования положена в основу пакетных систем.

# Системы пакетной обработки

Для этого в системах пакетной обработки формируется пакет заданий.

Каждое задание содержит требование к системным ресурсам, из этого пакета формируется *мультипрограммная смесь*, то есть множество «одновременно» выполняемых задач.

# Системы пакетной обработки

Для «одновременного» выполнения выбираются задачи, предъявляющие разные требования к ресурсам, так, чтобы обеспечивалась сбалансированная загрузка всех устройств вычислительной машины.

# Системы пакетной обработки

Например, в мультипрограммной смеси желательно одновременное присутствие вычислительных задач и задач с интенсивным вводом-выводом.

# Системы пакетной обработки

Таким образом, выбор нового задания из пакета заданий зависит от внутренней ситуации, складывающейся в системе, то есть выбирается «возможное выгодное» задание.

# Системы пакетной обработки

Следовательно, в вычислительных системах, работающих под управлением пакетных ОС, невозможно гарантировать выполнение того или иного задания в течение определенного периода времени.

# Системы пакетной обработки

Совмещение операций ввода-вывода и вычислений может достигаться разными способами.



# Организация операций ввода-вывода в вычислительных системах



# Системы пакетной обработки

1. Специализированный процессор ввода-вывода.  
В компьютерах класса мэйнфреймов такие процессоры называют каналами.

# Системы пакетной обработки

Канал имеет систему команд, отличающуюся от системы команд центрального процессора.

# Системы пакетной обработки

Команды канала специально предназначены для управления внешними устройствами, например «проверить состояние устройства», «установить магнитную головку», «установить начало листа», «напечатать строку».

# Системы пакетной обработки

Канальные программы могут храниться в той же оперативной памяти, что и программы ЦП.

# Системы пакетной обработки

Центральный процессор и канал работают параллельно.

# Системы пакетной обработки

2. Компьютеры с управлением контроллерами внешних устройств.

# Системы пакетной обработки

Каждое внешнее устройство (или группа внешних устройств одного типа) имеет свой собственный контроллер, который автономно обрабатывает команды, поступающие от ЦП.

# Системы пакетной обработки

При этом контроллер и центральный процессор работают асинхронно, поскольку многие внешние устройства включают электромеханические узлы, контроллер выполняет свои команды управления устройствами существенно медленнее, чем ЦП.



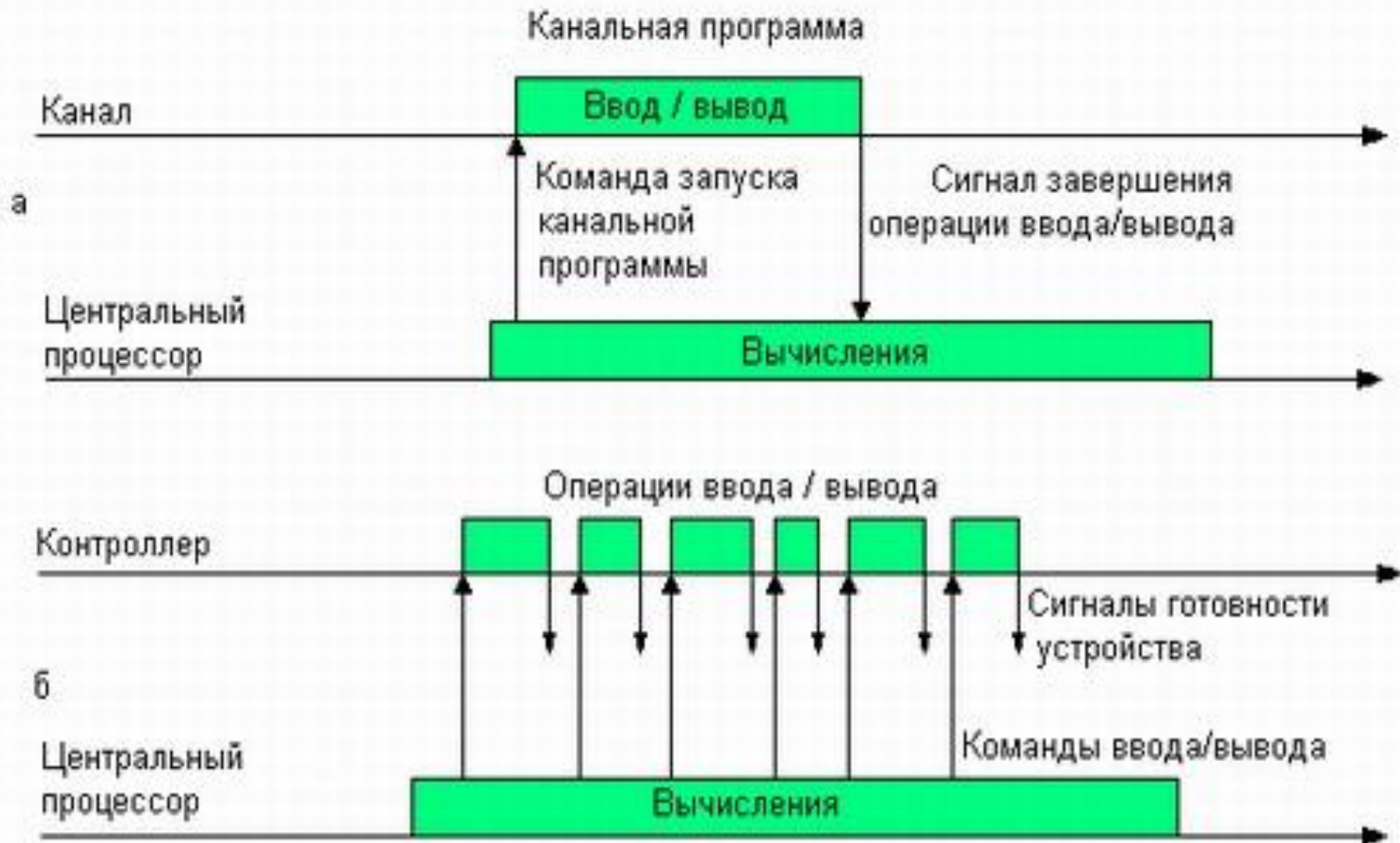
# Системы пакетной обработки

Это используется для организации параллельного выполнения вычислений и операций ввода-вывода: в промежутке между передачей команд контроллеру ЦП может выполняться вычисления.

# Системы пакетной обработки

Контроллер может «сообщить» ЦП о том, что он готов принять следующую команду, сигналом прерывания либо процессор узнает об этом, периодически опрашивая состояние контроллера.

# Системы пакетной обработки



# Системы пакетной обработки

Пока одна задача ожидает какого-либо события (завершение ввода-вывода, наступление определенного момента времени, разблокирование файла или загрузка с диска недостающей страницы программы), процессор выполняет другую задачу.

A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.

# Системы разделения времени

# Системы разделения времени

В системах разделения времени пользователям (или одному пользователю) предоставляется возможность интерактивной работы сразу с несколькими приложениями.

# Системы разделения времени

Каждое приложение должно регулярно получать возможность «общения» с пользователем.

Для этого ОС принудительно периодически приостанавливает приложения, не дожидаясь, когда они добровольно освободят процессор.

# Системы разделения времени

Всем приложениям попеременно выделяется квант процессорного времени, таким образом пользователи, запустившие программы на выполнение, получают возможность поддерживать с ними диалог.



# Системы разделения времени

Так как в системах разделения времени каждой задаче выделяется только квант процессорного времени, задачи не занимают процессор надолго.

# Системы разделения времени

Поскольку квант выбирается небольшим, то у всех пользователей, одновременно работающих на одной и той же машине, складывается впечатление, что каждый из них единолично использует машину.

# Системы разделения времени

Системы разделения времени обладают меньшей пропускной способностью, чем системы пакетной обработки, так как на выполнение принимается каждая запущенная пользователем задача, а не та, которая «выгодна» системе.

# Системы разделения времени

Кроме того, производительность системы снижается из-за более частого переключение процессора с задачи на задачу.

# Системы разделения времени

Это соответствует тому, что критерием эффективности систем разделения времени является не максимальная пропускная способность, а удобство и эффективность работы пользователя.



# Системы реального времени

# Системы реального времени

Системы реального времени, предназначены для управления техническими объектами (например, станком, спутником, научной экспериментальной установкой и т. п.) или технологическими процессами (гальванической линией, доменным процессом и т. п.).

# Системы реального времени

При таком управлении существует понятие предельно допустимого времени, в течение которого должна быть выполнена та или иная управляющая объектом программа.



# Системы реального времени

В противном случае может произойти авария: спутник выйдет из зоны видимости, данные, поступающие с датчиков, будут потеряны, толщина гальванического покрытия не будет соответствовать норме и т.п.

# Системы реального времени

Т.О., критерием эффективности здесь является способность выдерживать заранее заданные интервалы времени между запуском программы и получением результата (управляющего воздействия).

# Системы реального времени

Это время называется временем реакции системы, а соответствующее свойство системы — реактивностью.

# Системы реального времени

В системах реального времени мультипрограммная смесь представляет собой фиксированный набор заранее разработанных программ, а выбор программы на выполнение осуществляется или по прерываниям (исходя из текущего состояния объекта) или в соответствии с расписанием запланированных работ.

# Системы реального времени

Способность аппаратуры компьютера и ОС к быстрому ответу зависит от скорости переключения с одной задачи на другую и от скорости обработки сигналов прерывания.

# Системы реального времени

Время обработки прерывания в системах реального времени определяет требования к производительности процессора даже при небольшой его загрузке.

# Системы реального времени

Если при возникновении прерывания процессор должен опросить сотни потенциальных источников прерывания, то реакция системы будет слишком медленной.

# Системы реального времени

В системах реального времени не стремятся максимально загружать все устройства, наоборот, при проектировании ИС обычно закладывается резерв вычислительной мощности на случай пиковой нагрузки.



# Системы реального времени

Например, в системе управления атомной электростанцией в случае возникновения крупной аварии атомного реактора многие аварийные датчики сработают одновременно и создадут повышенную нагрузку.

# Системы реального времени

Если система реального времени не спроектирована для поддержки пиковой нагрузки, то может случиться так, что система не справится с работой именно тогда, когда она нужна в наибольшей степени.

# Системы реального времени

Основной особенностью таких систем является строго регламентированное время отклика на внешние события.

# Системы реального времени

Другим важным параметром является «одновременная» обработка - даже если одновременно происходит несколько событий, реакция системы на них не должна запаздывать.

# Системы реального времени

Компьютеры для управления самолетами, ядерными реакторами и подобными сложными системами обычно работают под управлением специализированных ОС реального времени.

# Системы реального времени

ОС реального времени принято делить на два класса: *жесткого* и *мягкого* реального времени.

# Системы реального времени

Признаки систем *жёсткого реального времени*:

- недопустимость никаких задержек ни при каких условиях;
- бесполезность результатов при опоздании;
- катастрофа при задержке реакции;
- цена опоздания бесконечно велика.

# Системы реального времени

Пример системы жесткого реального времени — бортовая система управления самолетом.

Среди систем с жестким реальным временем можно выделить распространённую коммерческую операционную систему QNX, которая основывается на UNIX.



# Системы реального времени

*Система мягкого реального времени* характеризуется следующими признаками:

- при опоздании результатов не возникает катастрофических последствий;
- снижение производительности системы, вызванное запаздыванием реакций, приемлемо.

# Системы реального времени

ОС мягкого реального времени могут использоваться в мобильных и коммуникационных системах — там, где цена опоздания не так велика.



# Мультипроцессорная обработка

# Мультипроцессорная обработка

Мультипроцессорная обработка — это способ организации вычислительного процесса в системах с несколькими процессорами, при котором несколько задач (процессов, потоков) могут одновременно выполняться на разных процессорах системы.

# Мультипроцессорная обработка

Мультипроцессорная обработка отличается от мультипрограммной обработки.

# Мультипроцессорная обработка

В мультипрограммных системах параллельная работа разных устройств позволяет одновременно вести обработку нескольких программ, но при этом в процессоре в каждый момент времени **выполняется только одна программа**, т. е. в этом случае несколько задач выполняются попеременно на одном процессоре, создавая лишь **видимость параллельного выполнения.**

# Мультипроцессорная обработка

В мультипроцессорных системах несколько задач выполняются действительно одновременно, так как имеется несколько обрабатывающих устройств — процессоров.

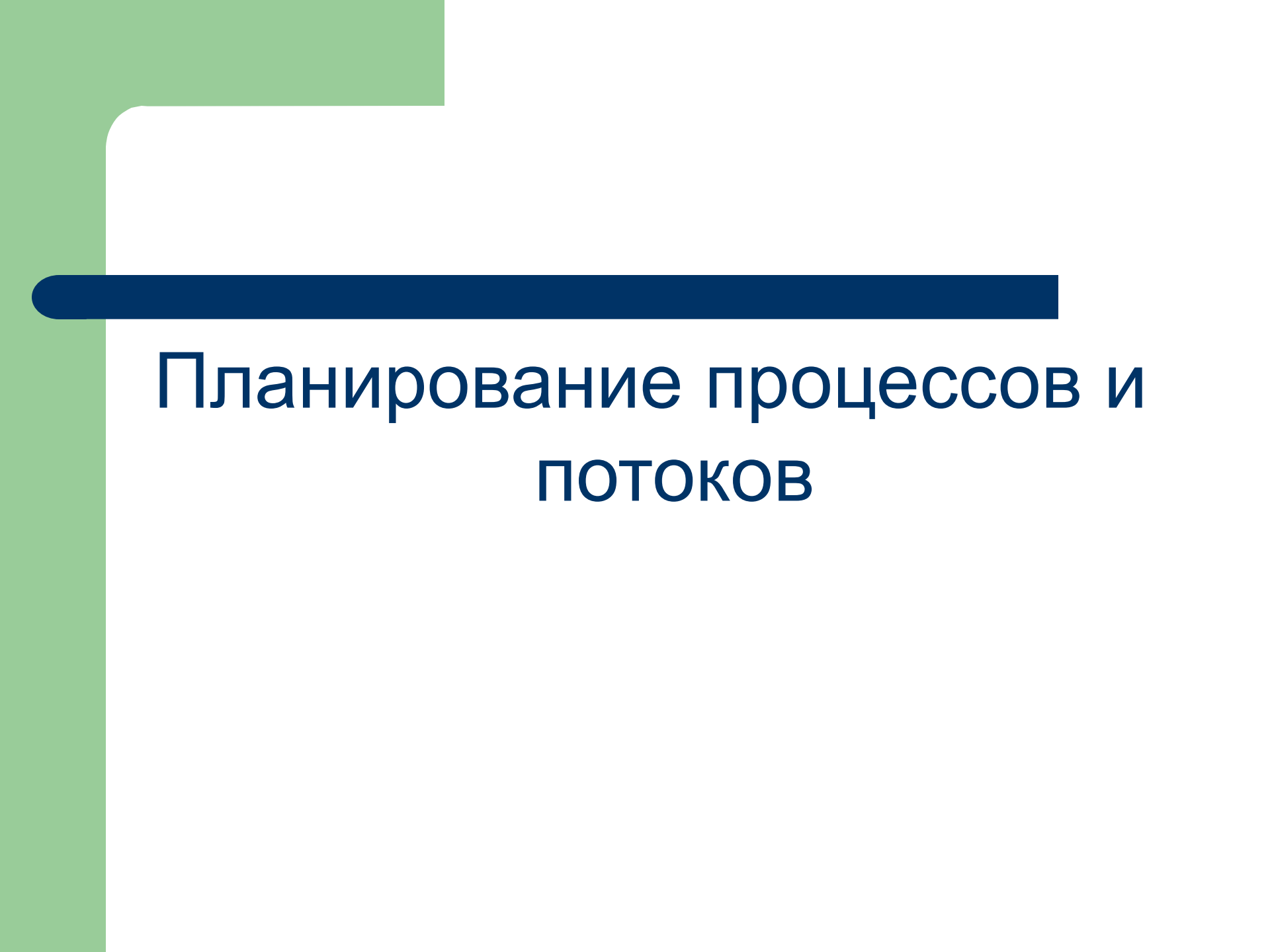
# Мультипроцессорная обработка

Мультипроцессирование не исключает мультипрограммирования: на каждом из процессоров может попеременно выполняться некоторый закрепленный за данным процессором набор задач.



# Мультипроцессорная обработка

Функции мультипроцессорной обработки данных имеются во всех популярных ОС, таких как Sun Solaris 2.x, Santa Crus Operations Open Server 3.x, IBM OS/2, Microsoft Windows NT и Novell NetWare, начиная с 4.1.



# Планирование процессов и ПОТОКОВ



# Потоки и процессы

# Потоки и процессы

Потоки и процессы — это связанные понятия в вычислительной технике.

Оба понятия представляют из себя последовательность инструкций, которые должны выполняться в определенном порядке.

# Потоки и процессы

Процессы существуют в ОС и соответствуют тому, что пользователи видят как программы или приложения.

# Потоки и процессы

Поток, с другой стороны, существует внутри процесса, по этой причине потоки иногда называются "облегченные процессы".

Каждый процесс состоит из одного или более потоков.

# Процессы и потоки

**Потоки выполнения** (англ. *Thread* — нить) отличаются от процессов многозадачной ОС тем, что:

- процессы независимы, тогда как потоки существуют как составные элементы процессов ;

# Процессы и потоки

- несколько потоков выполнения внутри процесса совместно используют информацию о состоянии, а также память и другие вычислительные ресурсы процесса;



# Процессы и потоки

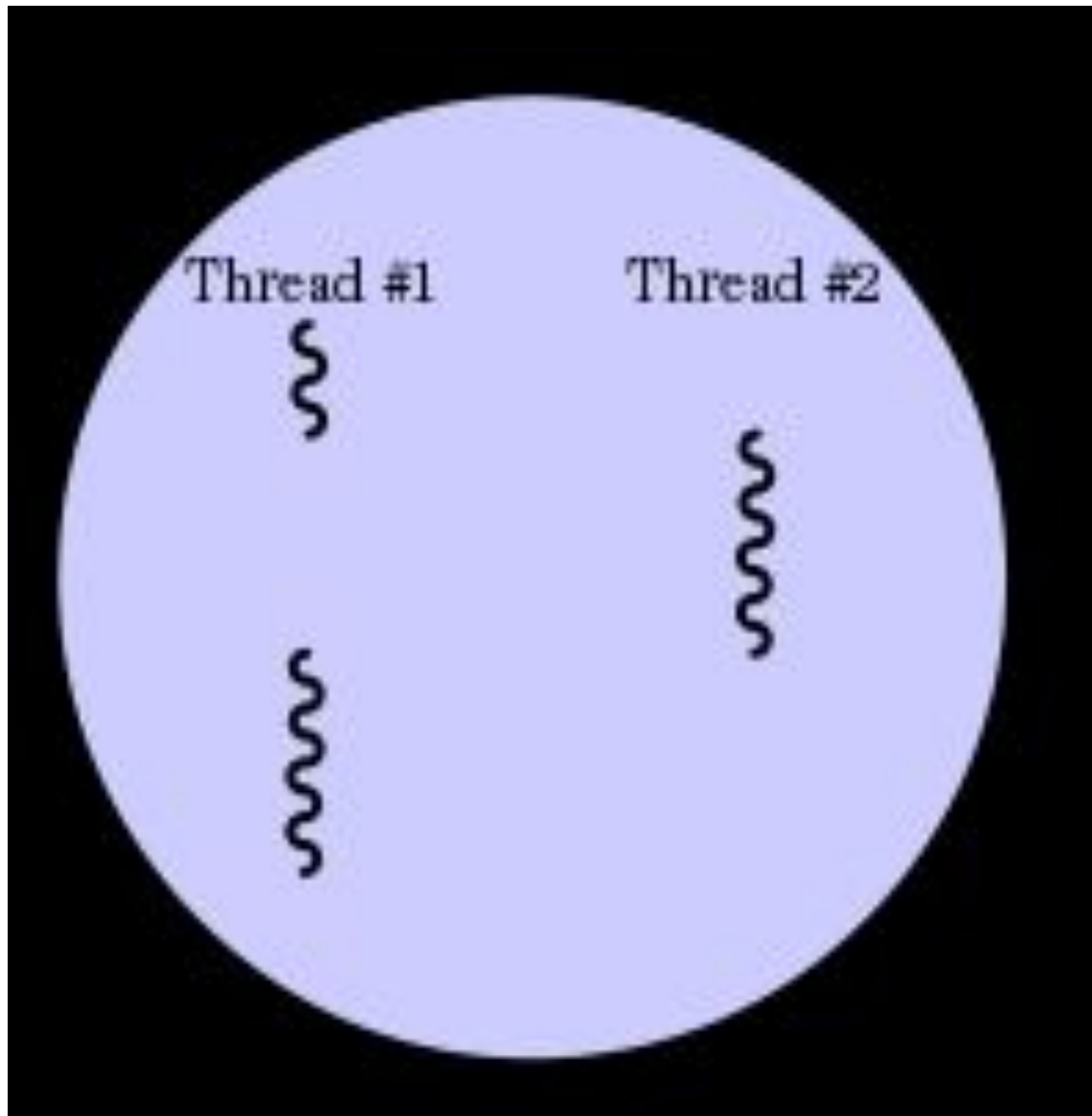
- процессы имеют отдельные адресные пространства, тогда как потоки выполнения совместно используют адресное пространство процессов;

# Процессы и потоки

- процессы взаимодействуют только через предоставляемые ОС механизмы связей между процессами;

# Процессы и потоки

- переключение контекста между потоками выполнения в одном процессе происходит быстрее, чем переключение контекста между процессами.



Процесс содержащий два потока выполнения

# Потоки и процессы

Существование нескольких процессов позволяет компьютеру "одновременно" выполнять несколько задач.

# Потоки и процессы

Существование нескольких потоков позволяет разделять работу процесса для параллельного выполнения.

# Потоки и процессы

На многопроцессорном компьютере процессы или потоки могут работать на разных процессорах.

Это позволяет выполнять параллельную работу.

# Потоки и процессы

Возникающие проблемы синхронизации являются причиной ошибок в многопоточных приложениях:

- Один поток может ожидать результата другого потока;
- одному потоку может понадобиться монопольный доступ к ресурсу, который используется другим потоком;



# Потоки и процессы

- Поток может закончиться, ожидая ресурс, который никогда не будет доступен.

Это состояние называется взаимоблокировкой.

# Планирование процессов и потоков

Одной из основных подсистем ОС, является подсистема управления процессами и потоками, которая

- занимается их созданием и уничтожением,
- поддерживает взаимодействие между ними,
- распределяет процессорное время между несколькими одновременно существующими в системе процессами и потоками.

# Планирование процессов и потоков

- ОС для реализации мультипрограммирования выполняет планирование и диспетчеризацию потоков (в ОС, не поддерживающих потоков, — диспетчеризацию процессов).

# Планирование процессов и потоков

- Планирование включает определение момента времени для смены текущего потока, а также выбор нового потока для выполнения.

# Планирование процессов и потоков

- Диспетчеризация заключается в реализации найденного в результате планирования решения, то есть в переключении процессора с одного потока на другой.

# Планирование процессов и потоков

- Планирование бывает динамическим, когда решения принимаются во время работы системы на основе анализа текущей ситуации, или статическим, если потоки запускаются на выполнение на основании заранее разработанного расписания.
- *Первый способ характерен для универсальных ОС, а второй — для специализированных ОС, (реального времени).*

# Планирование процессов и потоков

Динамический планировщик ОС может реализовывать различные алгоритмы планирования, которые делятся на классы:

- вытесняющие алгоритмы
- невытесняющие алгоритмы
- алгоритмы квантования
- приоритетные алгоритмы.

# Планирование процессов и потоков

Используемый алгоритм планирования зависит от назначения ОС.

Применяются также смешанные алгоритмы, объединяющие достоинства нескольких классов.



# Планирование процессов и потоков

Невытесняющие (non-preemptive) алгоритмы основаны на том, что активному потоку позволяется выполняться, пока он сам, по собственной инициативе, не отдаст управление ОС для того, чтобы та выбрала из очереди другой ГОТОВЫЙ к выполнению поток.

# Планирование процессов и потоков

Вытесняющие (preemptive) алгоритмы - решение о переключении процессора с выполнения одного потока на выполнение другого потока принимается ОС, а не активной задачей.

# Планирование процессов и потоков

В основе многих вытесняющих алгоритмов планирования лежит концепция квантования.

# Планирование процессов и потоков

В соответствии с этой концепцией каждому потоку поочередно для выполнения предоставляется ограниченный непрерывный период процессорного времени - квант.

# Планирование процессов и потоков

Смена активного потока происходит, если:

- поток завершился и покинул систему;
- произошла ошибка;
- поток перешел в состояние ожидания;
- исчерпан квант процессорного времени, отведенный данному потоку.

# Вытесняющий алгоритм



# Планирование процессов и потоков

Поток, который исчерпал свой квант, переводится в состояние готовности и ожидает, когда ему будет предоставлен новый квант процессорного времени, а на выполнение в соответствии с определенным правилом выбирается новый поток из очереди готовых.

# Планирование процессов и потоков

Другой важной концепцией, лежащей в основе многих вытесняющих алгоритмов планирования, является приоритетное обслуживание (приоритетный алгоритм).



# Планирование процессов и потоков

Приоритетное обслуживание предполагает наличие у потоков некоторой изначально известной характеристики — приоритета, на основании которой определяется порядок их выполнения.

# Планирование процессов и потоков

Приоритет - это число, характеризующее «важность» потока при использовании ресурсов вычислительной машины *(в частности процессорного времени)*: Чем выше приоритет, тем выше привилегии, тем меньше времени будет проводить поток в очередях.

# Планирование процессов и потоков

В большинстве ОС, поддерживающих потоки, приоритет потока непосредственно связан с приоритетом процесса, в рамках которого выполняется данный поток.

# Планирование процессов и потоков

Приоритет процесса назначается ОС при его создании.

Значение приоритета включается в дескриптор процесса и используется при назначении приоритета потокам этого процесса.



# Стандарты POSIX

# Стандарт POSIX

**POSIX Threads** — стандарт реализации потоков (нитей) выполнения, определяющий *API* для создания и управления ими.

# Что такое API

**Интерфейс программирования приложений** (иногда **интерфейс прикладного программирования**) (англ. *application programming interface, API* [эй-пи-ай]) — набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) для использования во внешних программных продуктах.

*Используется программистами для написания всевозможных приложений.*

# POSIX®

**POSIX®** (*P*ortable *O*perating *S*ystem *I*nterface for *U*nix — Переносимый интерфейс операционных систем *U*nix) — набор стандартов, описывающих интерфейсы между ОС и прикладной программой.



# POSIX®

Стандарт создан для обеспечения совместимости различных UNIX-подобных ОС и переносимости прикладных программ на уровне исходного кода, но может быть использован и для не-Unix систем.

# POSIX®

Серия стандартов POSIX была разработана комитетом 1003 IEEE.

Международная организация по стандартизации (ISOМеждународная организация по стандартизации (ISO) совместно с Международной электротехнической комиссией (IEC) приняли данный стандарт (POSIX) под названием **ISO/IEC 9945**.

# POSIX®

Название «POSIX» было предложено Ричардом Столлманом.

Введение в POSIX.1 гласит: «Ожидается произношение „поз-икс“ как „позитив“, а не „посикс“. Произношение опубликовано в целях обнародования стандартного способа ссылки на стандартный интерфейс операционной системы».

«POSIX» является зарегистрированным товарным знаком IEEE.

# Стандарт POSIX

- Библиотеки, реализующие этот стандарт (и функции этого стандарта), обычно называются **Pthreads** (функции имеют приставку «pthread\_»).
- Наиболее известны варианты для Unix-подобных ОС, таких как Linux или Solaris, но существует и реализация для Microsoft Windows (Pthreads-w32)
- *Pthreads определяет набор типов и функций на языке программирования Си.*



# Система прерываний

# Система прерываний

**Прерывание** ( *interrupt* ) — сигнал, сообщающий процессору о наступлении какого-либо события.

# Система прерываний

Выполнение текущей последовательности команд приостанавливается и управление передаётся обработчику прерывания, который реагирует на событие и обслуживает его, после чего возвращает управление в прерванный процесс.

# Система прерываний

В зависимости от источника возникновения сигнала прерывания делятся на:

1. асинхронные или внешние (аппаратные) — события, которые исходят от внешних источников (например, периферийных устройств)



# Система прерываний

Такие прерывания могут произойти в любой произвольный момент: сигнал от таймера, сетевой карты или дискового накопителя, нажатие клавиш клавиатуры, движение мыши.

Факт возникновения в системе такого прерывания называется *запрос на прерывание (Interrupt request, IRQ)*;

# Система прерываний

2. синхронные или внутренние — события в самом процессоре как результат нарушения каких-то условий при исполнении машинного кода: деление на ноль или переполнение, обращение к недопустимым адресам или недопустимый код операции;

# Система прерываний

3. программные (частный случай внутреннего прерывания) — инициируются исполнением специальной инструкции в коде программы.

# Система прерываний

Программные прерывания используются для обращения к функциям встроенного программного обеспечения (*firmware*), драйверов и ОС.

# Система прерываний

Прерывания являются основной движущей силой любой ОС.

# Система прерываний

Прерывания от таймера вызывают смену процессов в мультипрограммной ОС, а прерывания от устройств ввода-вывода управляют потоками данных, которыми вычислительная система обменивается с внешним миром.

# Система прерываний

Механизм прерываний поддерживается аппаратными средствами компьютера и программными средствами ОС и состоит из след. действий:

# Система прерываний

1. При возникновении сигнала (для аппаратных прерываний) или условия (для внутренних прерываний) прерывания происходит первичное аппаратное распознавание типа прерывания.



# Система прерываний

Если прерывание данного типа в настоящий момент запрещены (маскированы), то процессор игнорирует его и продолжает естественный ход выполнения команд.

# Система прерываний

Иначе происходит автоматический вызов процедуры обработки прерывания, адрес которой находится в специальной таблице ОС, размещаемой либо в регистрах процессора, либо в определенном месте ОЗУ.

# Система прерываний

2. Автоматически сохраняется контекст прерванного потока (*значения счетчика команд, слова состояния машины, а также регистров общего назначения*), **КОТОРЫЙ** позволит ОС возобновить исполнение этого потока после обработки прерывания.

# Система прерываний

*Может быть сохранен и полный контекст процесса, если ОС обслуживает данное прерывание со сменой процесса.*

*Однако в общем случае это не обязательно, часто обработка прерываний выполняется без вытеснения текущего процесса.*

# Система прерываний

Решение о перепланировании процессов может быть принято в ходе обработки прерывания, например, если это прерывание от таймера и после наращивания значения системных часов выясняется, что процесс исчерпал выделенный ему квант времени.

# Система прерываний

Прерывание может выполняться и без смены процесса, например, прием очередной порции данных от контроллера внешнего устройства чаще всего происходит в рамках текущего процесса, хотя данные могут быть предназначены другому процессу.

# Система прерываний

3. Одновременно с загрузкой адреса процедуры обработки прерываний в счетчик команд может автоматически загрузиться нового значения слова состояния машины, которое определяет режимы работы процессора при обработке прерывания, в том числе работу в привилегированном режиме.

# Система прерываний

Прерывания практически во всех мультипрограммных ОС обрабатываются в привилегированном режиме модулями ядра.



# Система прерываний

4. Временно запрещаются прерывания данного типа, чтобы не образовалась очередь вложенных друг в друга потоков одной и той же процедуры.

# Система прерываний

5. После того как прерывание обработано ядром ОС, прерванный контекст восстанавливается и работа потока возобновляется с прерванного места.

# Система прерываний

Часть контекста восстанавливается аппаратно по команде возврата из прерываний (например, адрес следующей команды и слово состояния машины), а часть — программным способом, с помощью явных команд извлечения данных из стека.

# Система прерываний

При возврате из прерывания  
блокировка повторных прерываний  
данного типа снимается.

# Система прерываний

Существуют два способа обработки прерывания:

- . векторный (vectored), когда в процессор передается номер вызываемой процедуры обработки прерывания,
- . опрашиваемый (polled), когда процессор вынужден последовательно опрашивать потенциальные источники прерывания.

# Система прерываний

- Для упорядочивания процессов обработки прерываний применяется система приоритетов: прерывания распределяются по нескольким приоритетным уровням, а роль арбитра выполняет диспетчер прерываний ОС.

# Система прерываний

Для синхронизации процессов и потоков, решающих общие задачи и совместно использующих ресурсы, в ОС существуют специальные средства: критические секции, семафоры, мьютексы, события, таймеры.

# Система прерываний

Отсутствие синхронизации может приводить к таким нежелательным последствиям, как гонки и тупики.



- ДЗ: Что такое критические секции, семафоры, мьютексы, события, таймеры, «Гонки и тупики».