



Part 1. Overview

第一章 操作系统导论

Note : These lecture materials are based on the lecture notes prepared by the authors of the book titled Operating System Concepts.



掌握计算机系统的基本组织结构，掌握操作系统的主要组成、操作和管理，了解其他计算机系统以及计算环境



1. 操作系统的定义
2. 计算机系统的结构
3. 计算机系统的组织
4. 计算机系统的体系结构
5. 操作系统的结构和操作
6. 操作系统管理内容
7. 介绍其他计算机系统



第一节 操作系统的定义

1. 操作系统的定义

Windows Unix, Linux,
openEuler, Mac OS

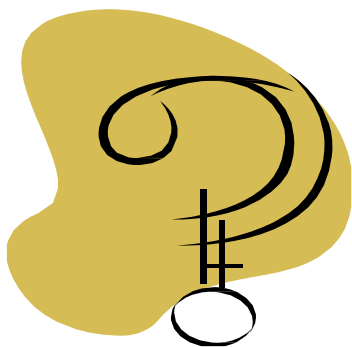
WinCE Android iOS 鸿蒙





1. 操作系统是什么？
2. 它的工作内容是什么？
3. 它是软件还是硬件？

计算机用户与计算机硬件之间运行的一个程序。



1. 操作系统是什么？
2. 它的工作内容是什么？
3. 它是软件还是硬件？

1. 资源管理服务：资源包括硬件资源和软件资源
2. 为用户提供服务：交互、提供应用程序的运行环境等



1. 硬件资源：中央处理器、内存、硬盘、键盘等
2. 软件资源：MS Office, eclipse, Visual Studio etc.
3. 服务提供：方便用户

操作系统的目标？

1. 通过运行计算机的应用程序，方便用户解决问题
2. 方便用户操作使用计算机
3. 有效使用计算机硬件资源



- 操作系统是一直运行在计算机上的程序，通常狭义上被称为内核（Kernel）程序，其他程序则为系统程序和应用程序。
- 它是资源管理平台，运行程序的平台，也是为用户提供服务的平台

Wikipedia Definition:

An operating system is system software that manages computer hardware and software resources and provides common services for computer programs.



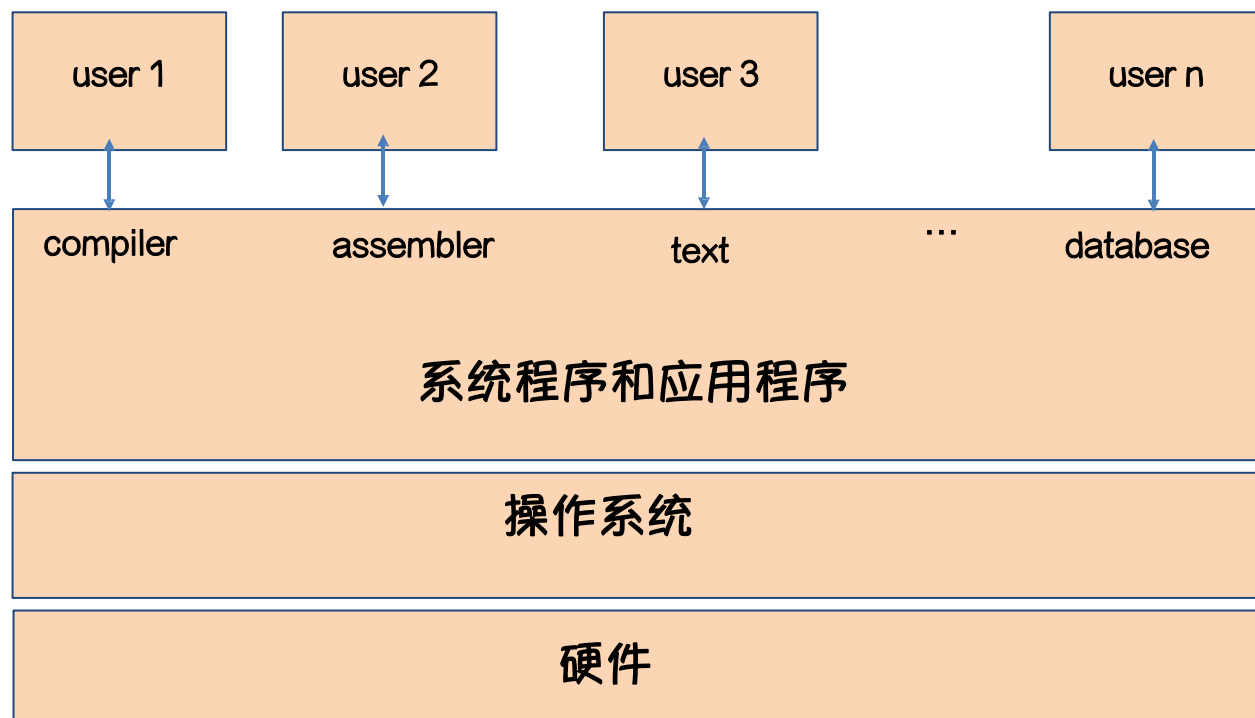
第二节

计算机系统的结构

(Computer System Structure)

2. 计算机系统的结构

计算机的系统结构是层次结构



2. 计算机系统的结构



计算机系统的结构是层次结构

1. 硬件

提供最基本的计算资源，如CPU，内存，I/O设备等

2. 操作系统

协调控制多个应用或多个用户对硬件的使用

3. 应用程序

为解决用户的问题，定义了解决方法程序，比如文本编辑器，网页浏览器，数据库，游戏等

4. 用户

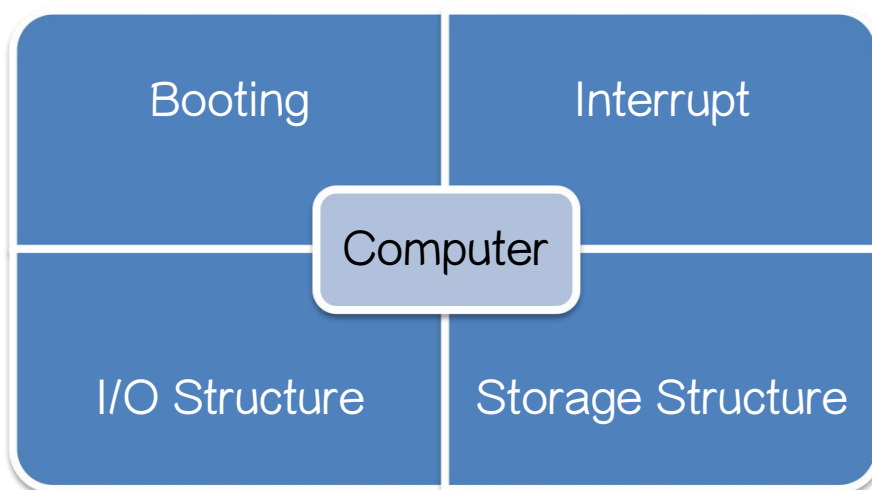
人、设备、其他计算机系统

第三节

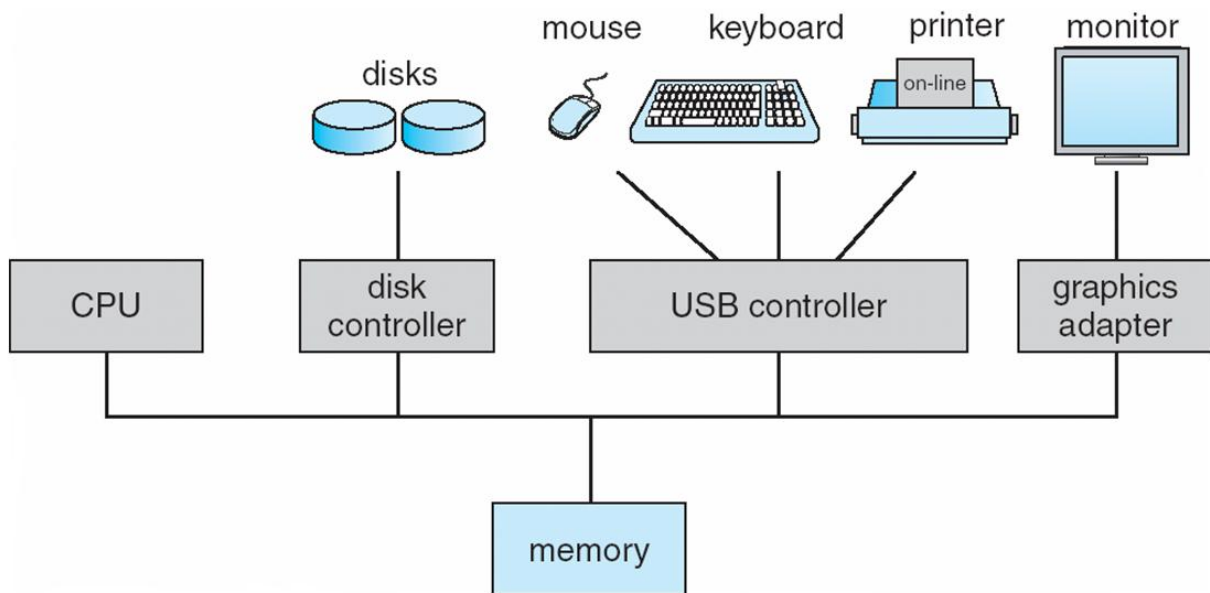
计算机系统的组织



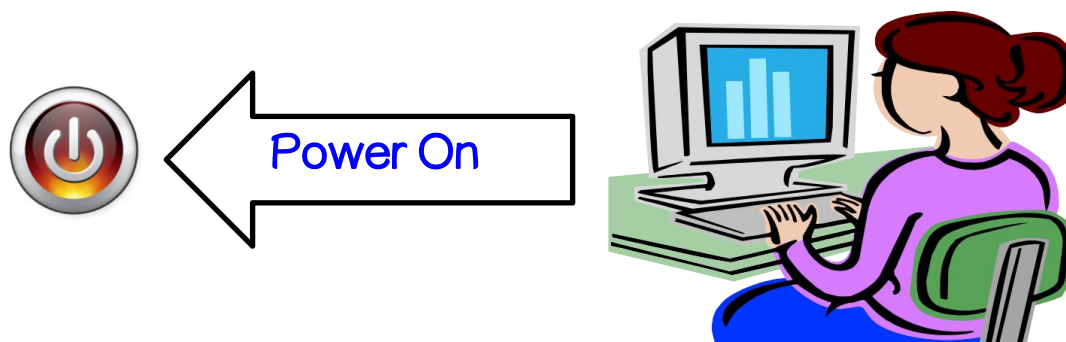
1. 启动
2. 中断
3. I/O 结构
4. 存储结构



3. 计算机系统的组织

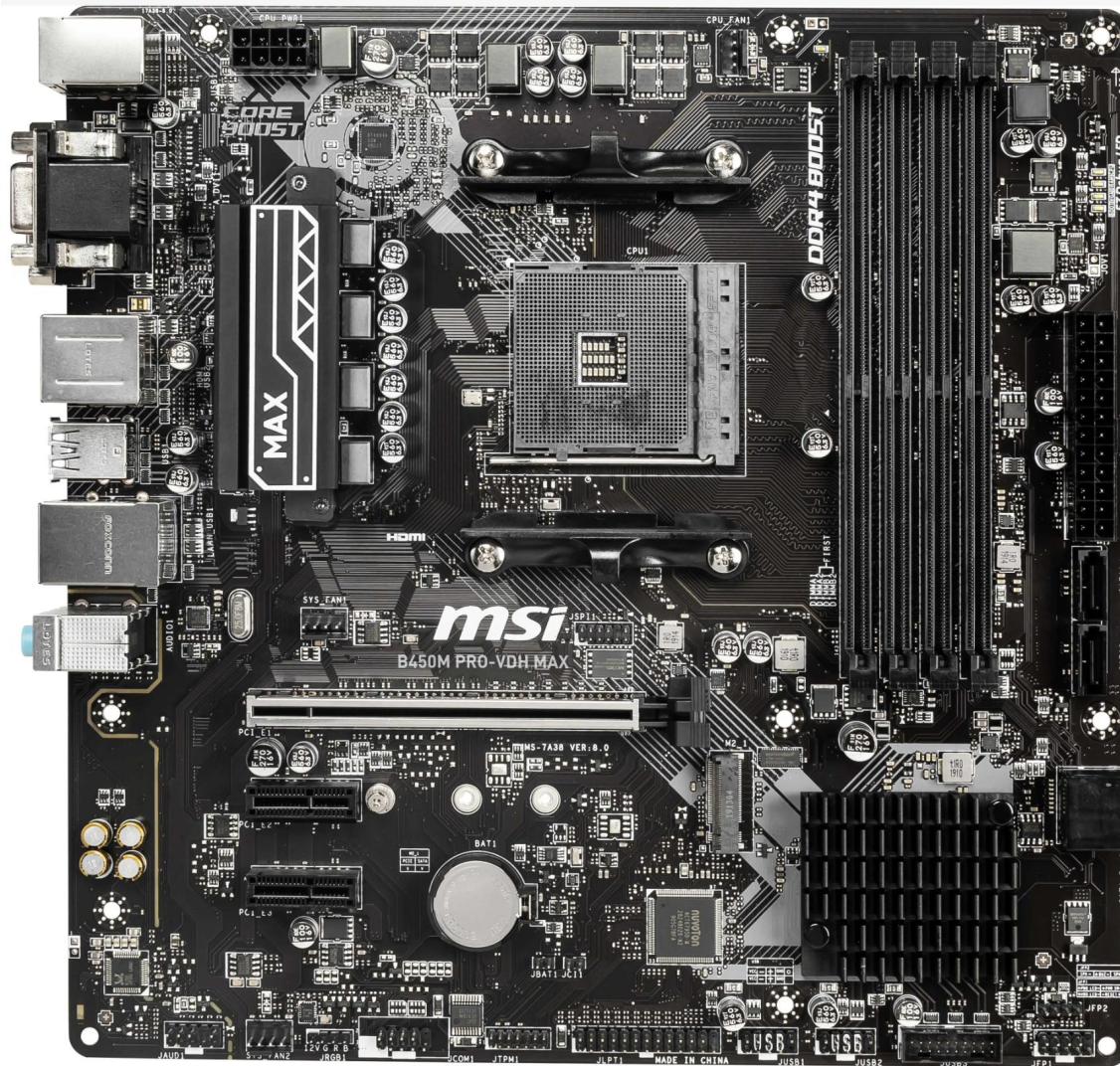


单个或多个中央处理器通过总线与内存和设备控制器相连。



What to do?

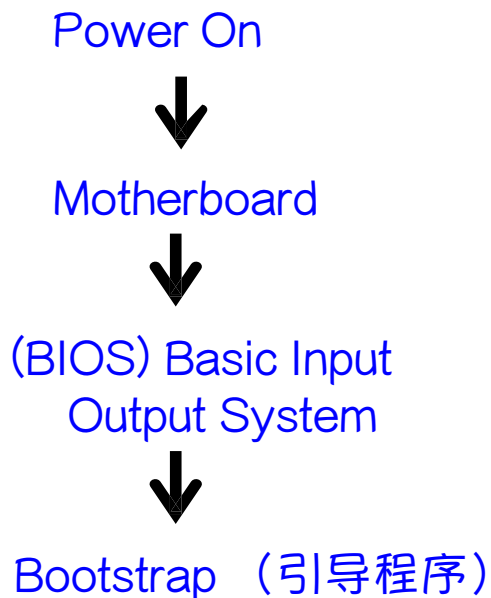
Motherboard



BIOS (Basic Input and Output)



(http://shunsheng.yeah.net)		Guxing Soft
CMOS SETUP UTILITY AWARD SOFTWARE, INC.		
STANDARD CMOS SETUP	INTEGRATED PERIPHERALS	
BIOS FEATURES SETUP	SUPERVISOR PASSWORD	
CHIPSET FEATURES SETUP	USER PASSWORD	
POWER MANAGEMENT SETUP	IDE HDD AUTO DETECTION	
PNP/PCI CONFIGURATION	SAVE & EXIT SETUP	
LOAD BIOS DEFAULTS	EXIT WITHOUT SAVING	
LOAD OPTIMUM SETTINGS		
Esc : Quit	↑ ↓ → ← : Select Item	
F10 : Save & Exit Setup	<Shift>F2 : Change Color	
Time, Date, Hard Disk, Type...		



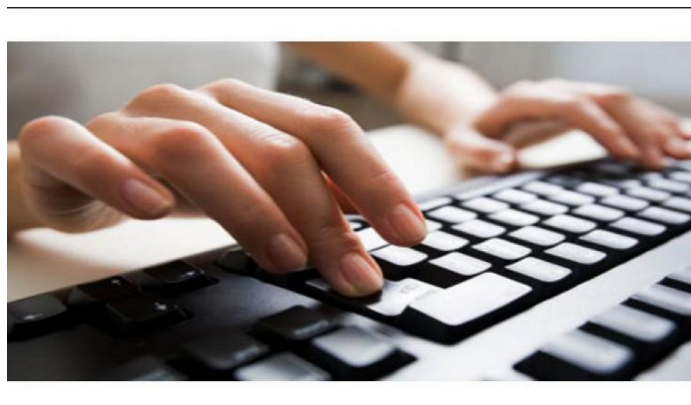
确认每个设备是否工作正常。确认无误后，开启引导程序 - bootstrap program

引导程序一般位于 ROM(Read-Only Memory) 或 EEROM(Eelectrically Easable Programable ROM) 中称为计算机硬件中的固件(firmware)。

引导程序的工作内容包括：

1. 初始化设备
2. 把操作系统载入到内存中
3. 运行第一个进程 `init()`，等待事件发生

How does the computer system work?



Touch , Watch , Tell - - Interrupt



- 一个事件的触发是通过硬件或软件中断来实现的
- 在实现操作系统功能时，中断是一个非常重要的实现机制

1. 软中断

即通过软件触发中断。系统调用 (System Call) 会触发软件中断。如异常，除零 (divide by zero)，**不可以屏蔽**

2. 硬中断

即硬件通过向 CPU 发送信号来触发中断，一般对用户不可见，但可触发程序的运行。如网卡接收数据包，**可以屏蔽**

A **trap** or **exception** is a software-generated interrupt caused either by an error or a user request

- 系统一旦发生中断，CPU会运行**中断服务程序(Interrupt Service Routine)**，且每个中断都有自己对应的中断服务程序 ISR
- 系统通过**中断向量表**（Interrupt Vector Table）来管理中
断请求与中断服务程序之间的对应关系， which contains the
addresses of all the service routines。
- 中断举例
 1. **如 I/O 设备发生的事件**，如按键盘、点击鼠标、磁盘
的读写等
 2. **异常事件、重要事件的发生**，如断电、部件失灵等
 3. **非法指令**，如除于零、下达不存在的指令等
- 例如，Linux 操作系统有 256个中断



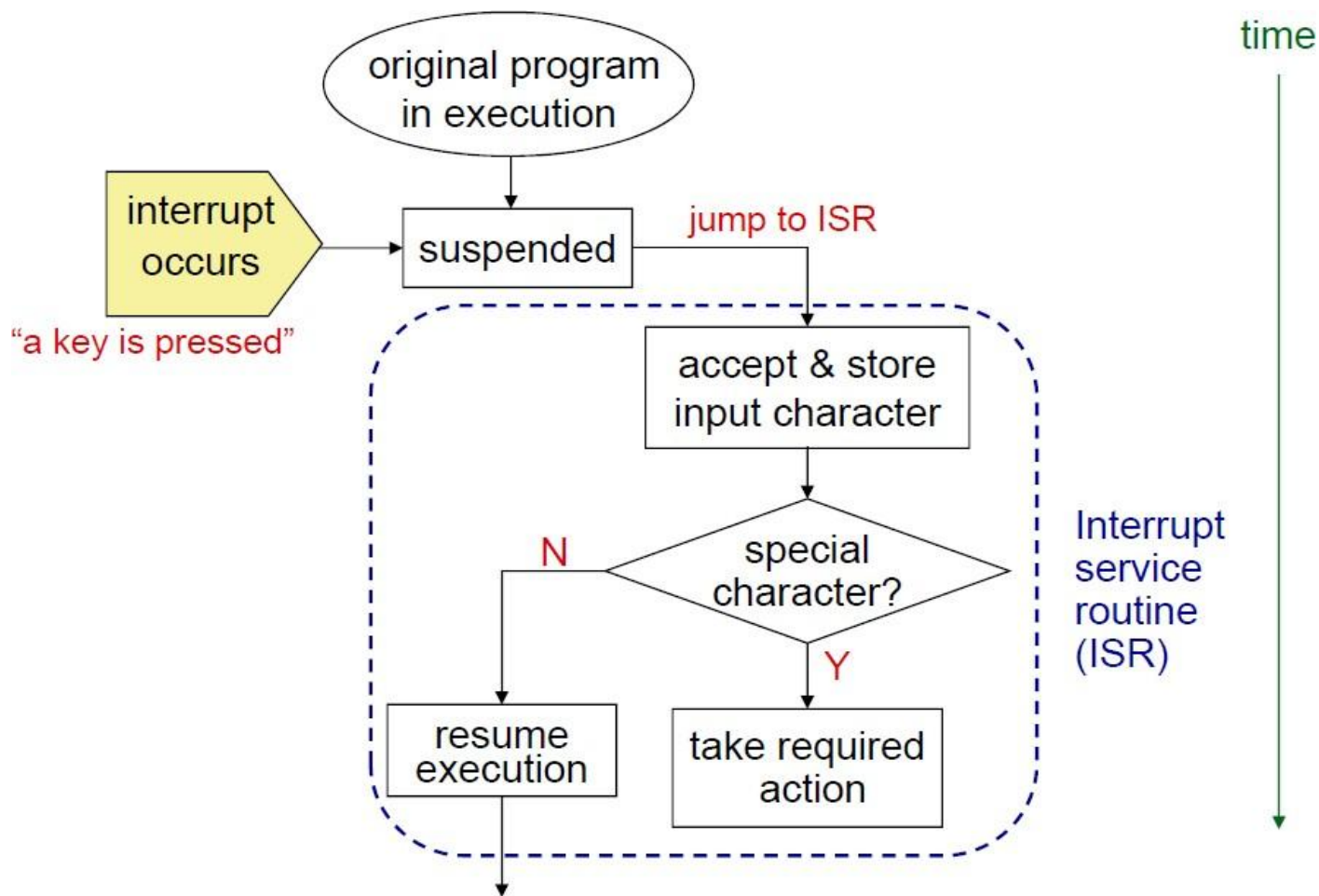
当发生中断时，

- 保存当前进程的计数器（Program Counter）
- 跳到相应的中断服务程序中，运行相应的中断服务程序

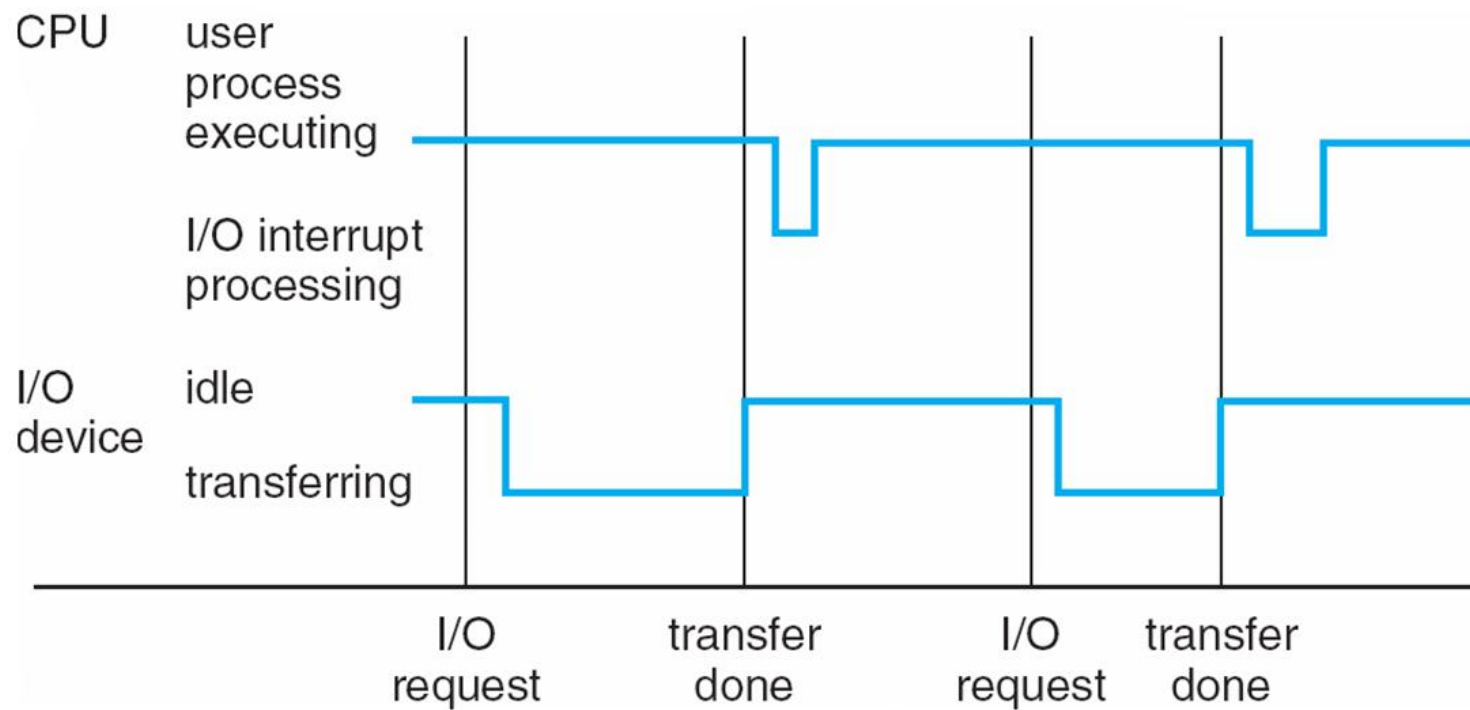
一旦中断服务程序运行结束，

- 返回到被中断的程序（by PC），并继续往下运行
- 或者返回到操作系统指定的程序中

3.2 中断- 按键盘为例



3.2 中断- I/O为例

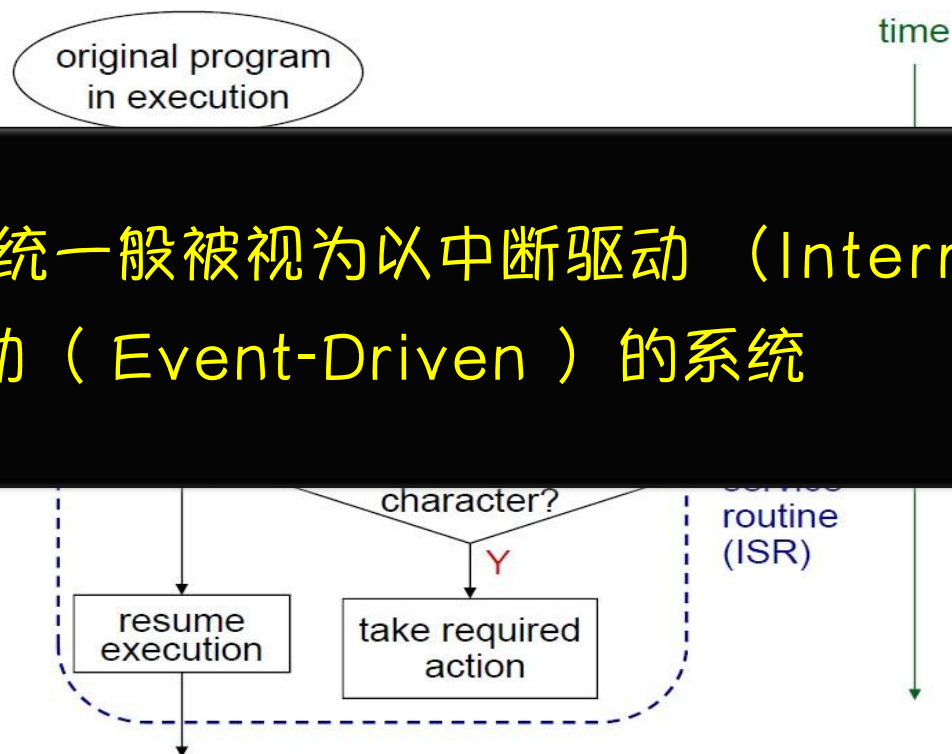


Interrupt Timeline

3.2 中断- 键盘为例



所以，现代的操作系统一般被视为以中断驱动（Interrupt-Driven）或事件驱动（Event-Driven）的系统





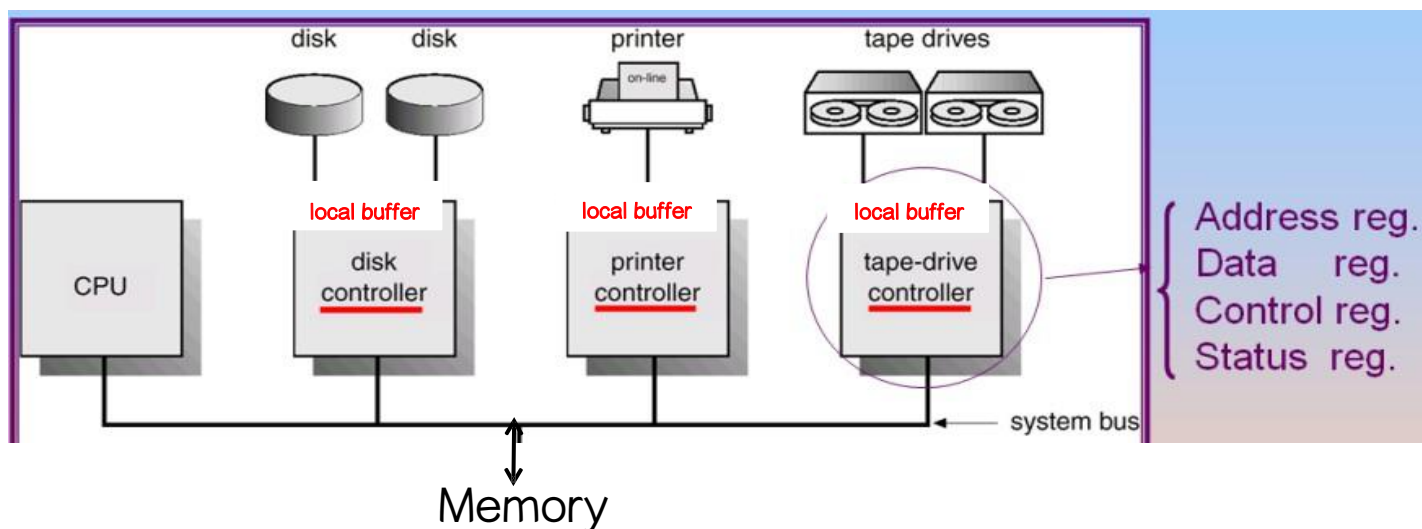
Stand I/O – “stdio.h” in C library

- | | | |
|---------|-----------------------------|-------------------------------|
| 1. 标准输入 | <code>#define stdin</code> | <code>__create_file(0)</code> |
| 2. 标准输出 | <code>#define stdout</code> | <code>__create_file(1)</code> |
| 3. 标准错误 | <code>#define stderr</code> | <code>__create_file(2)</code> |

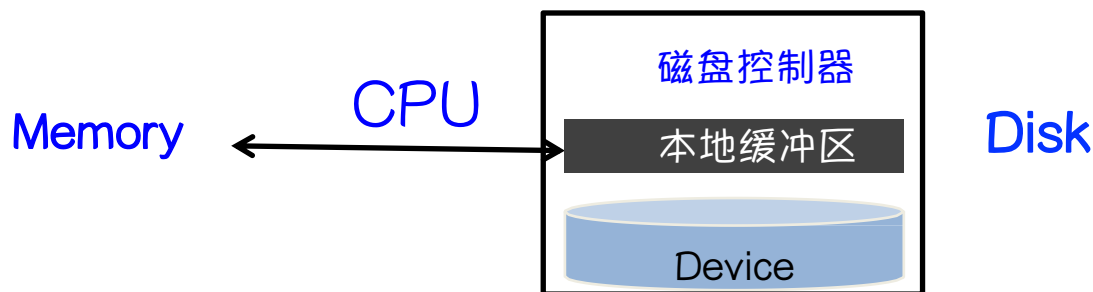
File descriptor (File pointer) 文件标识符

- 0 : 表示标准输入
- 1 : 表示标准输出
- 2 : 表示标准错误

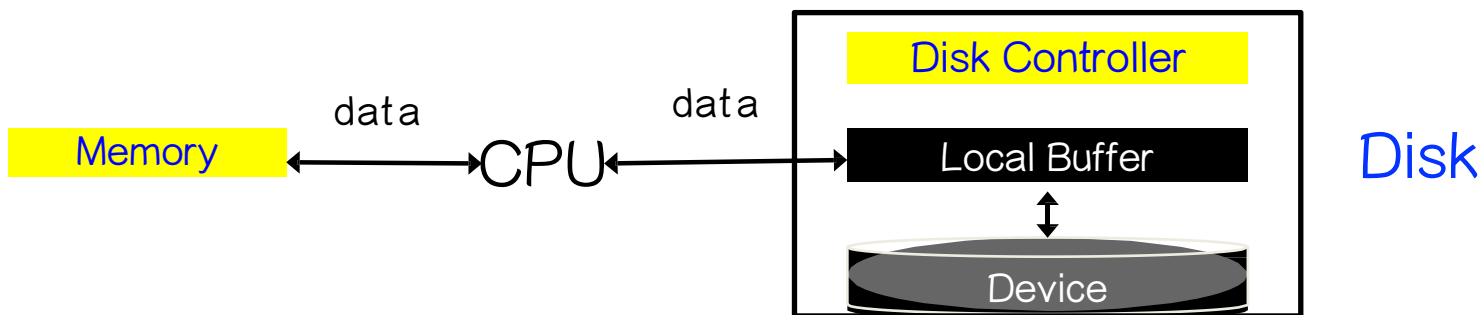
1. I/O设备与CPU可同时运行
2. 每个设备控制器负责相应类型的设备，如磁盘控制器负责磁盘
3. 每个设备控制器拥有自己的本地缓冲器（buffer）和寄存器



1. **CPU**负责内存与本地缓冲器之间的数据传递
2. **设备控制器**负责在其所控制的外部设备与本地缓冲存储之间进行数据传递
3. I/O 操作结束后，设备控制器通过中断通知 CPU 表示 I/O 结束



磁盘写操作的工作流程？



- 传输单位为字节(byte);
- CPU和磁盘控制器小容量的寄存器用于传输数据

会出现什么问题呢？

磁盘控制器每次传输数据时，通过中断通知CPU，表示I/O结束

Notice

- CPU and disk controller only have small registers for delivering data.
- Delivering unit is byte.

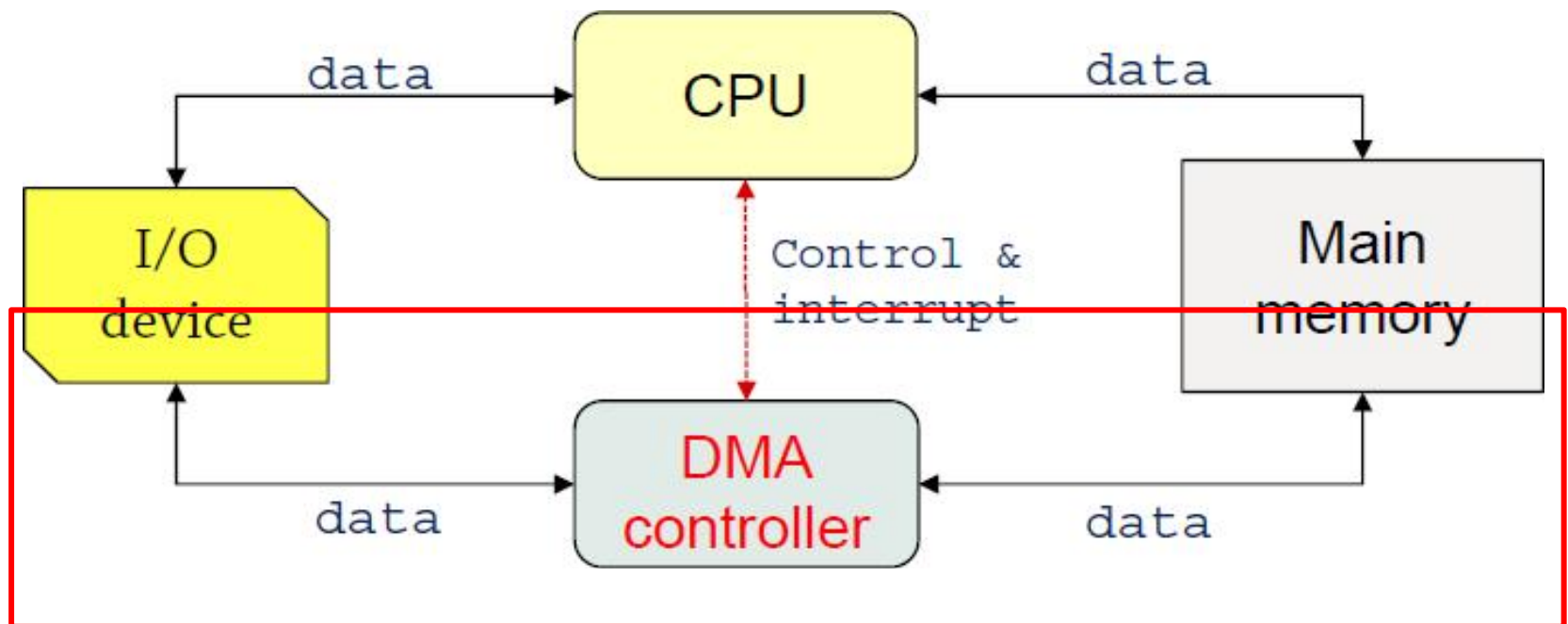


CPU说：你烦不烦？
能不能让我干点别的事儿呀？

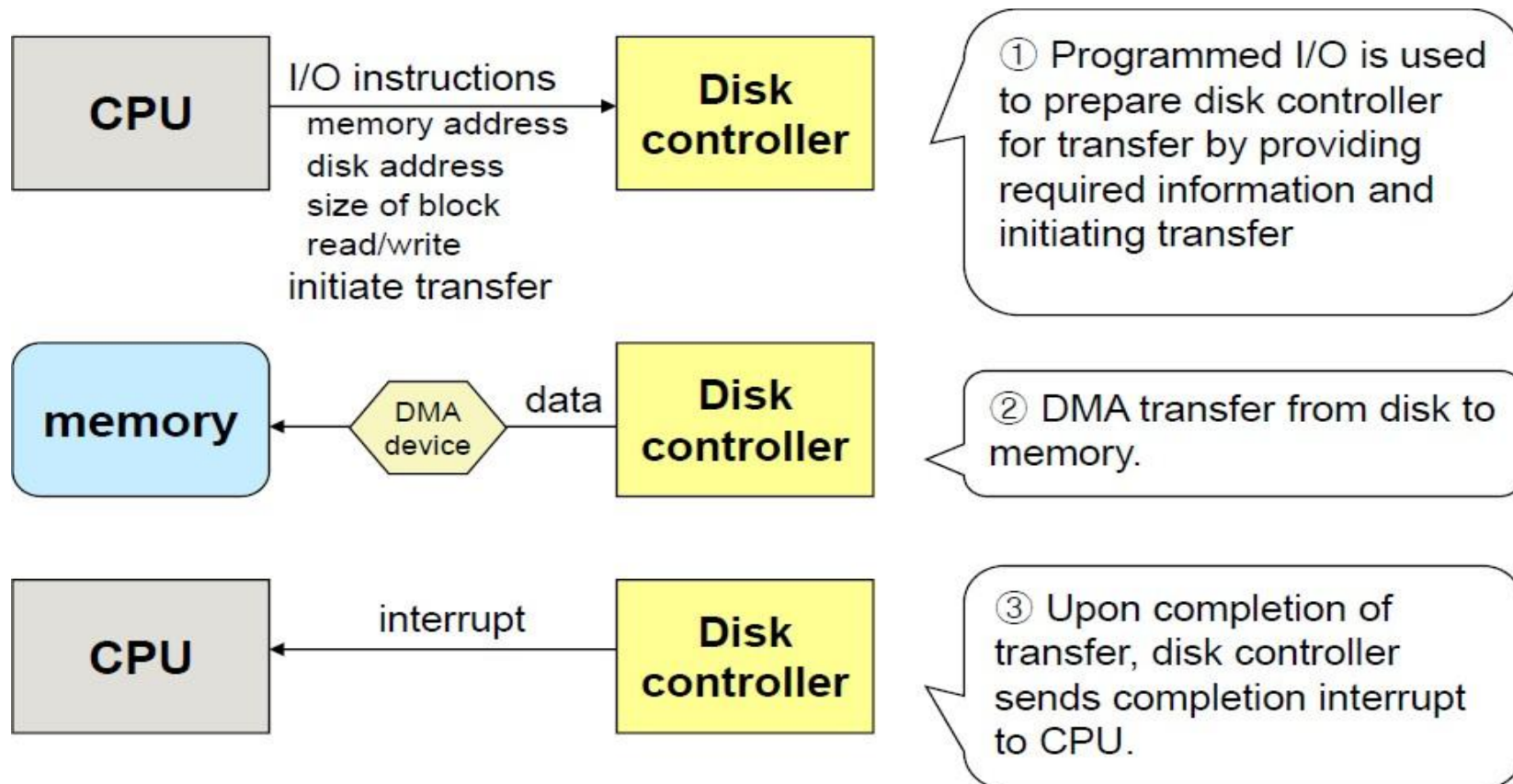


DMA (Direct Memory Access)直接访问内存是在专门的硬件控制下，实现高速外部设备与主存储器之间自动成批交换数据，目的就是**尽量减少CPU干预的I/O操作方式**

1. DMA控制器在没有CPU干涉的情况下，**以块/块(Block)为单位**，负责在设备缓冲器与主存之间的数据传送
2. 之前是每个字节传送完成后就触发中断表示完成，而DMA是每完成块/块传送后触发中断 **(字节 → 块)**
3. 有利于高速 I/O设备传送数据，接近于内存速度



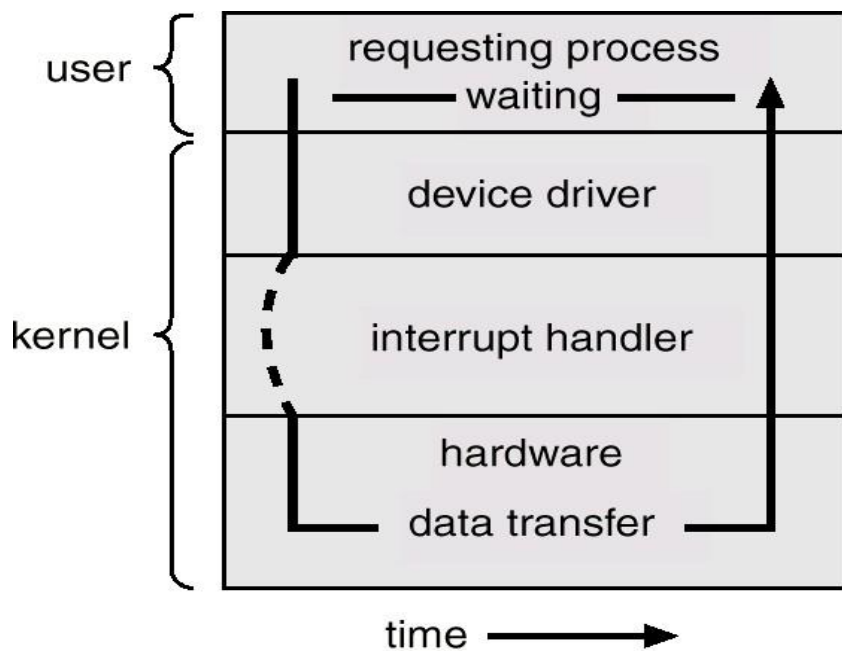
直接访问操作 - 读操作举例



3.3 I/O 结构 - I/O 操作方法

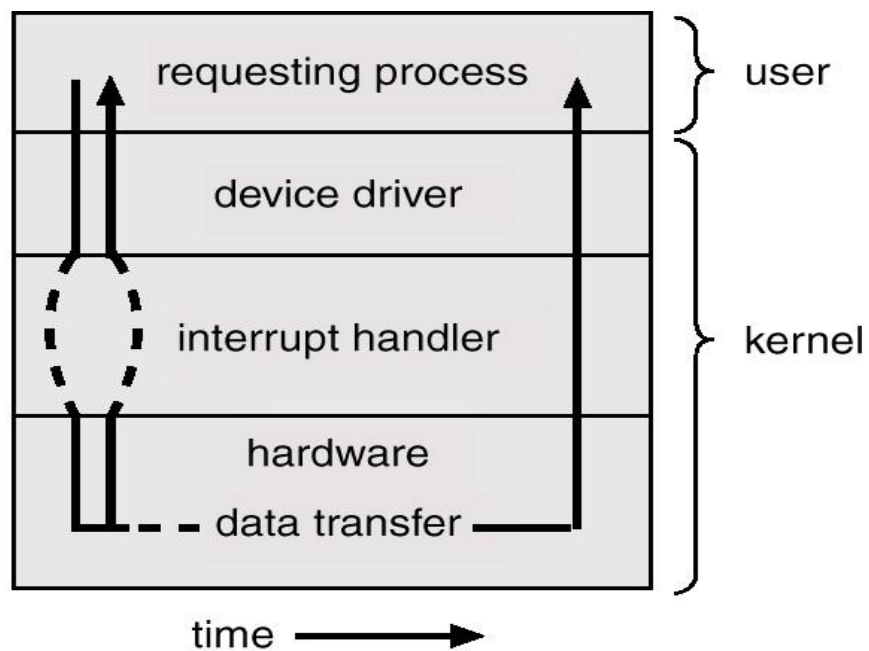


同步



(a)

异步



(b)



I/O 操作分为

1. **同步 (Synchronous) 操作**: 只有 I/O 结束后, 用户程序才能获得控制权, 即 I/O 进行期间, 用户程序无法继续运行。
2. **异步 (Asynchronous) 操作**: I/O 还没有结束的情况下, 用户程序可以获得控制权, 即 I/O 进行期间, 用户程序可以继续运行。

寄存器 { 1. 指令寄存器
2. 数据寄存器

CPU

Instruction Register

Data Register



memory

Disk

一级存储设备指内存（主存）

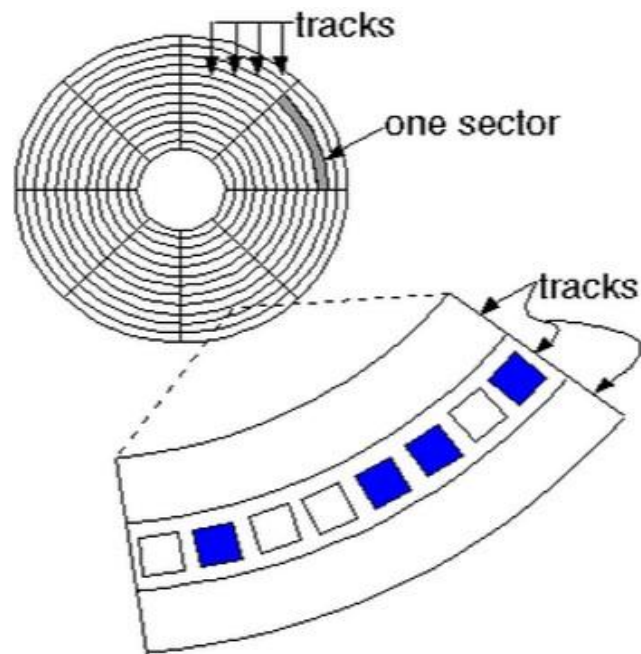
CPU可以直接随机访问的唯一大容量存储设备，**是易失存储设备**。

注：易失存储设备指的是断电后，数据会丢失的设备

二级存储设备：一般不易失的存储设备，一般是磁盘类

1. 磁盘(Magnetic Disks)

- 磁盘的表面可以逻辑划分为多个磁道(track)，而每个磁道在划分为扇区(sector)
- 磁盘控制器负责设备与计算机之间的逻辑交互



2. 闪存(Flash Memory)

- 芯片级别，移动存储设备
- 分为 Nand Flash 和 Nor Flash

3. 固态硬盘，简称固态硬盘SSD:Solide-State-Disk

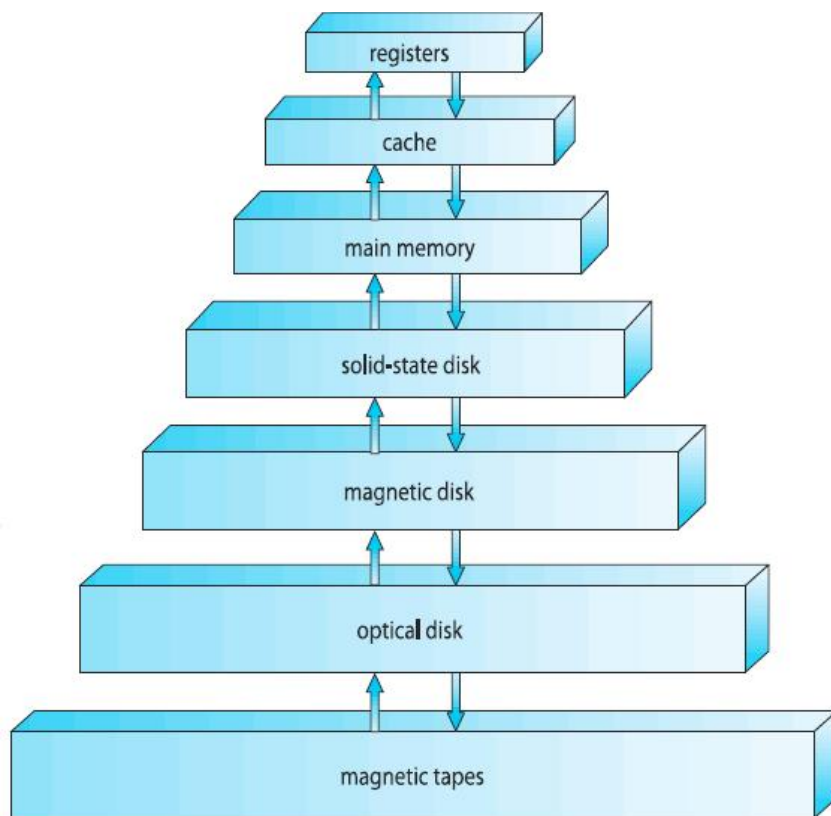
- 比磁盘速度快，普及
- 基于NAND Flash 的集成品



Compact flash (CF) & secure digital (SD) cards, a Sony memory stick, and a USB memory key.

3.4 存储结构- 层次结构

根据磁盘的速度、价格和易失性，表现为层次结构





Summary: Specific Units of Digital Information

IEC prefix		Representations			Customary prefix	
Name	Symbol	Base 2	Base 1024	Base 10	Name	Symbol
kibi	Ki	2^{10}	1024^1	$\approx 1.02 \times 10^3$	kilo	k, K
mebi	Mi	2^{20}	1024^2	$\approx 1.05 \times 10^6$	mega	M
gibi	Gi	2^{30}	1024^3	$\approx 1.07 \times 10^9$	giga	G
tebi	Ti	2^{40}	1024^4	$\approx 1.10 \times 10^1_2$	tera	T
pebi	Pi	2^{50}	1024^5	$\approx 1.13 \times 10^1_5$	peta	P
exbi	Ei	2^{60}	1024^6	$\approx 1.15 \times 10^1_8$	exa	E
zebi	Zi	2^{70}	1024^7	$\approx 1.18 \times 10^2_1$	zetta	Z
yobi	Yi	2^{80}	1024^8	$\approx 1.21 \times 10^2_4$	yotta	Y

- binary prefix = IEC prefix
- Source: http://en.wikipedia.org/wiki/Binary_prefix



第四节

计算机系统的体系结构



1. 单处理器系统 (Single Processor Systems)
2. 多处理器系统 (Multi Processor Systems)
3. 集群系统 (Cluster Systems)



系统中只有一个通用处理器（CPU），用来处理来自用户进程的指令

除了通用处理器，系统一般还包括其他专用处理器，如磁盘控制器、图形控制器等，**专用处理器不接受用户的指令，只接受系统的指令**

有的通用处理器集成了专用处理器，具有专用处理器功能



系统中有多处理器，又称为并联系统(parallel systems、tightly-coupled systems)，多个处理器共享一个内存
CPU之间通过共享内存来进行通讯

操作系统可以运行在某一个CPU上或多个CPU上

多处理器系统的优点

- 增加了吞吐量 (throughput)
- 方便扩展
- 增加了可靠性 (graceful degradation or fault tolerance)



1. 非对称处理器，又称异构多处理器， Asymmetric Multiprocessor(AMP)

- 处理器在结构上不同，一个处理器负责运行操作系统，其它处理器运行其他程序，**处理器之间有主从关系**

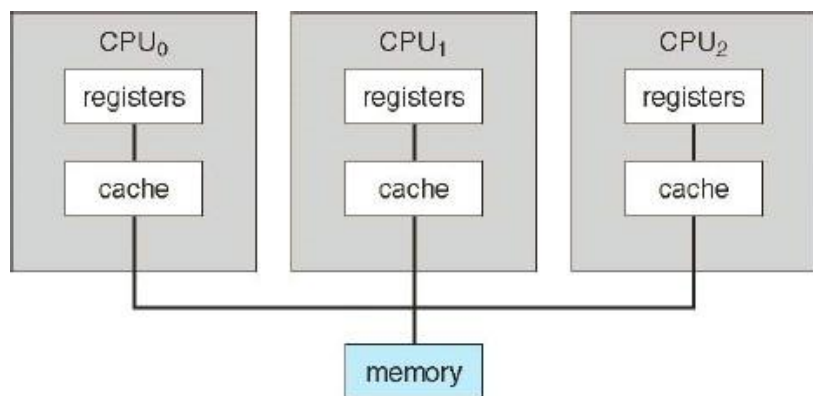
2. 对称处理器，又称同构多处理器 Symmetric Multiprocessor(SMP)

- 各处理器在结构上完全相同，操作系统可以运行在任何一个处理器上，**处理器之间没有主从关系。**

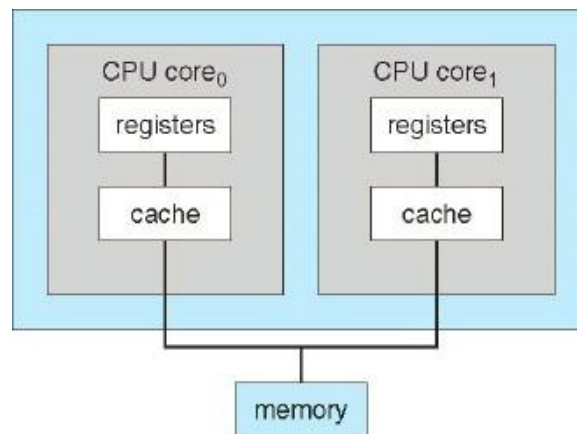
3. Multi-Core - 多核处理器

一个处理器上有多个CPU，也就是多个处理器集成到一个芯片上，可以是对称多处理器，也可以是非对称多处理器





Symmetric multiprocessing architecture

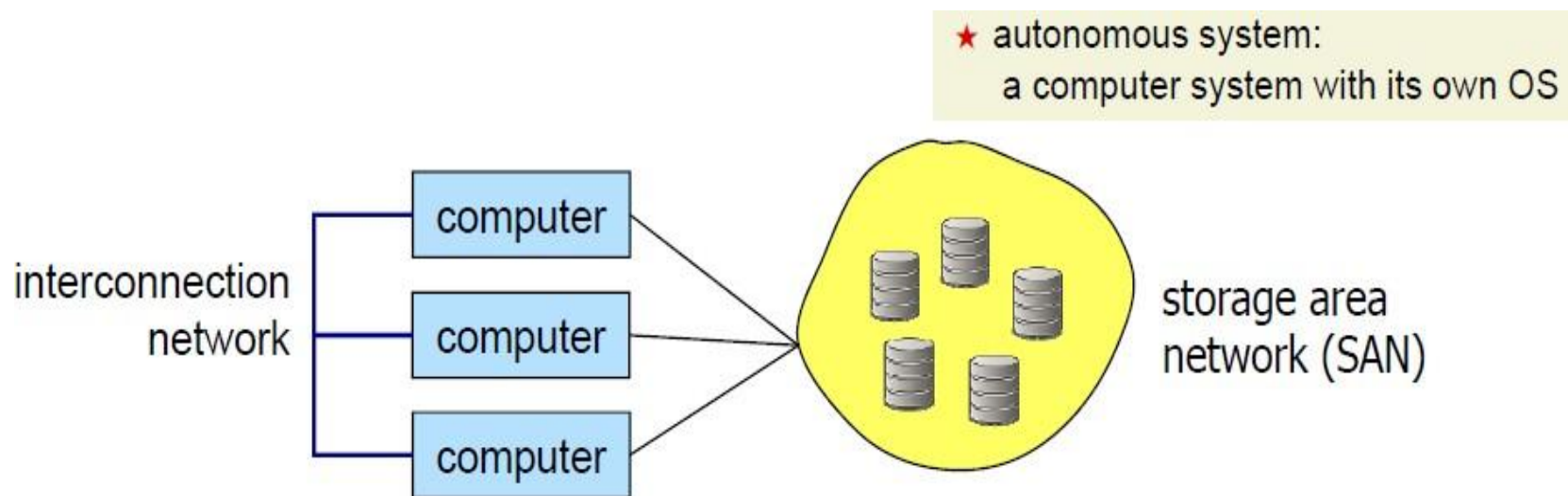


Dual-core microprocessor

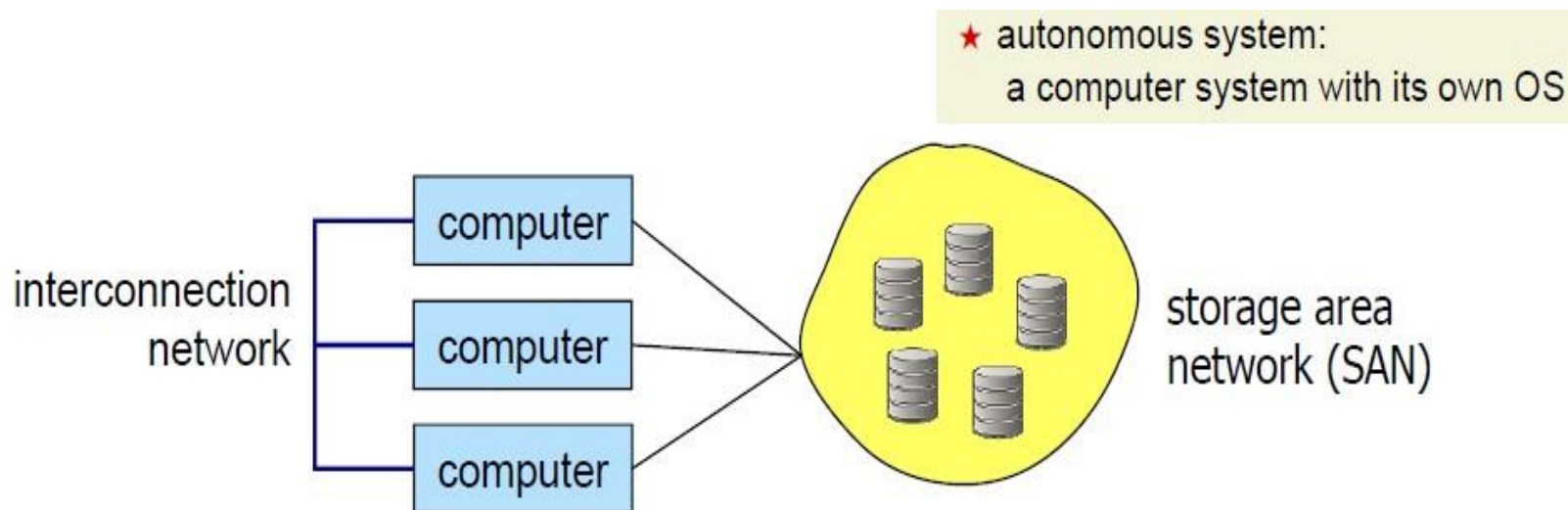
注意：多处理器环境下，不管多处理器还是多核处理器、每个处理器都需要有自己的寄存器和高速缓存

Why?

集群系统是一种由互相连接的计算机组成的并行分布式系统 (Parallel Distributed System)，并共享存储，可以作为单独、统一的计算资源来使用。



- 多个自治系统（autonomous system）协同工作。
- 自治系统指的是各计算机系统（节点）都有自己独立的操作系统的系统
- 一个集群系统可以看成是单个逻辑计算单元，它由多个节点通过网络连接组成





优点 - 高性能

计算机集群提供了 更快的处理速度，更大的存储容量，更好的数据完整性，更高的可靠性和更广泛的可用性资源

缺点 - 高费用

但是，需要更昂贵的实现和维护。相比一台计算机，这将导致更高的运行开销

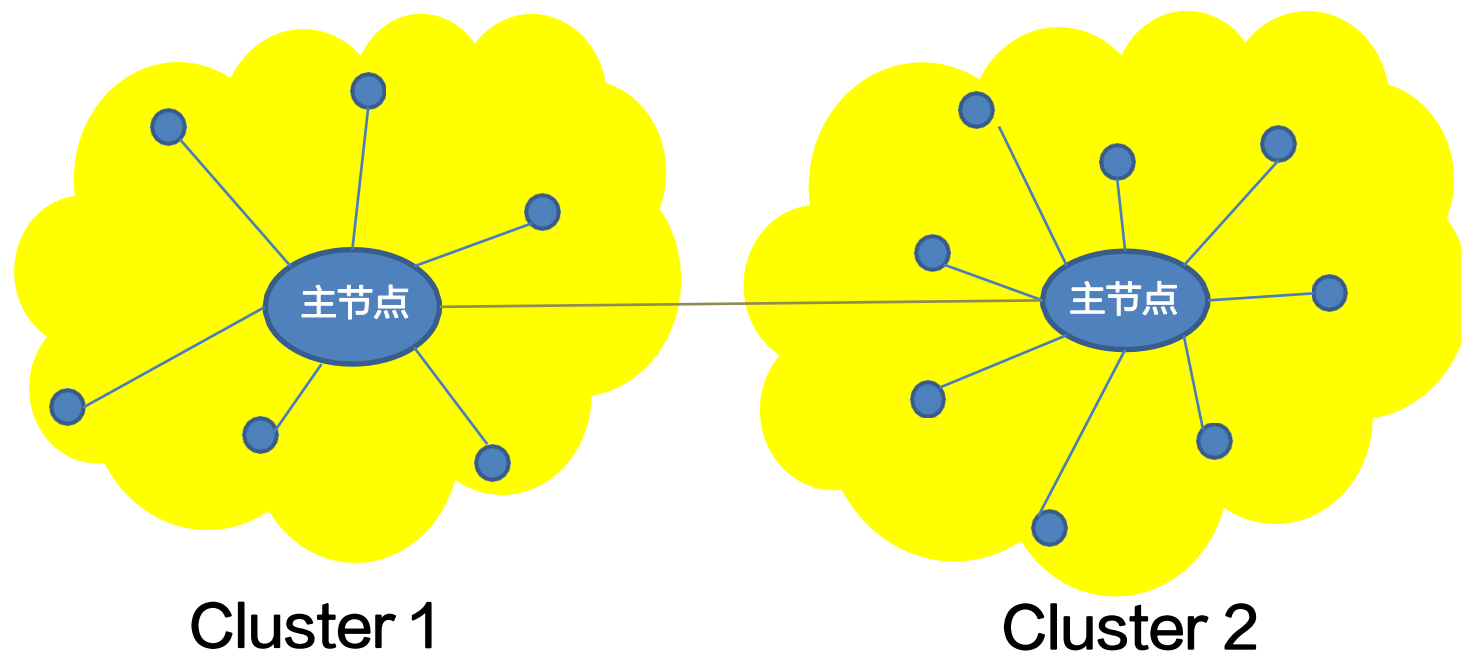


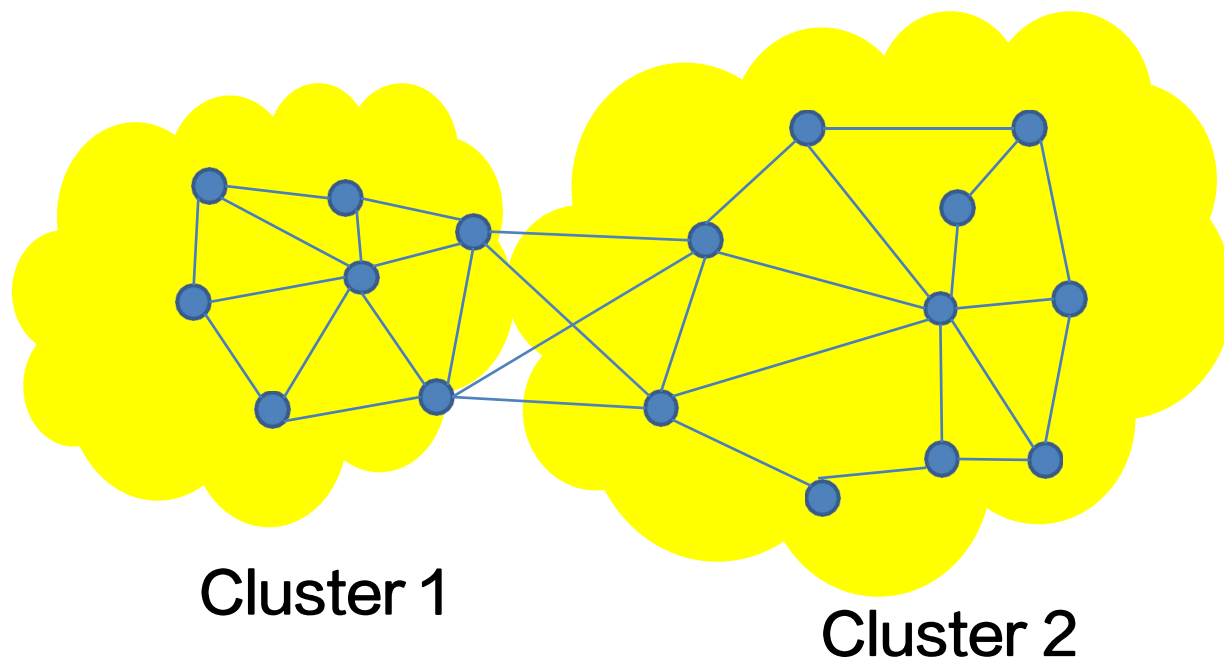
非对称集群 (Asymmetric Clustering) : 选出一个节点当主节点(header node), 即主机待机模式的节点, 有主从关系

Host standby node is a Header node,
Header node monitors all the other nodes

对称集群 (Symmetric Clustering) : 没有主节点, 全部节点地位相等, 没有主从关系

No host standby node,
All nodes are monitored each other







第五节

操作系统的结构及操作

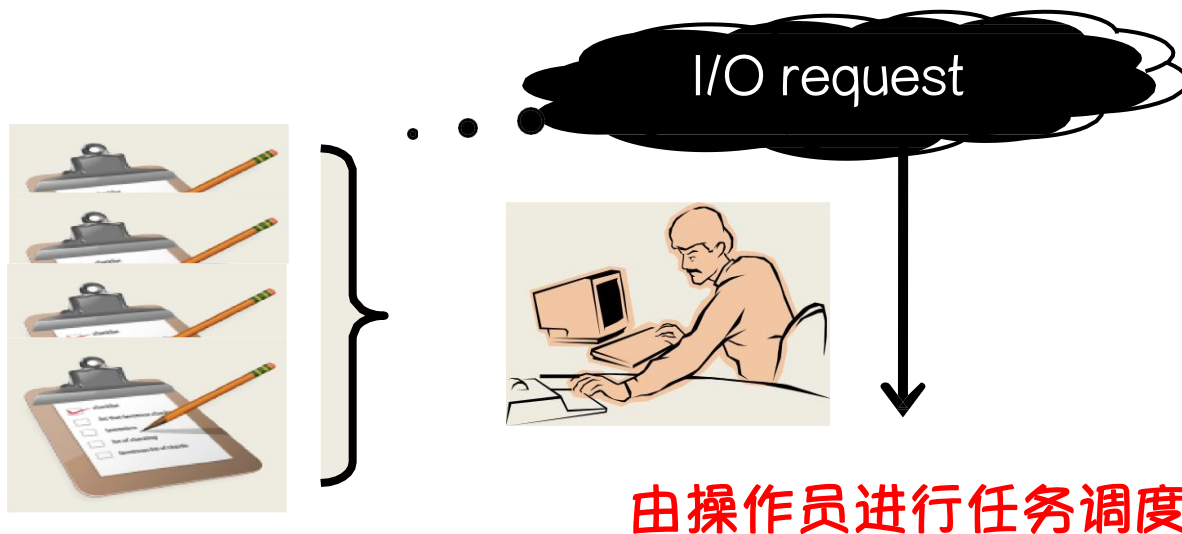


1. 操作系统的结构

- (1) 批处理系统 (batch processing system)
- (2) 多道程序系统 (multiprogramming system)
- (3) 分时系统, 又称多任务系统
(time sharing system, multitasking system)

2. 操作系统的操作

双重模式操作 (Dual Mode Operation)



当发生 I/O 操作时，由操作员调度另一个程序运行，从而提高CPU的使用率。



问：我们把操作员的工作内容写到操作系统里，由操作系统进行任务的调度，可不可以？

1. 当系统进行 I/O 操作时，调度器就会选择另一个任务并运行
2. 需要一个调度器，通过一定的调度机制选择任务并运行
3. 为保障CPU总有任务运行（提高CPU的使用率），多道程序系统必须把需要运行的多个任务载入到内存以被选择并运行

可能出现的问题？

Multiprogramming System



多用户交互式系统，需要对每个用户的响应时间应小于1秒钟

Q: 如何去确保响应时间，实现公平使用CPU?

A: 分时系统，又称多任务系统，Multitasking System, 又称 Time-Sharing System

给每个任务赋予一个给定的时间片（time slot, time slice）
CPU 在多任务之间相互切换，如20msec为单位切换另一个用户



两者目的不同

1. 多道程序系统 (Multiprogramming System) : 为提高CPU的使用率, 让CPU忙
2. 多任务系统(分时系统) - Multitasking System, Time-Sharing System : 让每个任务能公平使用CPU, 更体现了公平性

问：现在的操作系统是批处理系统，还是多道程序系统，还是多任务系统呀？



- 一个操作系统可以被多个用户、多个程序所共享
- 多个用户之间、多个程序之间可能相互影响
- 非法或不正确的操作会导致系统崩溃或破坏
- 如无限循环会导致其他程序无法运行，会导致系统失灵或崩溃

操作系统需要有一个保护机制

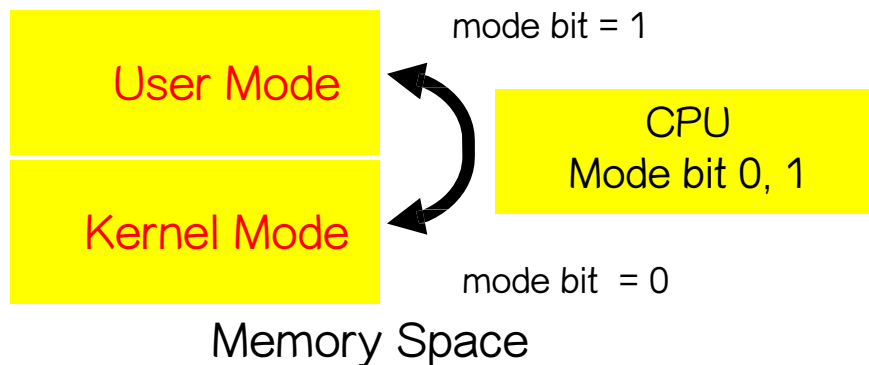
双重模式操作

Solution: Dual Mode Operation



- 为了确保操作系统的正常运行，**必须区分操作系统的代码和用户代码的执行**
- 大部分采用的方法是提供硬件支持的双重模式操作，即**用户模式和内核模式**
 - 用户模式时，**用户掌握计算机的控制权**，运行用户代码
 - 内核模式时，**操作系统掌握计算机的控制权**，运行系统代码
- 通过**模式位（Mode bit）来区分**即用户模式和内核模式

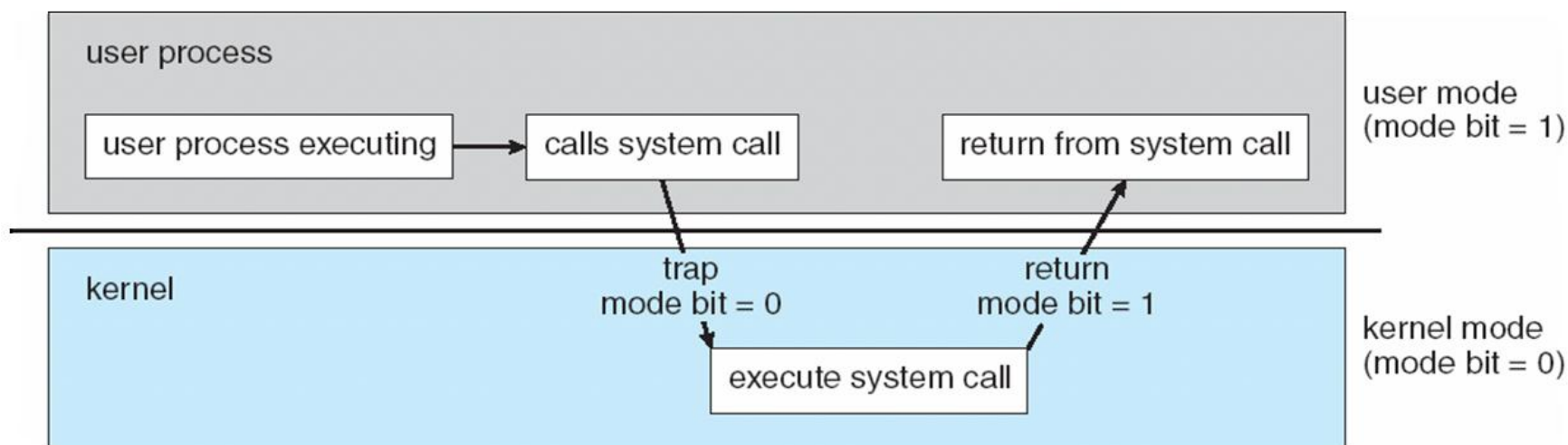
用户模式和内核模式在程序运行过程中相互切换



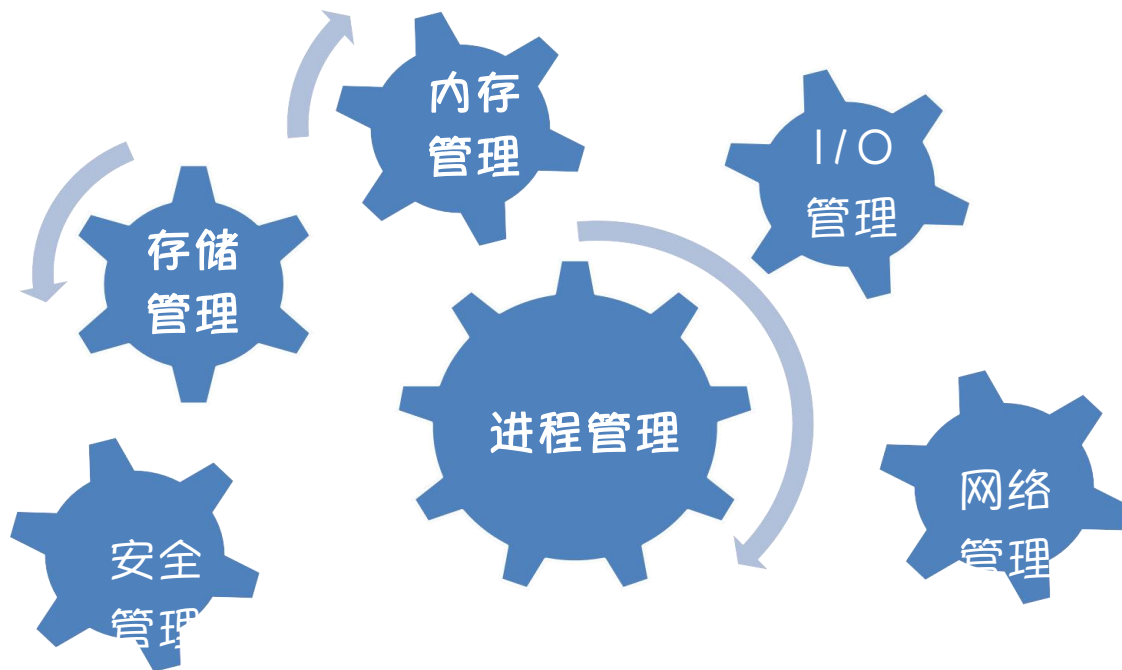
又称管理模式（Supervisor Mode），系统模式（System Mode），
特权模式（Privileged Mode），监督程序模式（Monitor Mode）

5.2 操作系统的操作

如系统调用 (System Call)，当用户程序调用系统调用函数的时候，操作系统的运行模式从用户模式转变成内核模式，如调用 `printf()` 函数



第六节 操作系统的管理内容





- 进程是运行中的程序，是系统的运行单元

程序(program) vs 进程(process)

程序是被动实体 (inactive)，而进程是活动实体(active)

1. 在一个系统中有无数个进程在同时运行，运行在一个或多个CPU上，进程之间通过复用CPU并发运行
2. 运行进程需要分配一定的资源，如CPU, 内存, I/O设备, 存储设备等。当进程结束时，也应收回已分配的资源，从而让进程有效使用这些资源



进程的管理活动包括以下内容：

1. 创建和删除进程
2. 挂起和重启进程
3. 需提供进程同步机制
4. 需提供进程通讯机制
5. 需提供死锁处理机制
6. 。 。 。



- 内存管理的主要目的就是提高内存的使用率，从而有效使用内存
- 管理内存中的数据的数据的存储、指令的运行
- 内存管理活动包括以下内容：
 - 当有内存有空闲空间时，决定哪些进程可以载入内存、载入到哪里等
 - 记录内存的哪些部分正在使用以及被谁使用
 - 根据需要分配和释放内存空间，即分配的方式



为了便于使用计算机系统，操作系统提供了统一的逻辑信息存储观点，即文件系统

1. 操作系统对**存储设备的物理属性进行了抽象的定义**，即文件，它是存储的逻辑单元。
2. 计算机可以在**多种类型的物理介质**上存储信息
 - 每种介质通过一个设备来控制，如磁盘驱动器、磁带驱动器
 - 每个介质有不同的访问速度、容量、数据传输率和访问方法
3. **文件通常组成目录**以方便使用
4. 多用户访问文件时，**需要控制权限问题**



文件系统管理活动包括以下内容:

1. 创建和删除文件
2. 创建和删除目录来组织文件
3. 提供操作文件和目录的原语
4. 将文件映射到二级存储设备上
5. 在稳定存储介质上备份文件
6. . . .



大容量存储系统(Mass-Storage System)

一般指的二级存储设备，如硬盘，它的管理活动包括以下内容：

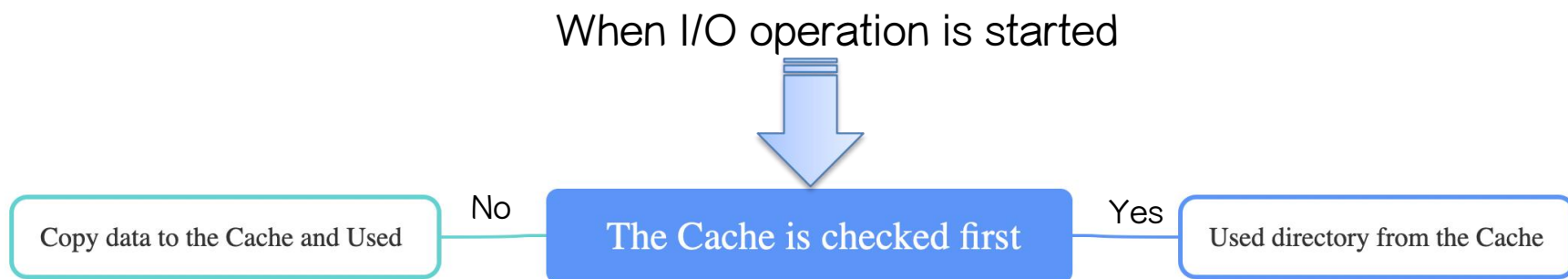
1. 空闲空间的管理
2. 存储空间的分配
3. 硬盘的调度



Storage hierarchy

Level	1	2	3	4	5
Name	registers	cache	main memory	solid state disk	magnetic disk
Typical size	< 1 KB	< 16MB	< 64GB	< 1 TB	< 10 TB
Implementation technology	custom memory with multiple ports CMOS	on-chip or off-chip CMOS SRAM	CMOS SRAM	flash memory	magnetic disk
Access time (ns)	0.25 - 0.5	0.5 - 25	80 - 250	25,000 - 50,000	5,000,000
Bandwidth (MB/sec)	20,000 - 100,000	5,000 - 10,000	1,000 - 5,000	500	20 - 150
Managed by	compiler	hardware	operating system	operating system	operating system
Backed by	cache	main memory	disk	disk	disk or tape

1. 高速缓存是临时存储设备，一般设置在**高速设备与低速设备之间**
2. 当高速设备（内存）从低速设备（磁盘）上读取数据时，会把数据临时**复制到高速缓存上**
3. **如I/O操作开始进行的时候，操作系统先到高速缓存里查找**





Questions:

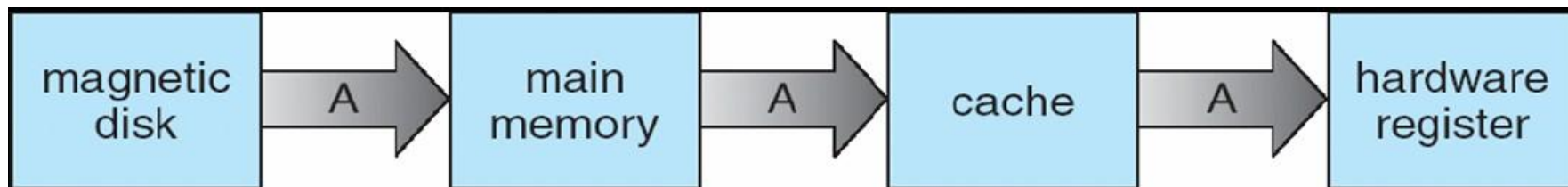
1. 缓存能代替存储设备吗?
2. 而且缓存容量一般很小, 缓存满了怎么办?
3. 一种解决办法是置换缓存里的内容, 怎么置换?

? 缓存管理

? 怎么决定缓存大小

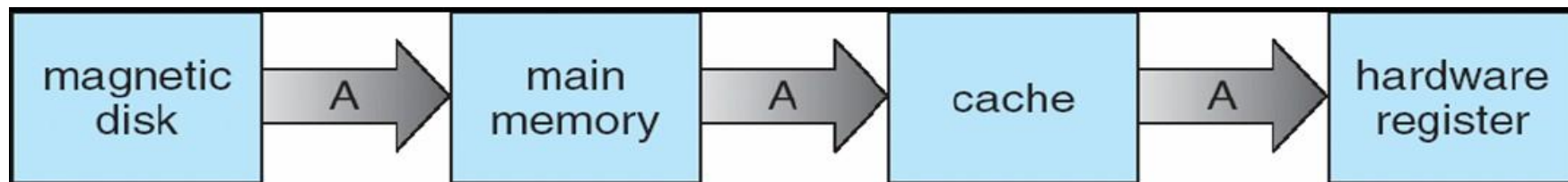
? 缓存内容置换

举例：把整数A 从磁盘移动到寄存器



把磁盘上的 $A=1$ 读到内存，然后加一，把结果存储到硬盘的流程

1. 先从磁盘上读取 A 到内存上
2. 复制 A 到高速缓存上
3. 为了加一，复制 A 到 CPU 的寄存器上
4. 加一后，把结果写入磁盘

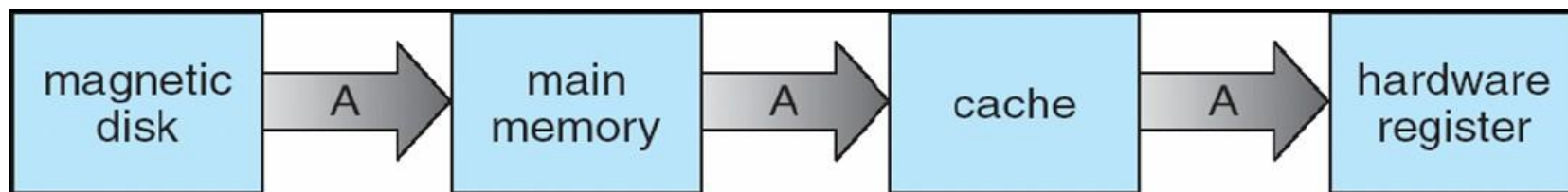


加一操作结束后(写入磁盘之前), 各个存储器内的 A 的值会不同。

当前状态, 磁盘: 1, 内存: 1, 缓存: 1, 寄存器: 2

假设, 在多任务环境下, CPU 切换到另一个进程读取 A 的值的时候, 第一个先到高速缓存上读取, 而缓存上的 A 的值不是最新值。

在多处理器环境下表现为更复杂, 因为每个处理器都有自己的高速缓存和寄存器



那么, 在多任务环境下、多处理器环境下, 分布式环境下必须要保证数据的一致性

////一致性问題Coherency Solution: 用硬件、软件来解决



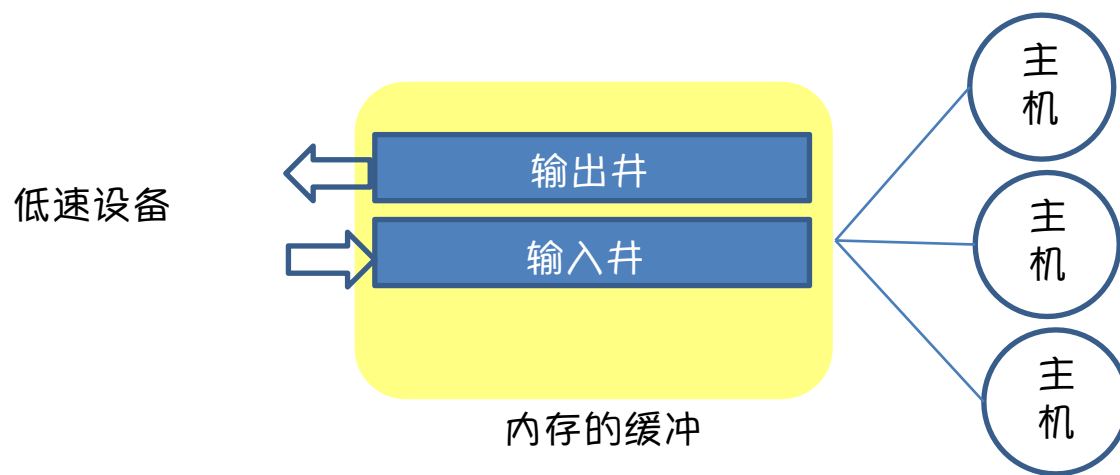
I/O子系统的目的是针对用户隐藏具体硬件设备的特性，它包括以下几个部分

1. 一个包括缓冲(buffer)、高速缓存(cache)和假脱机(spooling)的内存管理部分
2. 通用设备驱动器接口
3. 特定硬件设备的驱动程序

Notice ,

In Linux/Unix, a device is accessed through the use of file management interface

1. 缓冲：为传输数据，暂时存储数据
2. 缓存：为性能提高，暂时存储数据
3. 假脱机：是关于低速字符设备与计算机主机交换信息的一种技术，
又称外部设备联机并行操作。可以提高I/O速度、设备并没有单独
分配给任何一个任务、通过实现虚拟设备可以共享设备





第七节

其他计算机系统



- 分布式系统(Distributed Systems)将物理上分开的、可能异构的计算机系统通过网络连接在一起，为用户提供系统所维护的各种资源的计算机集合
- 其他专用系统，如实时系统(real-time system)，嵌入式系统(embedded system)，多媒体系统，手持系统
- 实时系统指的是系统中的任务有时间节点(截止时间，deadline)的系统，如军事设备、医疗设备上使用的系统



1. 客户机/服务器计算(C/S Computing)

- 计算服务器系统
- 文件服务器系统



存在瓶颈
bottle-neck problem

2. 基于Web 计算(Web-based Computing)

- 负载均衡器(load balancer)



2. 点对点计算(Peer-to-Peer Computing)

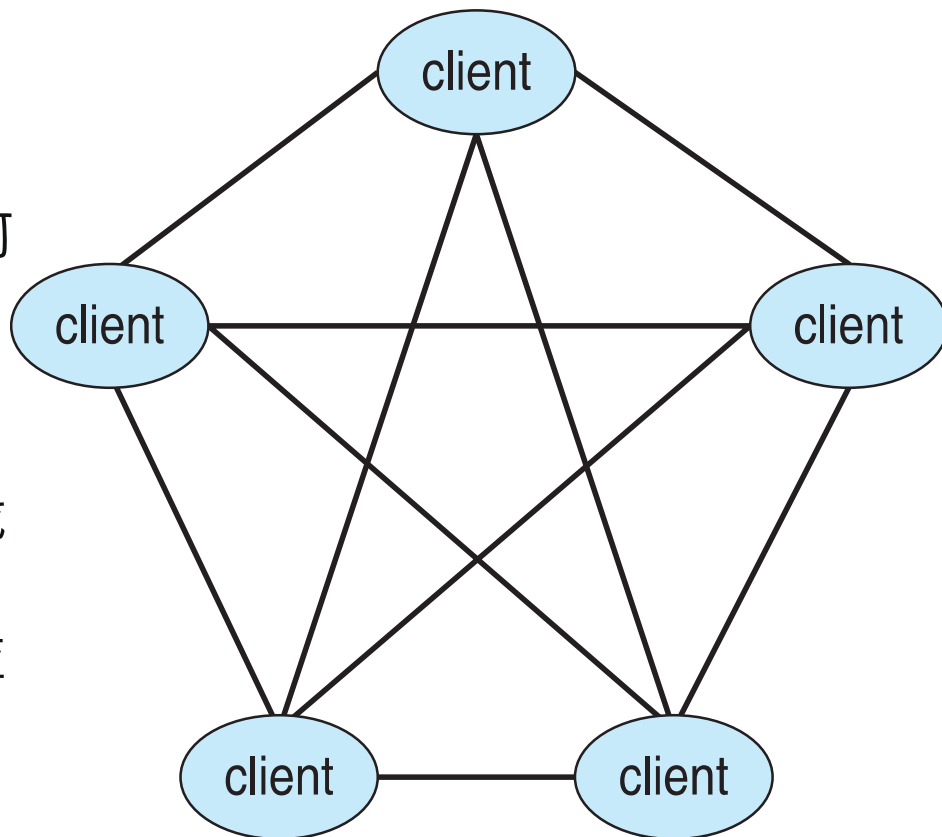
- 又是客户机、又是服务器
- 复杂

负载均衡器
load-balancer

P2P不区分客户端和服务端

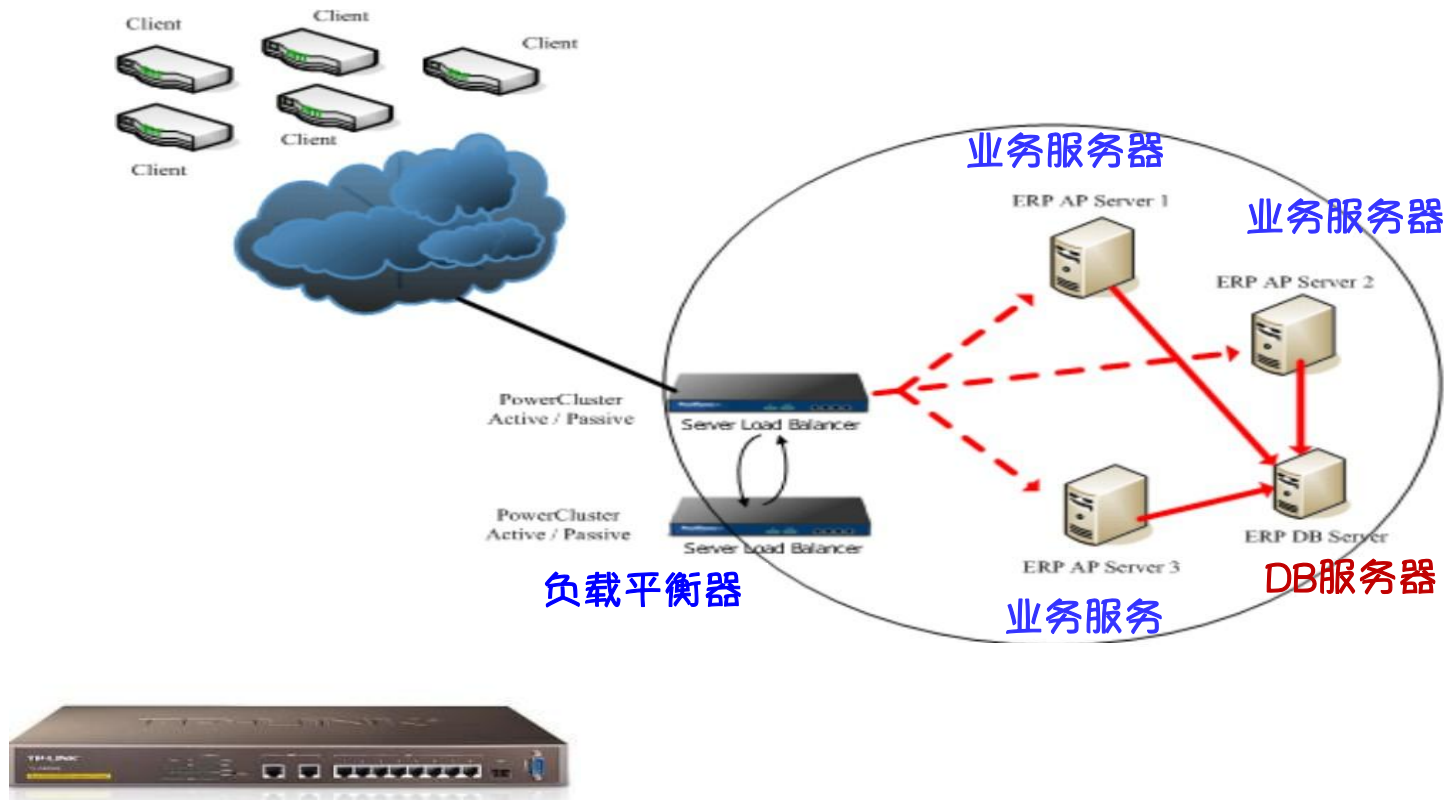
所有节点都被视为对等节点，每个可以作为客户端，服务器或两者节点必须加入P2P网络

- 将其服务注册到网络上的中央查找服务，或者
- 广播服务请求并通过发现协议响应服务请求
- 例子包括Napster和Gnutella, VoIP (VoIP)如Skype



8. 负载均衡器 (load balancer)

用户

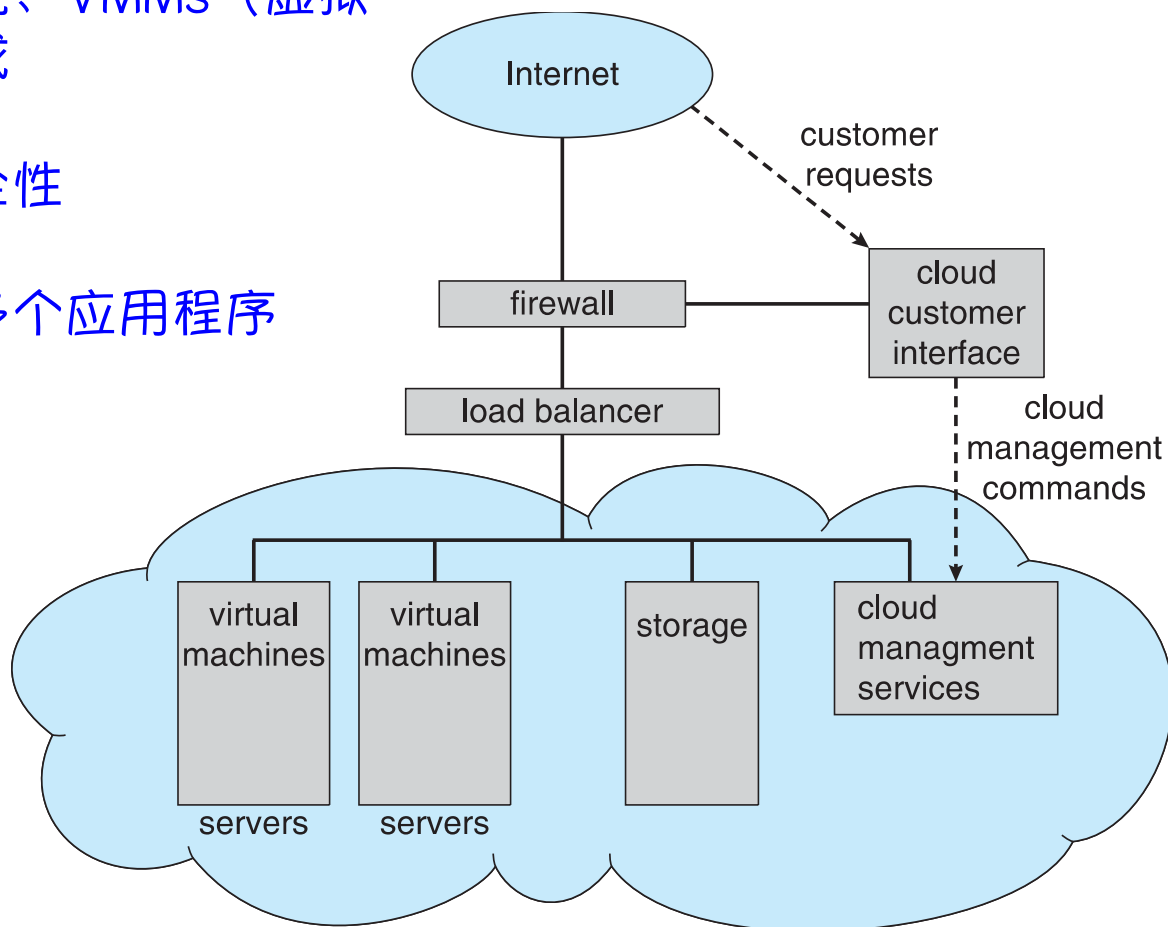


- 通过网络提供计算、存储、甚至应用作为服务
 - 基于虚拟化的虚拟化逻辑扩展
 - Amazon EC2有数千台服务器，数百万个虚拟机，互联网上可用的PBs存储，根据使用情况付费
-
- ① 公共云 (Public) : 任何人都可以通过互联网支付
 - ② 私有云 (Private) : 由公司运营，供公司自己使用
 - ③ 混合云 (Hybrid) : 包括公共和私有云组件
-
- ① 软件即服务(SaaS) : 通过互联网提供的一个或多个应用程序(如文体编辑器)
 - ② 平台即服务(PaaS) : 可以通过Internet(即数据库服务器)使用的软件栈。
 - ③ 基础设施即服务(IaaS) : 通过互联网提供的服务器或存储(即用于备份的存储)

云计算环境由传统操作系统、VMMs（虚拟机管理）和云管理工具组成

网络连接需要防火墙等安全性

负载均衡器将流量分散到多个应用程序



Q&A