

操作系统原理

第一章 绪论



- **操作系统与计算机体系结构的关系**
- **操作系统的形成与发展**
- **操作系统的定义**
- **操作系统的资源管理功能**

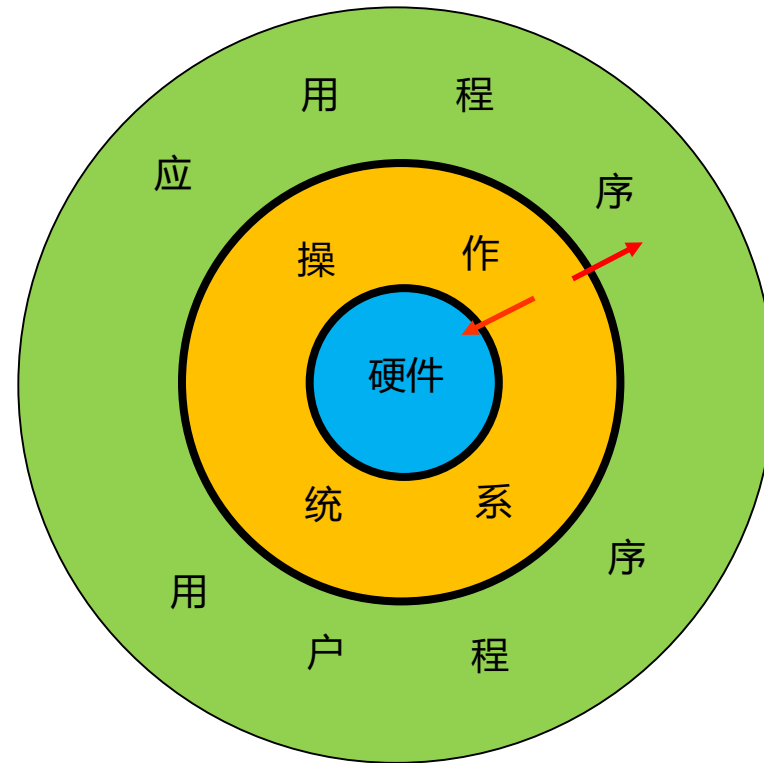
操作系统在计算系统中的地位

■ 与硬件的关系

- 控制CPU的工作
- 访问存储器
- 中断处理、设备驱动
- 高效利用硬件资源

■ 与用户及其他软件的关系

- 控制、管理
- 提供方便的用户界面
- 提供优质的服务



计算机系统的组成

■ 硬件环境对OS的限制

- 提供OS运行环境
- 限制了OS的功能实现

■ 用户和上层软件的需求

- 图形界面/文本界面/不需要交互手段
- 单任务/多任务
- 应用间的消息通讯与交互

■ 思考：是硬件决定了软件，还是反过来？

- 操作系统与计算机体系结构的关系
- 操作系统的形成与发展
- 操作系统的定义
- 操作系统的资源管理功能

1. 手工操作阶段

■ 特点

- 人工操作 (1946~1960)
- 作业独占计算机
- 作业串行执行



■ 出现的问题

- 随着计算机性能的提高, 人的速度 vs. 计算机的速度

| 机器速度 | 程序执行 (计算) 时间 | 人工操作时间 | 人工时间/计算时间 |
|--------|--------------|--------|-----------|
| 1万次/秒 | 1小时 | 3分钟 | 1: 20 |
| 60万次/秒 | 1分钟 | 3分钟 | 3: 1 |

2. 批处理阶段

■ 原理

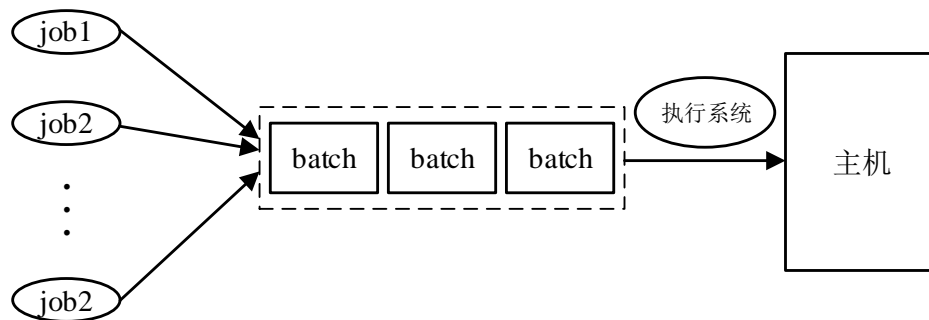
- 将多个程序打包 (batch)
- 将打好的程序包投入运行

■ 特点

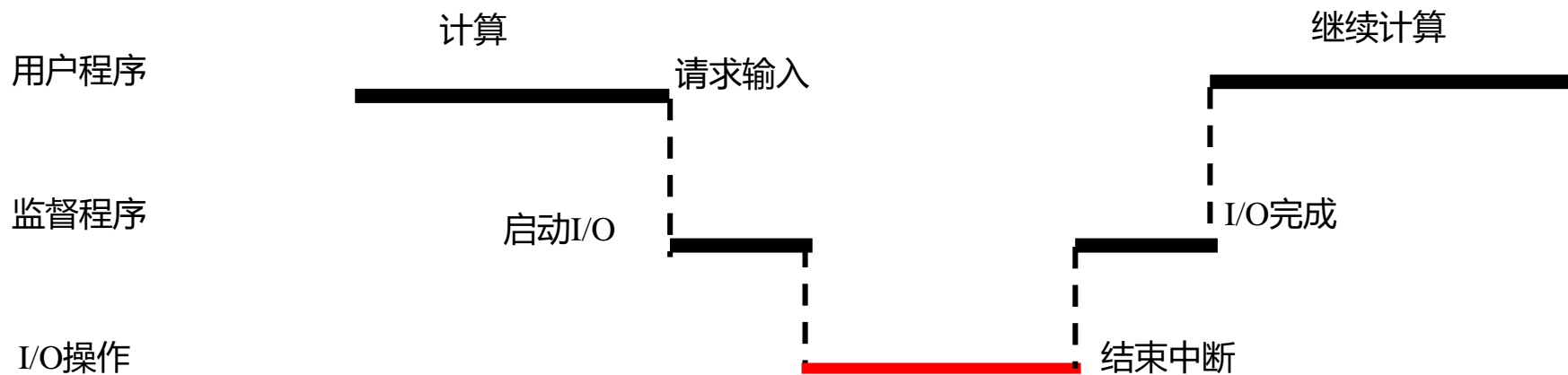
- 程序包中的上一个程序执行完毕后，下一个程序自动投入运行
- 程序包中的程序执行切换无需人工干预

■ 出现的问题

- (电子) CPU速度提高 \gg I/O速度的提高

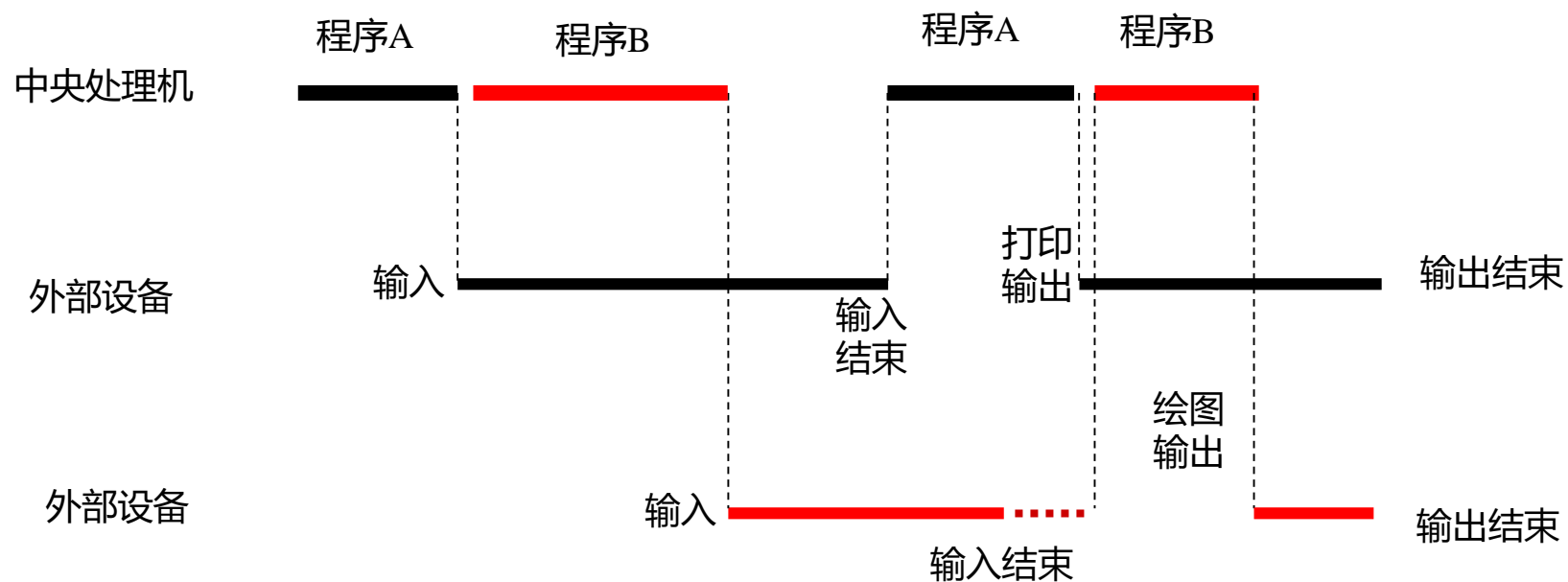


■ 单道程序程序的工作情况



单道程序工作

■ 多道程序工作的情况



多道程序工作

■ 什么是多道程序设计（multi-programming）技术

- 在计算机主存中同时存放几道相互独立的程序。这些程序在执行系统控制之下，相互穿插地运行。当某道程序因某种原因不能继续运行下去时(如等待外部设备传输数据)，管理程序便将另一道程序投入运行。

■ 多道运行的特征

- 多道（内存中存放多个相互独立的程序）
- 宏观上并行（相互穿插运行）
- 微观上串行（CPU上永远只能执行单个程序）

■ 物质基础：足够容量的内存

■ 什么是执行系统

- 借助于通道与中断技术，由主机控制I/O工作。原有的监督程序不仅要负责调度作业自动地运行，而且还要提供I/O控制功能。它常驻主存，称为执行系统。

■ 特点

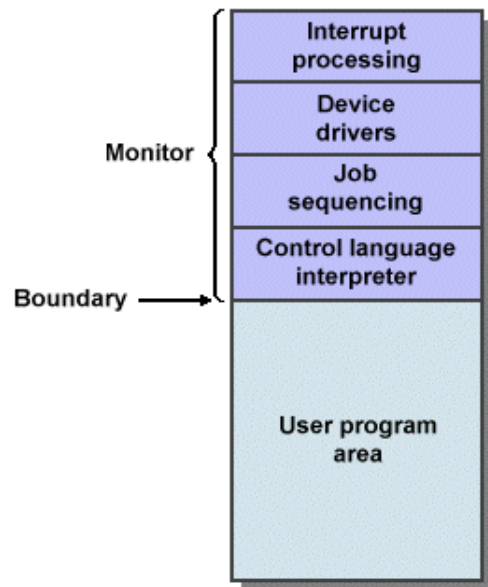
- 主机、外设并行操作；采用多道程序设计技术调度作业运行

■ 基本功能

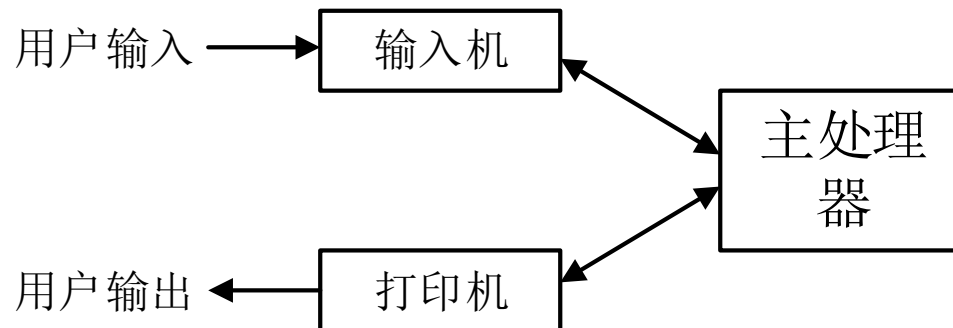
- I/O控制功能
- 调度

■ 问题

- 主机与外设的并行是有限度的，还依赖于程序运行的特征。



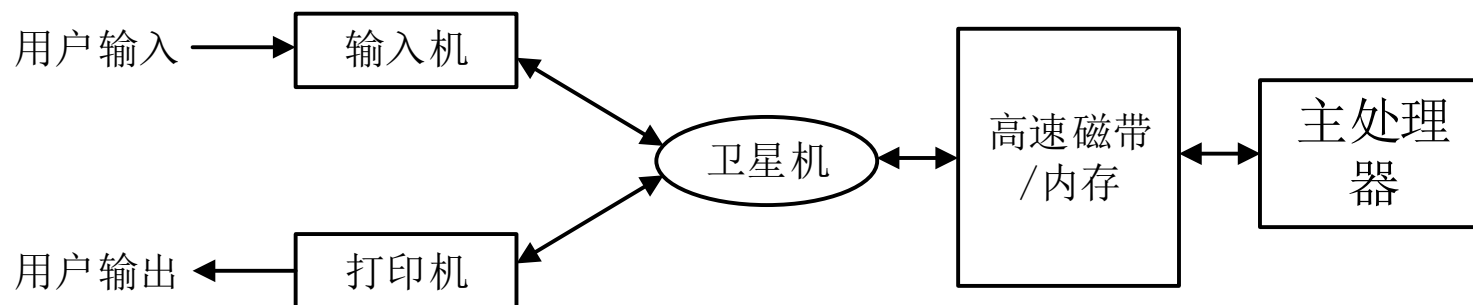
■ 作业的输入/输出由CPU来处理



■ 问题

□ 慢速I/O设备严重浪费宝贵的主处理器时间

■ 作业的输入/输出由卫星机来处理



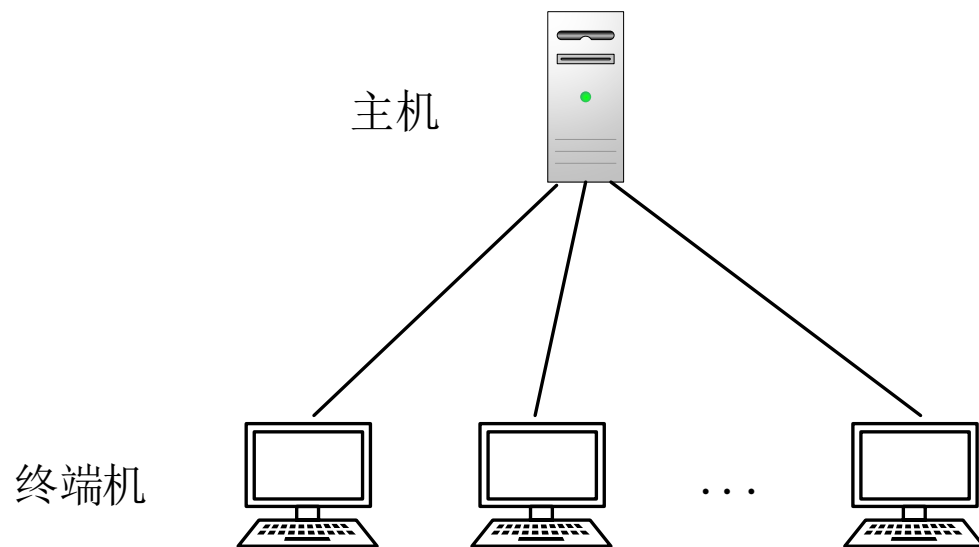
■ 特点

- 引入卫星机（采用廉价处理器，如MCU）来负责I/O，主处理器只与高速存储设备交互
- 今天广泛使用的DMA（Direct Memory Access）技术原型

3. 分时系统阶段

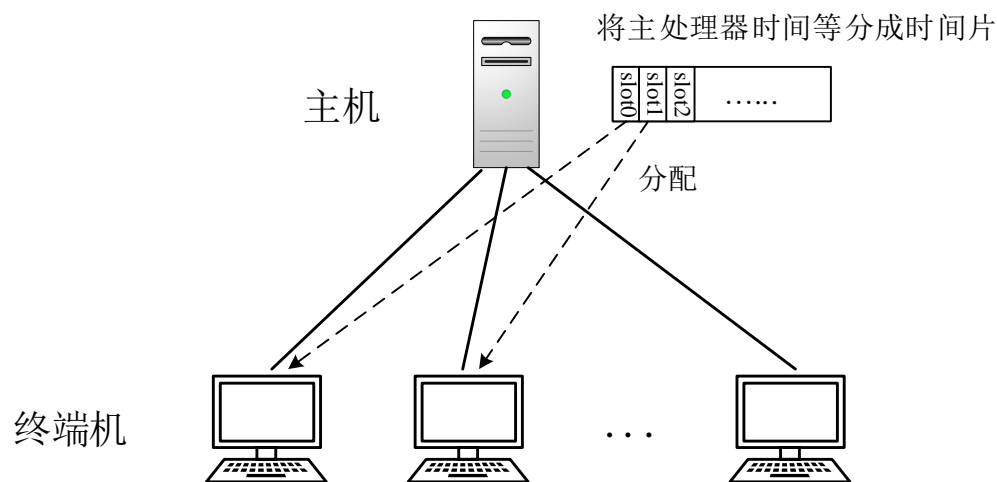
■ 出现的时代背景

- 计算机非常昂贵 (1960~1970)
- 多个用户需要共享单个计算机资源



■ 什么是分时 (time-sharing) 技术?

- 所谓分时技术, 是把处理机时间划分成**很短的时间片**(如几百毫秒)**轮流**地分配给各个应用程序使用, 如果某个程序在分配的时间片用完之前计算还未完成, 该程序就**暂停执行**, 等待下一次获得时间片后再**继续**计算。



- **实时 (real-time) 系统**是指系统能及时响应外部事件的请求，在规定的时间内 (deadline) 范围内完成对该事件的处理，并控制实时任务协调一致地运行。
- **硬实时 (hard real-time) 系统**
 - 若实时约束不被满足，则会导致灾难性后果
- **软实时 (soft real-time) 系统**
 - 若实时约束不被满足，则会导致服务质量的降级

5. 衍生和多样化阶段

- 嵌入式操作系统
- 个人计算机操作系统
- 网络操作系统
- 分布式操作系统

■ 资源受限环境下工作的操作系统

- 计算能力受限
- 能耗受限
- 网络连接情况受限等

■ 典型应用场景

- 各类（小）电器设备，如掌上PDA、手机、微波炉、数码相机、自动售货机、工业自动化仪表与医疗仪器等。

■ 面向个人电脑（PC）设计的操作系统

- 算力一般
- 外设丰富
- 与人直接交互

■ 主要特点

- 单用户多任务环境
- 界面友好

■ 支持网络通讯的操作系统

- 支持网络通讯协议
- 通过网络实现资源的远程访问

■ 主要特点

- 支撑各类网络应用（如Web、FTP等）
- 安全性得到加强

■ 分布式系统

- 分布式系统是指多个分散的处理单元经互连网络连接而形成的系统，其中每个处理单元既具有高度自治性又相互协同，能在系统范围内实现资源管理、任务动态分配，并能并行地运行分布式程序。

■ 分布式操作系统的特点

- 资源协调和调度
- 单一系统镜像 (SSI)
- 其他

- 操作系统与计算机体系结构的关系
- 操作系统的形成与发展
- **操作系统的定义**
- 操作系统的资源管理功能

■ 操作系统的定义

- 操作系统（OS）是一个大型的软件系统，它负责计算机系统软、硬件资源的分配，控制和协调并发活动，提供给用户和其他软件方便的接口和环境，使用户获得良好的工作环境。

■ 操作系统的特点

- 并发（concurrency）
- 共享（sharing）
- 虚拟（virtualization）
- 不确定

|| 并发 (concurrency)

- 并发是指两个或多个事件在同一时间间隔内发生。
- 并发 (concurrency) vs. 并行 (parallel)
 - 小明一边吃雪糕一边唱歌 (唱一句再吃一口) 。—— 并发
 - 小明在走路, 而与此同时, 小红在吃雪糕。——并行
- 并行往往 (特别是当两件事情互相关联的时候) 意味着并发, 而并发不一定意味着并行。
- 操作系统需要支持、并提供机构来协调并发的事件或进程

共享 (sharing)

■ 共享是指系统中的资源可供多个并发执行的进程共同使用。

■ 两种资源共享方式

➤ 时分复用 (time multiplexing)

◆ 例如多道系统中，两个作业共享CPU。

➤ 空分复用 (space multiplexing)

◆ 在内存中存放多道程序。

■ 操作系统需要实现、管理和协调进程对资源的共享

■ 虚拟是指把一个物理上的实体变为若干个逻辑上的对应物。

■ 例如

- 多道技术将程序装入内存中不同的地址空间，使得对于每个程序而言，以为它在独占内存资源。
- 分时技术通过将CPU时间分片，然后将时间片尽量公平地分配给多个进程的办法，使得对于每个进程而言，以为它在“独占”CPU。

■ 操作系统的重要任务之一是实现逻辑到物理的转换

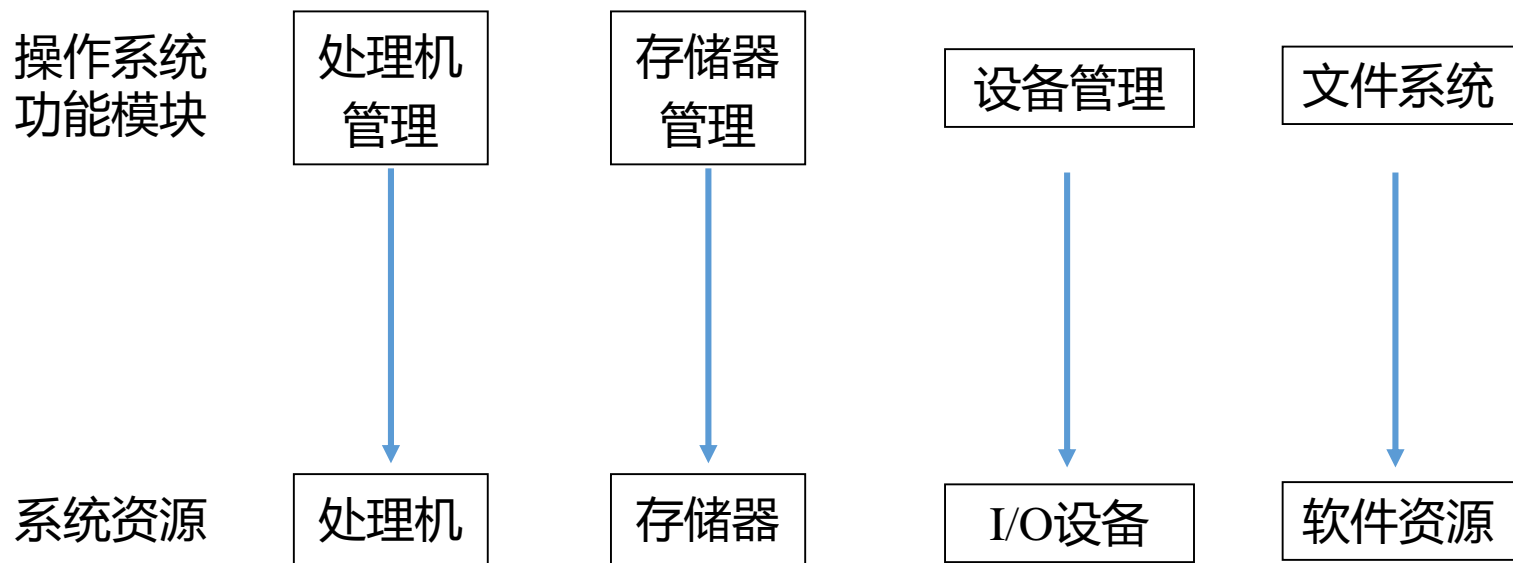
- 不确定性表现为多个作业的执行顺序和每个作业的执行时间是不确定的。
- 不确定性也被称为异步性 (asynchrony) 。
- 操作系统需要协调异步活动，同时提高系统资源利用效率

- 操作系统与计算机体系结构的关系
- 操作系统的形成与发展
- 操作系统的定义
- 操作系统的资源管理功能

操作系统的资源管理功能

- 1.处理机管理
- 2.存储器管理
- 3.设备管理
- 4.信息管理

资源管理功能和资源的对应关系



■ 重点概念

- 多道程序设计
- 分时技术
- 联机、脱机批处理
- 操作系统的定义和特点

■ 理解

- 存储程序式计算机的结构和特点
- 操作系统的演化历史和分类
- 操作系统的基本功能