

普通高等院校计算机类专业系列教材

*Computer Operating Systems*

# 计算机操作系统

孙雅如 房鼎益 丁剑锋 编著

西安电子科技大学出版社  
<http://www.xduph.com>



普通高等院校计算机类专业系列教材

# 计算机操作系统

Computer Operating Systems

孙雅如 房鼎益 丁剑锋 编著

西安电子科技大学出版社

2003

## 内 容 简 介

本书介绍了现代计算机系统最重要的系统软件——操作系统。全书着重讲述计算机操作系统的基本概念、基本原理、设计方法、实现技术及其最新发展。全书共分7章，第1章介绍了操作系统的基本知识；第2~5章分别讲述了处理机管理、存储器管理、设备管理和文件管理；第6章对操作系统的—个实例——UNIX System V版本进行了剖析；第7章对多处理机操作系统、网络操作系统和分布式操作系统进行了介绍。

本书可以作为高等学校计算机专业教材，也可以作为有关科技人员学习操作系统的参考书。

★ 本书配有电子教案，需要者可与出版社联系，免费索取。

### 图书在版编目 (CIP) 数据

计算机操作系统 / 孙雅如等编著.

—西安: 西安电子科技大学出版社, 2003.7

(普通高等院校计算机类专业系列教材)

ISBN 7-5606-1249-0

I. 计… II. 孙… III. 操作系统—高等学校—教材 IV. TP316

中国版本图书馆 CIP 数据核字 (2003) 第 039856 号

策 划 臧延新

责任编辑 臧延新

出版发行 西安电子科技大学出版社 (西安市太白南路2号)

电 话 (029)8242885 8201467 邮 编 710071

http://www.xduph.com

E-mail: xdupfxb@pub.xaonline.com

经 销 新华书店

印 刷 陕西画报社印刷厂

版 次 2003年7月第1版 2003年7月第1次印刷

开 本 787毫米×1092毫米 1/16 印张 14.5

字 数 338千字

印 数 1~4000册

定 价 15.00元

ISBN 7-5606-1249-0 / TP·0657(课)

**XDUP 1520001 - 1**

\*\*\* 如有印装问题可调换 \*\*\*

# 普通高等院校计算机类专业系列教材

## 编审专家委员会名单

**主任委员：**冯博琴（陕西省计算机教育学会理事长，  
西安交通大学计算机教学实验中心主任，教授）

**副主任委员：**陈建铎（陕西省计算机教育学会副理事长，  
西安石油学院计算机系教授）

**李伟华**（陕西省计算机教育学会副理事长，  
西北工业大学计算机系副主任，教授）

**武波**（陕西省计算机教育学会副理事长，  
西安电子科技大学计算机学院副院长，教授）

**李荣才**（西安电子科技大学出版社总编辑，教授）

**委员：**（按姓氏笔划排列）

巨永锋（长安大学信息工程学院副院长，教授）

冯德民（陕西师范大学计算机科学学院院长，教授）

石美红（西安工程科技学院信息控制系教授）

朱明放（陕西理工学院计算机系副主任，副教授）

何东健（西北农林科技大学信息工程学院院长，教授）

陈桦（陕西科技大学计算机与信息科学系主任，教授）

李长河（西安理工大学计算机科学与工程系主任，副教授）

李晋惠（西安工业学院计算机系副主任，副教授）

李银兴（宝鸡文理学院计算机系副主任，副教授）

张俊兰（延安大学计算机系教授）

孟东升（西安石油学院计算机系副主任，副教授）

赵文静（西安建筑科技大学信息与控制工程学院副院长，教授）

耿国华（西北大学软件开发中心主任，教授）

龚尚福（西安科技学院计算机系主任，教授）

**项目策划** 陈宇光 马乐惠

**策 划** 云立实 马武装 臧延新 马晓娟

**电子教案** 马武装

# 前 言

本书在广泛参考了国内外有关计算机操作系统教材和文献的基础上,总结了作者多年的教学实践经验,对计算机操作系统中庞杂、抽象而难理解的内容进行了梳理与取舍,在取材上保持了操作系统的精华与核心,着重把最基本、最实用的内容包含进来,力求简洁明了。

本书通过对操作系统的基本理论、设计方法和实现技术的分析,使读者对操作系统有一个系统的、全面的整体概念;针对操作系统中出现的新理论、新技术和新的设计思想所做的介绍,使读者对操作系统新的研究方向有一定的了解。

本书第1章对操作系统的梗概做了介绍;第2~5章讲述了资源管理,包括处理机管理、存储器管理、设备管理和文件管理,这是操作系统的主要内容;第6章从进程管理、文件管理、存储管理和设备管理等方面对UNIX操作系统进行了剖析;第7章属于高级操作系统部分,介绍了多处理机操作系统、网络操作系统和分布式操作系统。

本书配有相应的教学软件,通过对操作系统内部活动的动态演示,使读者在掌握操作系统基本理论与实现方法的同时,更加深入地了解操作系统内部结构与系统的活动过程,从而加深对操作系统抽象理论的理解。需要该软件的读者,可直接与作者联系。

由于时间仓促及作者水平所限,书中难免有错误与不妥之处,恳请读者批评指正。

编著者

2003年4月

# 目 录

<b>第 1 章 引论</b> .....	1	2.4.1 概述 .....	48
1.1 操作系统的基本概念 .....	1	2.4.2 进程调度的实现 .....	49
1.1.1 操作系统的定义 .....	1	2.4.3 进程调度算法 .....	50
1.1.2 操作系统在计算机系统中的作用 .....	2	2.5 线程 .....	52
1.1.3 研究操作系统的几种观点 .....	2	2.5.1 线程的引入 .....	52
1.2 操作系统的发展 .....	4	2.5.2 线程的类型 .....	53
1.2.1 问题的提出 .....	4	2.5.3 线程与进程的比较 .....	55
1.2.2 解决办法 .....	5	2.6 死锁 .....	56
1.2.3 基本操作系统类型 .....	6	2.6.1 死锁定义 .....	56
1.2.4 新型操作系统 .....	8	2.6.2 死锁的类型 .....	56
1.3 操作系统的特征与功能 .....	9	2.6.3 死锁发生的条件 .....	58
1.3.1 操作系统的特征 .....	9	2.6.4 解决死锁的方法 .....	58
1.3.2 操作系统的功能 .....	10	2.7 作业管理 .....	64
1.4 操作系统的结构设计 .....	13	2.7.1 概念 .....	64
1.4.1 整体式系统 .....	13	2.7.2 作业状态 .....	64
1.4.2 模块层次式系统 .....	14	2.7.3 作业调度 .....	65
1.4.3 客户/服务器系统 .....	15	2.8 操作系统的接口 .....	70
习题 .....	17	2.8.1 命令接口 .....	70
<b>第 2 章 处理机管理</b> .....	18	2.8.2 程序接口 .....	71
2.1 进程概念 .....	18	2.8.3 图形用户接口 .....	73
2.1.1 进程的引入 .....	18	习题 .....	73
2.1.2 进程的描述 .....	23	<b>第 3 章 存储器管理</b> .....	76
2.2 OS 内核及进程控制 .....	28	3.1 存储管理概述 .....	76
2.2.1 OS 内核 .....	28	3.1.1 存储管理的概念 .....	76
2.2.2 进程控制 .....	28	3.1.2 地址重定位 .....	76
2.2.3 进程控制原语 .....	29	3.1.3 虚拟存储器概念的引入 .....	78
2.3 进程通信 .....	33	3.2 分区式存储管理 .....	78
2.3.1 进程同步与互斥 .....	33	3.2.1 单一连续分配 .....	79
2.3.2 经典同步问题 .....	41	3.2.2 固定式分区分配 .....	79
2.3.3 进程通信 .....	44	3.2.3 动态分区分配 .....	80
2.4 进程调度 .....	48	3.2.4 动态重定位分区分配 .....	83

3.2.5 分区分配方案的评价 .....	83	4.3.2 单缓冲与双缓冲 .....	106
3.3 覆盖和对换 .....	84	4.3.3 循环缓冲 .....	107
3.3.1 覆盖(Overlay).....	84	4.3.4 缓冲池 .....	108
3.3.2 对换(Swapping).....	84	4.4 设备分配 .....	109
3.4 分页存储管理 .....	85	4.4.1 设备分配的原则与策略 .....	110
3.4.1 分页原理 .....	85	4.4.2 设备分配中使用的数据结构 .....	110
3.4.2 分页存储分配与回收算法 .....	86	4.4.3 设备分配算法 .....	112
3.4.3 地址映射机构 .....	87	4.5 设备虚拟性与独立性 .....	113
3.4.4 多级页表 .....	88	4.5.1 SPOOLing 定义.....	113
3.4.5 分页存储管理方案的评价 .....	89	4.5.2 SPOOLing 系统的组成 .....	114
3.5 请求分页存储管理 .....	90	4.5.3 SPOOLing 系统的应用 .....	114
3.5.1 请求分页原理 .....	90	4.5.4 设备独立性 .....	115
3.5.2 页面置换算法 .....	91	4.6 设备处理过程 .....	115
3.5.3 页面分配 .....	93	4.6.1 设备驱动程序功能与特点 .....	115
3.5.4 抖动 .....	94	4.6.2 设备处理方式与设备处理流程 .....	116
3.5.5 工作集 .....	95	习题 .....	117
3.5.6 性能分析 .....	95		
3.6 分段存储管理 .....	96	<b>第 5 章 文件系统</b> .....	118
3.6.1 分段原理 .....	96	5.1 文件系统概述 .....	118
3.6.2 地址映射机构 .....	96	5.1.1 文件 .....	118
3.6.3 请求分段存储管理 .....	97	5.1.2 文件系统 .....	120
3.6.4 段的保护与共享 .....	98	5.2 文件结构和文件存取 .....	121
3.6.5 分段存储管理方案的评价 .....	98	5.2.1 文件逻辑结构及文件存取 .....	121
3.7 段页式存储管理 .....	98	5.2.2 文件的物理结构(系统观点) .....	125
3.7.1 段页式存储管理的实现 .....	99	5.2.3 记录成组技术 .....	129
3.7.2 段页式存储管理的评价 .....	100	5.2.4 文件存储空间管理 .....	130
习题 .....	100	5.3 文件目录和目录检索 .....	136
		5.3.1 文件控制块 .....	136
<b>第 4 章 设备管理</b> .....	102	5.3.2 目录结构 .....	136
4.1 概述 .....	102	5.3.3 文件目录检索 .....	140
4.1.1 设备的类别 .....	102	5.3.4 文件目录实例 .....	141
4.1.2 设备管理的功能和任务 .....	103	5.4 文件共享与保护的实现 .....	142
4.2 数据传送控制方式 .....	103	5.4.1 文件共享 .....	142
4.2.1 程序直接控制方式 .....	103	5.4.2 文件保护 .....	143
4.2.2 中断方式 .....	104	5.5 文件系统的用户接口 .....	146
4.2.3 DMA 方式 .....	104	5.5.1 文件操作 .....	146
4.2.4 通道控制方式 .....	105	5.5.2 目录操作 .....	148
4.3 缓冲管理 .....	106	习题 .....	148
4.3.1 缓冲的引入 .....	106		

<b>第 6 章 UNIX 系统内核</b> .....	150	6.5.3 字符设备管理 .....	178
6.1 概述 .....	150	习题 .....	180
6.1.1 UNIX 系统简介 .....	150	<b>第 7 章 高级操作系统</b> .....	182
6.1.2 UNIX 系统的特点 .....	152	7.1 多处理机操作系统 .....	182
6.1.3 UNIX 系统的核心结构 .....	154	7.1.1 多处理机系统的分类 .....	182
6.2 进程管理 .....	155	7.1.2 多处理机系统的结构 .....	184
6.2.1 UNIX 进程映象和进程状态 .....	155	7.1.3 多处理机操作系统 .....	187
6.2.2 进程控制 .....	157	7.2 网络操作系统 .....	188
6.2.3 进程调度 .....	159	7.2.1 计算机网络的诞生和定义 .....	188
6.2.4 进程通信 .....	161	7.2.2 计算机网络的结构与组成 .....	189
6.3 文件系统 .....	166	7.2.3 计算机网络的分类 .....	191
6.3.1 文件系统结构 .....	166	7.2.4 网络协议体系结构 .....	193
6.3.2 文件控制块——i 节点的分配与 释放 .....	167	7.2.5 网络操作系统 .....	197
6.3.3 文件的打开与关闭 .....	168	7.3 分布式操作系统 .....	200
6.3.4 文件的物理组织——多重索引 结构 .....	170	7.3.1 分布式系统概念 .....	200
6.3.5 文件存储空间管理——空闲块 成组链接法 .....	171	7.3.2 分布式系统的设计 .....	201
6.3.6 主要数据结构之间的联系 .....	172	7.3.3 分布式操作系统 .....	204
6.4 存储管理 .....	173	7.3.4 分布式操作系统中的进程通信 .....	209
6.4.1 对换 .....	173	7.3.5 分布式操作系统中的进程同步 .....	212
6.4.2 请求分页 .....	174	7.3.6 分布式系统的资源管理 .....	215
6.5 设备管理 .....	175	7.3.7 分布式系统中的死锁 .....	217
6.5.1 UNIX 系统设备开关表 .....	175	习题 .....	220
6.5.2 UNIX 系统的块设备管理 .....	176	<b>参考文献</b> .....	222





# 第 1 章 引 论



## 本章要点

本章给出操作系统的定义；介绍操作系统的发展过程及基本操作系统类型和新型操作系统；重点对操作系统的功能和特征简要地予以分析。

## 1.1 操作系统的基本概念

### 1.1.1 操作系统的定义

我们将未配置任何软件的计算机称为裸机，即由中央处理器（CPU）、存储器、输入/输出（I/O）设备等硬件组成的计算机。裸机是无法充分发挥硬件性能的，不适合一般的用户使用，只有配备了软件以后，计算机才可以更好地完成对信息的存储、检索和处理等一系列工作。所以我们可以把计算机系统划分为两大部分：硬件和软件。硬件是计算机系统的物质基础，它包括多种多样的物理设备。而软件可以分为应用软件和系统软件。应用软件，例如字处理软件 Word、图像处理软件 Photoshop、Web 浏览器 IE、编译软件等，可以为用户解决具体的问题。而系统软件是指使计算机能够工作的一些基础软件，只有在系统软件的支持下，应用软件才能够正常的运行。我们将要研究的操作系统（Operating System）就是系统软件的典型代表。

引入操作系统以后，计算机的组成如图 1-1 所示。

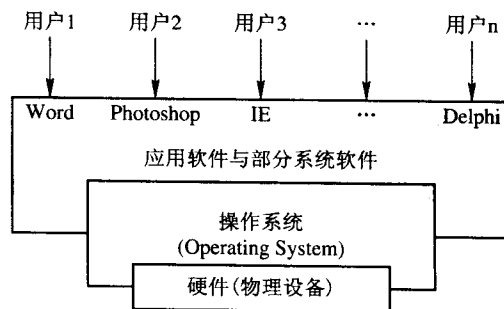


图 1-1 计算机系统组成

从操作系统在计算机系统中所处的位置，我们可以看到，操作系统处在硬件与其它软件之间的一个特殊位置上，引入操作系统主要可完成以下两个方面的工作：



(1) 方便用户使用。操作系统应该提供给用户一个好的界面，用户不必了解硬件和其它软件的细节，就可以方便地使用计算机。

(2) 充分利用资源。操作系统应该最大限度地发挥计算机系统资源的使用效率，合理地组织工作流程，使得计算机资源能为多用户共享。

据此，我们可以给出操作系统的定义：操作系统是为了方便用户和提高计算机的利用率，而对计算机资源进行组织和管理的程序集合。其中，用户是一个广义的概念，包括一般用户和软件开发人员等；资源包括处理器、存储器、输入/输出设备等硬件资源和程序与数据等软件资源。

### 1.1.2 操作系统在计算机系统中的作用

操作系统在计算机中起着重要的作用，一方面它提供用户与计算机之间的接口，方便用户的使用；另一方面它管理和控制计算机系统中的软硬件资源，使之得到有效的利用。这两个方面的作用也称为对外职能和对内职能。

#### 1. 对外职能——用户与计算机之间的接口

操作系统作为用户与计算机之间的接口，必须为用户提供良好的界面，使用户能够感觉计算机是可用而且易用的。这是一个最主要的职能。早期的计算机没有操作系统，使用计算机时只能使用机器语言，既繁琐又费时，没有计算机专业知识的人使用计算机时会感到困难重重。有了操作系统，用户通过操作系统提供的操作命令和系统调用命令就可以方便、快捷、安全、可靠地操纵计算机，运行自己的程序。这些命令对于用户来说要容易掌握得多。而随着图形用户接口的出现，使用户尤其是普通用户操作计算机变得更为简单。

#### 2. 对内职能——组织和管理计算机资源，充分发挥资源的效能，提高利用率

在一个计算机系统中，通常包含了多种硬件与软件资源，为了实现某种功能需要使用相应的资源，在分配资源时首先不能“张冠李戴”，比如某用户程序要求打印一篇文档，系统就不应该为其分配一台扫描仪。另外，系统的资源数目是有限的，因此在一些环境下用户程序在执行过程中会发生资源竞争的现象，引起冲突，造成系统的一些工作无法继续进行，这就需要有一个仲裁者在相互竞争资源的程序间有序地控制资源的分配，操作系统就起这样的作用。操作系统可以跟踪资源的使用情况，满足用户资源请求，做到合理分配，尽可能使得系统中的各类资源处于“忙”状态，缩短任务完成时间，提高资源利用率。

### 1.1.3 研究操作系统的几种观点

操作系统本身是一个相当复杂的系统，为了彻底了解它，我们可以从不同的角度研究操作系统。

#### 1. 用户观点

用户是计算机的使用主体，软件开发的主要目的就是让用户能够方便地使用计算机。用户观点是指站在用户立场上对计算机提出要求。用户对操作系统的内部结构没有太大兴趣，他们主要从功能要求、服务质量、使用维护及安装价格等方面提出要求。作为使用者，最关心的是如何利用操作系统提供的服务有效地使用计算机，即操作系统提供什么样的用



户界面。

### 2. 资源管理的观点

资源管理的观点是从系统资源的角度刻画整个操作系统的组成及其功能。通常把计算机系统资源按性质归纳为四类：CPU、内存、外部设备以及信息文件，这些资源是操作系统和用户程序赖以活动的物质基础和工作环境。用户使用计算机的过程就是通过使用这些资源达到解决问题的一个过程。

资源管理是指对上述硬、软件资源进行管理。一般说来，一个计算机系统中常常同时运行多个程序，操作系统的任务就是在相互竞争资源的程序之间有序地控制资源的分配。对于操作系统来说，资源管理要做如下工作：

- (1) 对资源使用情况进行登记，这是资源管理的依据。
- (2) 决定分配策略。根据资源使用情况分析分配的可行性，并决定分配所应采取的策略。
- (3) 提供分配与回收算法。通过资源分配的具体程序达到资源分配的目的，为了保证资源的后续利用，在当前使用完该资源后应该立即释放，即执行资源回收程序。

如图 1-2 所示，系统中存在两个不同的队列，分别为请求分配资源队列和请求释放资源队列。这两个队列按照某种原则形成，系统首先根据它们不同的请求方式加以判断。如果是请求分配资源方式，系统将进一步判断所需资源是否可用，如果资源可用并且够用，就进行资源分配并修改资源分配状态表，否则这次请求就没有被成功响应，该请求将被插入到请求分配资源队列，等待下一次系统响应。如果不是请求分配资源方式，那么就有两种可能：一种是请求释放资源方式，这时系统正常回收资源；另一种是异常方式，系统将调用出错处理程序进行相应的处理。

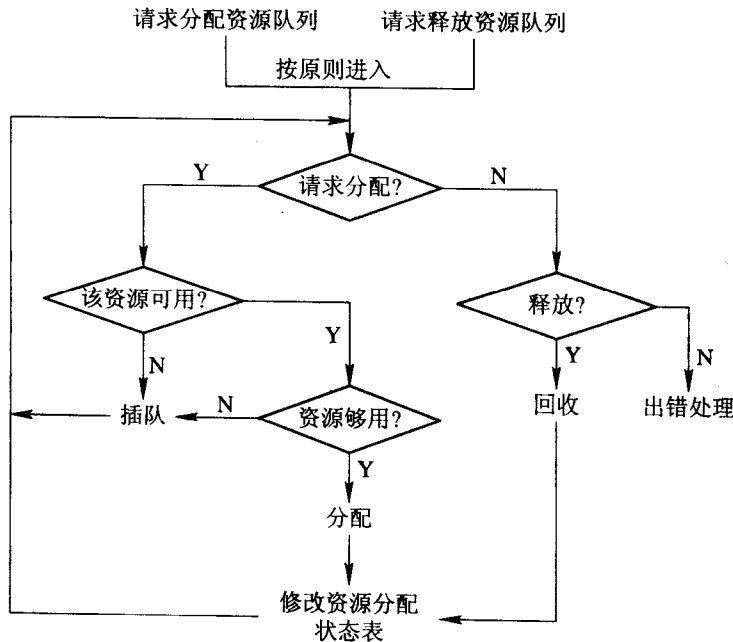


图 1-2 资源分配与释放过程



### 3. 进程观点

通过以前的学习，我们了解到计算机是通过运行程序来满足用户要求的，但是我们并不清楚程序在计算机内运行的基本过程和一些规则，只是“知其然而不知其所以然”，操作系统将会解决一些存在于我们思想中的疑问。在澄清这些疑问的过程中，有一个非常重要的动态对象——进程，它是分析计算机操作系统在处理过程中的基本对象，是系统中的活动实体。我们在后面的章节中会详细描述并经常用到它来研究系统的工作状况。

## 1.2 操作系统的发展

操作系统是由于计算机应用的客观需要而产生的，它与计算机的体系结构之间有着非常密切的联系，一方面它随着充分发挥计算机硬件性能的要求不断发展，另一方面它也促进了计算机硬件种类的扩充，使得计算机体系结构得到不断改进，功能越来越强大。通过对操作系统发展的回顾，我们可以了解到一些操作系统的基本要素。

### 1.2.1 问题的提出

操作系统是计算机系统中硬件与其它软件之间的第一层接口，硬件和软件的发展引导操作系统的变化和发展。操作系统产生的主要原因有以下两个方面。

#### 1. 速度问题

20世纪40年代出现的第一台计算机由电子真空管和插板组成。这种机器运算速度很低。在计算机出现的早期，每台计算机都有一个小组专门进行设计、编程和维护。所有的程序都采用机器语言，计算机上没有操作系统，更没有其它软件。机器的使用方式是程序员提前预约机时，上机时用手方式启动设备。将记录程序的卡片或纸带用输入机送入计算机，接着在控制台上用按钮启动程序运行。在程序运行的几个小时里用户独占计算机，他既是程序员又是操作员。

20世纪60年代，晶体管的发明改变了计算机的状况，使处理机的速度提高了几十倍。当时计算机的价格十分昂贵，大学或研究所的机房已经配有专门的操作员。程序员要使用计算机时，将程序写在纸上，利用穿孔机制成卡片，称为一个作业，然后交给操作员。由于在计算机的使用过程中操作员需要手工操作的工作很多，例如取一个作业，操作运行，打印输出，再取下一个作业等，造成机器长时间空闲，人工操作的速度和处理机的计算速度严重不匹配。由此可见，减少人工干预对提高机器利用率有着重要意义，人们开始希望有专门的监督程序或管理程序来进行作业的自动依次处理。

#### 2. 作业差异

除了速度问题之外，由于每个作业使用的语言不同，也给操作员带来了很大的麻烦。因为不同语言的作业在进行处理前，首先必须将相应的编译系统装入计算机并置换旧的程序，这样就使计算机的大量时间耗费在输入/输出上。例如，为了完成一道也许计算机处理只需要两分钟的程序，而将相应的编译系统装入也许会花费两个小时，且在这两个小时中，宝贵的CPU资源却一直处于空闲状态，白白浪费，另外也使操作员不胜其烦。因此设计使



用专门程序进行作业自动转换和处理的办法被提了出来。

另外,作业也有 I/O 繁忙型与 CPU 繁忙型之分。对于 I/O 繁忙型作业,在完成过程中大部分时间利用外部设备进行 I/O 操作,在这些时间内,宝贵的 CPU 资源却处于空闲状态,成为一种严重的资源浪费。对于 CPU 繁忙型作业,虽然 CPU 利用率比较高,但是外部设备却经常处于空闲状态,仍然属于一种资源浪费。也就是说,对于不同应用的作业来说,计算机系统给予了“一视同仁”的对待。虽然在控制上比较简单,却付出了宝贵的时间代价。

## 1.2.2 解决办法

### 1. 减少人工干预

为了尽量减少多个作业交接时的人工负担,批处理技术应运而生。操作员把用户提交的作业分类,把一批作业编成一个作业执行序列,成批地提交给计算机进行输入、运行和输出,同时也专门编制了监督程序或管理程序,对它们进行自动地依次处理,实现作业的自动转换。在早期的批处理中,操作员通过输入设备将多个作业一次存入磁带,然后监督程序先从磁带上读入一个作业,接着又调入编译程序来将用户作业翻译成目标代码,经连接和装配后执行并输出结果。紧接着又调入下一个作业,不断地循环反复操作。这样大大减少了人工操作的时间,这种处理方式被称为联机批处理方式。

### 2. 改善速度匹配

批处理方式解决了作业的自动转接,但其缺点是无论成批输入还是作业的转接执行都是在主机 CPU 控制下完成的。在系统进行慢速的输入/输出工作的时候,CPU 只好停止工作,处于等待状态。为了改善速度匹配问题,产生了脱机批处理技术。该技术是增加不与主机相连而专门用于输入/输出的外围计算机,由外围输入计算机负责将作业送入磁带,由主机和磁带机配合读入作业,主机完成程序的执行后将输出结果送到磁带,再由外围输出计算机将结果送到输出设备并显示给用户。

这样,主机不是直接与慢速的输入/输出设备打交道,而是与速度相对快得多的磁带机发生联系。主机和外围计算机分工明确,在监督程序的管理下,充分发挥了主机的高速度处理能力,大大减少了因手工操作而带来的主机机时浪费,提高了主机的使用效率,比较明显地缓解了高速主机和低速输入/输出设备之间的矛盾。

### 3. 实现多道程序系统

20 世纪 60 年代中期,处理机速度进一步提高,主存容量加大,出现了大容量的辅助存储器。计算机的系统结构从以中央处理机为中心改变为以主存为中心。其中最重要的是中断技术和通道技术的出现和发展,使操作系统有了进一步发展。借助通道和中断技术,主机可以在中断的时刻向 I/O 通道中的缓冲区进行读/写操作,更迅速地完成任务。但是包括 CPU 和 I/O 设备在内的各种资源不能并行工作仍然是造成系统性能低效的主要原因,所以在硬件系统的支持下,通过软件技术使得系统中的 CPU 和其它的 I/O 设备并行工作,推出了多道程序系统。我们从图 1-3 中可以了解 CPU 与 I/O 并行的大致协调过程。

所谓多道程序系统,是指允许多个相互独立的程序同时存在于内存中,而且处于同时运行的过程中。各道程序轮流占用 CPU,交替执行。这里需要解决一些技术问题,例如,如何共享硬件、软件资源,怎样解决同步和互斥的问题,如何提高内存的使用效率,如何



保证程序的安全等，这些将是操作系统这门课程要介绍的重要内容。

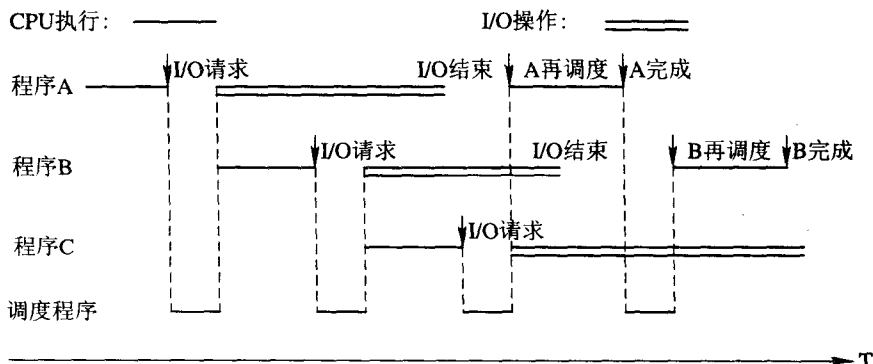


图 1-3 CPU 与 I/O 并行图

### 1.2.3 基本操作系统类型

#### 1. 批处理系统

批处理系统中，用户的作业分批提交并处理，即系统将作业成批地输入系统并暂存在外存中，组成后备作业队列，每次按一定的调度原则从后备作业中选择一个或多个装入主存进行处理，作业完成后退出。这些操作由系统自动实现，大大缩短了两个作业之间的转接时间，在系统中形成了一个自动转接的作业流，当一批作业运行完毕，输出结果后，系统便接收下一批作业。

在批处理系统中，用户不能直接干预作业的运行过程，而是将其对作业的控制意图在作业提交之前用作业控制语言编制成作业说明书或作业控制卡，这些控制意图可以是作业运行时的资源请求，或是对可能产生的运行错误的相应处理等。作业说明书或作业控制卡在提交作业时，与程序和数据一起提交给系统，由作业控制程序或命令解释程序解释执行，并且提供相应的服务。

在引入多道程序设计技术后，批处理系统具有以下特征：

- (1) 多道性：在内存中同时驻留多道程序，它们并发执行，以提高系统的资源利用率和系统的吞吐量。
- (2) 无序性：作业的执行顺序与作业进入系统的先后顺序没有严格的对应关系。
- (3) 调度性：作业从提交到运行完成需要经过两次调度，即作业调度和进程调度。作业调度是指按照一定作业调度算法，从后备作业队列中选择一个或几个作业调入内存。进程调度是指按照一定进程调度算法，从在内存的进程中选择一个进程，将处理机分配给它，使之执行。

批处理系统，特别是多道批处理系统的主要优点是资源利用率高。由于在内存中装入了多道程序，系统中的资源被它们充分共享和利用，而且它们之间的切换所造成的系统开销较小，使系统在单位时间内所完成的总的工作量增加，即提高了系统的吞吐量。

批处理系统的缺点主要体现在以下两个方面：

- (1) 平均周转时间长。作业的周转时间是指从作业进入系统开始，直到作业完成并退出



系统为止所经历的时间。在批处理系统中，由于作业需要排队来依次进行处理，因而作业的周转时间较长。

(2) 无交互能力。在作业提交后，用户不能与自己的作业进行交互，不便于对作业的控制。

## 2. 分时系统

所谓“分时”，就是多个用户对系统资源进行时间上的分享。在分时操作系统中，一个计算机系统与若干台本地或远程终端相连，每个用户可以在所使用的终端上以人一机会话的交互方式使用计算机。微观上，每个用户作业轮流运行一个时间片；宏观上，多个用户同时工作，共享系统资源。当用户通过键盘或者鼠标向系统提交一个请求后，会在 1 s（甚至更短）之内给出响应，实现了及时交互。

“分时”的设想最初是在 20 世纪 60 年代提出的，但是具体的分时系统的设计实现是很困难的，成本非常昂贵，所以直到 20 世纪 70 年代才得到逐步推广。目前尽管批处理系统仍然在某些方面继续使用，但是分时系统作为多道程序系统的一个典型代表，集中体现了多道程序系统的一些技术特征，成为当今的计算机操作系统主流，所以我们将以分时系统为主要对象来学习操作系统。

## 3. 实时系统

“实时”是指系统能够及时响应发生的外部事件(一般为一些随机事件)，并以足够快的速度完成对事件的处理。从这个定义中我们可以看到，在对时间响应的要求上，实时系统比分时系统要严格得多，一般在毫秒级、微秒级；而批处理系统甚至可以不作响应时间的要求。为了保证程序可靠运行，系统应提供安全措施，比如多级容错、硬件冗余等，避免因发生错误或丢失信息而造成重大经济损失甚至导致灾难性的后果。实时系统相对于其它操作系统来说，其优点是系统的及时响应以及系统的可靠性。

实时系统可以分为两种类型：一种称作硬实时系统，另一种称作软实时系统。硬实时系统对响应时间的要求最为严格，一个实时任务绝对不允许出现等待的情况。一旦有实时任务到来，系统核心必须立即中止正在处理的任务而转去处理它，一定要在规定的延迟时间之内作出信息反馈。从辅存中存取数据的方式在硬实时系统中是被限制使用甚至被禁止使用的，数据被存放在可以高速访问的内存或者只读存储器 ROM 中。ROM 是一种系统断电后仍然能够保留其信息的存储介质，而内存中存储的信息在系统断电后将会丢失。实际上大部分先进的系统并不采用这种硬实时方式，因为在这些系统中更倾向于使用户能够与计算机硬件尽可能分离，这种分离会导致延迟时间难以确定。比如我们将要在存储器管理中学习的虚拟存储技术，在实时系统当中几乎都没有使用，而且硬实时系统与分时系统会产生冲突，没有一个通用的分时系统支持硬实时功能，所以我们在本书中对这种技术不予分析。

在软实时系统中，系统通过用户任务的优先级来保证实时性。一个实时任务需要在一定的截止时间内作出反馈，它的优先级通常是高于其它任务的。软实时系统可以应用于多媒体、虚拟现实、海底探测等科学领域以及一些工业自动化控制领域和军事自动化控制领域中。软实时系统有相对“宽松”的延迟，允许它很好地与其它类型的操作系统相混合，比如在 UNIX 中我们就可以看到软实时的控制技术。



以上介绍的是三种基本的操作系统类型，实际的操作系统可能兼有三者或其中两种系统的功能。

#### 1.2.4 新型操作系统

操作系统经过一个大发展时期，到 20 世纪 80 年代已趋于成熟，随着集成电路技术和计算机体系结构的发展，操作系统也在不断地发展，先后形成了微型计算机操作系统、多处理机操作系统、网络操作系统和分布式操作系统。

##### 1. 微机操作系统

随着超大规模集成电路（简称 VLSI）的发展而产生了微型计算机，配置在微型计算机上的操作系统称为微型计算机操作系统（简称微机操作系统）。微机操作系统按照处理机的字长可分成 8 位、16 位以及 32 位等操作系统。典型的操作系统有 DOS、Windows、OS/2，以及近年来引起广泛关注的 Linux 等一些自由软件操作系统。

MS - DOS 是单用户单任务操作系统的代表。单用户单任务的操作系统每次只允许一个用户上机，且一个用户程序只能作为一个任务运行。早期的 DOS 因为配置在 IBM - PC 机上，而随着该机种及其兼容机的畅销，MS - DOS 操作系统事实上也就成了当时的个人计算机操作系统的标准。DOS 中提供了许多命令接口，该操作系统具有较强的功能及性能优良的文件系统。

1990 年 Microsoft 公司在 MS - DOS 的基础上推出了 Windows 3.0 操作系统，它凭借友好的图形用户界面、支持多任务等优点而迅速流行开来，占领了操作系统大部分的市场份额，成为单用户多任务操作系统的代表。单用户多任务操作系统是指每次只允许一个用户上机，但允许将一个用户的工作分为若干个任务，使它们并发执行，从而提高系统的性能。Microsoft 公司于 1993 年推出了全新的 32 位操作系统 Windows NT，它具有很强的网络功能。随着新一代 Windows 系列操作系统的陆续推出与应用，它们已经成为目前个人计算机操作系统的主流。

前面介绍的 DOS、Windows 操作系统都是单用户操作系统，每次只允许一个用户上机，上机的用户独自占有了系统中的所有资源。UNIX 操作系统克服了这种资源被独占的问题。UNIX 是一种多用户多任务的操作系统，允许多个用户通过各自的终端使用一台主机，共享主机系统的各类资源，而每个用户工作又可以分为几个任务，使它们并行执行，从而可以提高资源的利用率和增加系统的吞吐量。目前许多大、中、小型计算机上配置的都是多用户多任务的操作系统。

##### 2. 多处理机操作系统

较早的计算机系统基本上都是单处理机系统，后来出现了多处理机系统 MPS (Multiprocessor System)，从计算机体系结构上来改善系统性能。随着系统中处理机数目的增多，既提高了系统的吞吐量，又可以共用其它部件以节约成本。而且多处理机系统的系统重构功能提高了系统的可靠性，即当其中任何一台处理机发生故障时，系统能立即将该处理机上所处理的任务迁移到其它的一个或多个处理机上去处理，整个系统仍能正常运行，仅使系统的性能略有降低。根据各处理机之间的紧密程度，我们可以把多处理机系统从硬件结构上分为紧密耦合型和松散耦合型两大类。





多处理机系统所配置的操作系统可以分为对称式和非对称式（又称为主-从式）的。对称式是每个处理机上运行着同一操作系统的拷贝，这些拷贝可以与其它拷贝之间根据任务需要进行通信。非对称式是指每个处理机执行特定的任务，由一台主处理机控制整个系统，其它处理机根据主处理机的指令完成指定的任务，主处理机和其它处理机之间具有主-从关系。目前大多数多处理机系统采用的是对称式操作系统。

### 3. 网络操作系统

计算机网络定义为一些具有独立处理能力的计算机在网络协议控制下实现通信和相互合作的计算机复合系统，是在计算机技术和通信技术高度发展的基础上相互结合的产物。

计算机网络的发展，尤其是 Internet 和 WWW (World Wide Web) 的出现，对当今操作系统的发展起到了深远影响。当 PC 机在 20 世纪 70 年代出现时，仅仅面向个人独立使用。20 世纪 80 年代，随着人们开始通过 Internet 发送电子邮件 E-mail、进行文件传输 FTP 等工作，许多 PC 机被连接到计算机网络上。到 20 世纪 90 年代中期，网络连接被认为是计算机系统最基本的功能之一，于是操作系统开始提供有关网络的管理，比如 TCP/IP 和 PPP 协议、网络浏览器等。

如同在单机上安装操作系统可以方便用户合理组织与管理计算机资源一样，在网络系统之上覆盖一层网络操作系统后，用户可以方便地使用网络，实现用户通信和资源共享，提高网络资源的利用率和网络的吞吐量。

### 4. 分布式操作系统

分布式处理系统是指由多个分散的处理单元经互连网络的连接而形成的，且可以实现分布处理的系统。其中，每个处理单元既具有高度的自治性，又相互协调，能在系统范围内实现资源管理、动态地分配任务，并能并行地运行分布式程序。

在分布式系统上配置的操作系统，称为分布式操作系统。网络操作系统与分布式操作系统的特点如表 1-1 所示。

表 1-1 网络操作系统与分布式系统比较

操作系统	分布性	并行性	透明性	共享性	健壮性
网络操作系统	分布处理，集中控制	任务在本地计算机上运行	操作透明，但需指明物理位置	一般只能共享服务器上的资源	控制集中于服务器，可靠性弱
分布式操作系统	分布处理，分布控制	多任务在多处理单元中并行执行	操作透明，而且物理位置透明	各站点资源可供全系统共享	容错能力强，可靠性高

## 1.3 操作系统的特征与功能

### 1.3.1 操作系统的特征

虽然前面介绍的三种基本操作系统各有自己的特征，如批处理系统具有成批处理的特征，分时系统强调交互特征，实时系统强调实时特征，但它们也都具有以下四个基本特征。