

计算机与信息技术专业应用教材

操作系统教程

(第2版)

曾平 郑鹏 金晶 编著



清华大学出版社

TP316/351=2

2008

► 计算机与信息技术专业应用教材

操作系统教程（第2版）

曾平 郑鹏 金晶 编著

清华大学出版社

北京

内 容 简 介

操作系统是计算机专业的一门重要专业基础课程。本书是《操作系统教程》的第2版,是作者对原书内容的升级,体例安排更加合理,讲解更加顺畅,易于初学者轻松掌握。

本书全面系统地介绍了现代操作系统的基本概念、原理和实现方法。本书共分为8章,第1章介绍了操作系统的发展历史、类型及特征,第2章~第4章介绍了进程管理,第5章介绍了存储管理,第6章介绍了设备管理,第7章介绍了文件管理,第8章介绍了UNIX操作系统。本书的每一章都有小结和练习题,在全书的附录中还给出了各章习题的参考答案。

本书可以作为高等院校相关专业本、专科学生学习操作系统课程的教材或参考书,也可以作为从事计算机工作人员的自学用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,翻印必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

操作系统教程/曾平,郑鹏,金晶编著. —2版. —北京:清华大学出版社,2008.5

ISBN 978-7-302-17367-0

I. 操… II. ①曾…②郑…③金… 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(2008)第054731号

责任编辑:沙作伟

封面设计:林 陶

出 版 者:清华大学出版社

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京市鑫山源印刷有限公司

经 销:新华书店总店北京发行所

开 本:185×260 印张:17.5

字 数:426千字

版 次:2008年5月第1版

印 次:2008年5月第1次印刷

印 数:1~4 000

定 价:26.00元

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:010-62770177 转 3103 产品编号:028977-01

丛 书 序

为适应信息社会高速发展的需求，目前全国各类高等院校都在进行计算机教学的全方位改革，目的是规划出一整套面向计算机与信息技术专业、具有中国高校计算机教育特色的课程计划和教材体系。本丛书就是在这一背景下应运而生的。我们组织了由全国高校计算机专业的专家教授组成的“计算机与信息技术专业应用教材”课题研究组，通过对计算机和信息技术专业全方位的研讨，并结合我国当前的实际情况，编写了这套系统性、科学性和实践性都很强的丛书。

丛书特色

☑ 先进性：力求介绍最新的技术和方法

先进性和时代性是教材的生命，计算机与信息技术专业的教学具有更新快、内容多的特点，本丛书在体例安排和实际讲述过程中都力求介绍最新的技术和方法，并注重拓宽学生的知识面，激发他们的学习的热情和创新欲望。

☑ 理论与实践并重：阐明基础理论，强调实践应用

理论是实践的基础，实践是理论的升华；不能有效指导实践的理论是空头理论，没有理论指导的实践是盲目的实践。对于时代呼唤的信息化人才而言，二者缺一不可。本丛书以知识点为主线，穿插演示性案例于理论讲解之中，使枯燥的理论变得更易于理解、易于接受；此外，还在每一章的末尾提供大量的实习题和综合练习题，目的是提高学生综合利用所学知识解决实际问题的能力。

☑ 易教易学：创新体例，合理布局，通俗易懂

本丛书结构清晰，内容系统详实，布局合理，体例较好；力求把握各门课程的核心，通俗易懂，便于教学的展开，也便于学生学习。

丛书组成

本次推出的计算机与信息技术专业应用教材，涵盖计算机基础、程序设计和数据库三大领域，共 19 本：

- 计算机系统结构教程
- Java 语言程序设计
- C++程序设计（第 2 版）
- C++程序设计学习与上机实验指导
- 数据结构与算法教程（第 2 版）

- C 程序设计教程（基于 Visual C++ 平台）
- C 程序设计教程学习与上机指导（基于 Visual C++ 平台）
- C 程序设计教程（基于 Turbo C 平台）
- Access 数据库程序设计
- Access 2003 程序设计教程（第 2 版）
- 数据库原理与应用——基于 Visual FoxPro（第 2 版）
- 数据库原理与应用——基于 SQL Server 2000（第 2 版）
- Visual FoxPro 程序设计（第 2 版）
- Visual Basic 程序设计（第 2 版）
- 操作系统教程（第 2 版）
- SQL Server 2000 应用系统开发教程（第 2 版）
- SQL Server 2000 学习与上机实验指导
- 数据库原理与应用——基于 Access
- 数据库原理与应用——基于 Access 2003（第 2 版）

服务之窗

本丛书的出版者和作者竭诚为读者提供服务。

本套丛书免费为教师提供 PowerPoint 演示文档，该文档可将书中的内容及图片以幻灯片的形式呈现在学生面前，在很大程度上减轻了教师的备课负担，所以深受广大教师的欢迎。请致电：010-82896438 或发 E-mail: feedback@khp.com.cn 获取电子教案。

丛书编委会

主任委员：	李春葆					
副主任委员：	苏光奎	朱福喜				
委员：	尹为民	尹朝庆	李春葆	伍春香	朱福喜	
	苏光奎	胡新启	徐爱萍	曾平	曾慧	

编者寄语

如果说科学技术的飞速发展是 21 世纪的一个重要特征的话，那么教学改革将是 21 世纪教育工作不变的主题。要紧跟教学改革，不断创新，真正编写出满足新形势下教学需求的教材，还需要我们不断地努力实践、探索和完善。本丛书虽然经过细致的编写与校订，仍难免有疏漏和不足，需要不断地补充、修订和完善。我们热情欢迎使用本丛书的教师、学生和读者朋友提出宝贵意见和建议，使之更臻成熟。

本丛书作者的电子邮件：licb1964@126.com

本丛书出版者的电子邮件：feedback@khp.com.cn

前 言

操作系统是现代计算机系统中不可缺少的基本系统软件。如果让用户去使用一台没有配置操作系统的计算机，那是难以想象的。操作系统控制和管理整个计算机系统软硬件资源，并为用户使用计算机提供一个方便灵活、安全可靠的工作环境。

本书遵循操作系统课程的教学大纲要求，内容共分8章：第1章是操作系统引论，主要介绍操作系统的发展历史，操作系统的类型、特征及功能；第2章是进程描述与控制，主要介绍进程的定义，进程的状态及进程控制；第3章是进程同步与通信，主要介绍进程的同步与通信设施；第4章是调度与死锁，主要介绍处理机调度算法，死锁的定义及处理方法；第5章是存储管理，主要介绍存储管理的基本概念，分区存储管理，分页存储管理，分段存储管理及虚拟存储器技术；第6章是设备管理，主要介绍设备管理的基本概念，I/O控制方式，设备分配及I/O软件的层次结构；第7章是文件管理，主要介绍文件管理的基本概念，文件的结构，文件存储空间的管理，文件共享和保护；第8章是UNIX操作系统，主要介绍UNIX操作系统的内核设计原理，包括进程管理、内存管理、设备管理及文件管理功能的实现算法及相关数据结构。

本书的每一章都有小结和练习题，小结中列出了每章内容的重要知识点，练习题可以帮助读者巩固所学知识。本书附录中给出了各章习题的参考答案，以方便读者自学。

本书可以作为高等院校相关专业本、专科学生操作系统课程的教材或参考书，也可以作为从事计算机工作人员的自学用书。

尽管作者讲授该课程多年并试图努力编写好此书，但由于水平有限，书中难免存在错误或不足之处，敬请读者和同行批评指正。

作者联系方式：

zengping@public.wh.hb.cn

pzheng51@163.com

zn_jinjing@163.com

编 者
2008年4月



第1章 操作系统引论.....1

1.1 计算机系统概述.....1	1
1.2 操作系统的发展历史.....2	2
1.2.1 手工操作阶段.....2	2
1.2.2 早期批处理.....3	3
1.2.3 多道程序设计技术.....4	4
1.2.4 操作系统的形成及发展.....6	6
1.2.5 推动操作系统发展的动力.....6	6
1.3 操作系统的类型.....7	7
1.3.1 批处理系统.....7	7
1.3.2 分时系统.....8	8
1.3.3 实时系统.....9	9
1.3.4 其他操作系统类型.....9	9
1.4 操作系统的特征.....11	11
1.5 操作系统的功能.....12	12
1.6 操作系统的接口.....14	14
1.6.1 命令接口.....14	14
1.6.2 程序接口.....15	15
1.6.3 图形用户接口.....16	16
1.7 研究操作系统的几种观点.....16	16
1.8 小结.....17	17
练习题1.....19	19

第2章 进程描述与控制.....21

2.1 进程的引入.....21	21
2.1.1 前趋图.....21	21
2.1.2 程序的顺序执行.....22	22
2.1.3 程序的并发执行.....22	22
2.1.4 程序并发执行的条件.....23	23
2.2 进程的定义及特征.....24	24

2.2.1 进程的定义.....24	24
2.2.2 进程的特征.....24	24
2.2.3 进程和程序的关系.....25	25
2.3 进程的状态及描述.....25	25
2.3.1 进程的三种基本状态.....26	26
2.3.2 进程的创建状态和终止状态.....26	26
2.3.3 进程的挂起状态.....27	27
2.3.4 进程控制块.....28	28
2.4 进程控制.....30	30
2.4.1 进程创建.....31	31
2.4.2 进程撤销.....32	32
2.4.3 进程阻塞与唤醒.....32	32
2.4.4 进程的挂起与激活.....33	33
2.5 线程.....34	34
2.5.1 线程的概念.....34	34
2.5.2 线程与进程的比较.....36	36
2.6 小结.....36	36
练习题2.....38	38

第3章 进程同步与通信.....41

3.1 同步与互斥的基本概念.....41	41
3.1.1 临界资源与临界区.....41	41
3.1.2 同步与互斥的概念.....43	43
3.2 互斥的实现方法.....44	44
3.2.1 互斥算法.....44	44
3.2.2 硬件方法.....46	46
3.2.3 锁机制.....48	48
3.3 信号量.....49	49
3.3.1 信号量的描述.....49	49
3.3.2 利用信号量实现进程互斥.....50	50

3.3.3 利用信号量实现前趋关系.....	51	第5章 存储管理	101
3.3.4 经典进程同步问题.....	53	5.1 存储管理的基本概念	101
3.4 管程.....	58	5.1.1 存储管理的功能.....	101
3.4.1 管程的定义.....	58	5.1.2 存储分配的方式	101
3.4.2 用管程实现同步.....	59	5.1.3 重定位.....	102
3.5 进程通信.....	61	5.2 单一连续分配	104
3.5.1 进程通信的类型.....	61	5.3 分区存储管理	105
3.5.2 消息传递系统.....	61	5.3.1 固定分区.....	105
3.6 小结.....	64	5.3.2 动态分区分配	106
练习题3.....	66	5.3.3 可重定位分区分配	109
第4章 调度与死锁	69	5.3.4 分区的存储保护	111
4.1 调度的层次与性能评价	69	5.4 覆盖与交换	112
4.1.1 调度的层次.....	69	5.4.1 覆盖技术.....	112
4.1.2 调度性能的评价.....	70	5.4.2 交换技术.....	113
4.2 作业调度.....	72	5.5 分页存储管理	113
4.2.1 作业的状态及转换.....	72	5.5.1 分页实现思想	114
4.2.2 作业控制块及作业调度的功能	74	5.5.2 页表.....	114
4.3 进程调度.....	75	5.5.3 基本地址变换机构	115
4.3.1 进程调度的功能.....	75	5.5.4 具有快表的地址变换机构	116
4.3.2 进程调度的方式.....	76	5.5.5 多级页表.....	117
4.4 调度算法.....	76	5.6 分段存储管理与段页式存储管理	119
4.4.1 先来先服务调度算法.....	76	5.6.1 分段存储管理.....	119
4.4.2 短作业优先调度算法.....	78	5.6.2 段页式存储管理	121
4.4.3 优先级调度算法.....	79	5.7 虚拟存储器	122
4.4.4 时间片轮转调度算法.....	80	5.7.1 虚拟存储器定义	122
4.4.5 高响应比优先调度算法.....	82	5.7.2 请求分页存储管理	123
4.4.6 多级队列调度算法.....	82	5.7.3 请求分段存储管理	130
4.4.7 多级反馈队列调度算法.....	82	5.8 小结.....	131
4.5 死锁.....	84	练习题5.....	133
4.5.1 死锁的概念.....	84	第6章 设备管理	138
4.5.2 死锁产生的原因和必要条件.....	84	6.1 设备管理概述	138
4.5.3 处理死锁的基本方法.....	87	6.1.1 设备分类.....	138
4.5.4 死锁的预防.....	87	6.1.2 设备管理的任务和功能	139
4.5.5 死锁的避免.....	88	6.1.3 设备控制器与I/O通道.....	139
4.5.6 死锁的检测和解除.....	93	6.1.4 I/O系统结构.....	141
4.6 小结.....	95	6.2 输入/输出控制方式	143
练习题4.....	97	6.2.1 程序直接控制方式	143

6.2.2 中断控制方式	143
6.2.3 DMA 控制方式	143
6.2.4 通道控制方式	144
6.3 中断技术	144
6.3.1 中断的基本概念	144
6.3.2 中断的分类与优先级	145
6.3.3 中断处理过程	146
6.4 缓冲技术	146
6.4.1 缓冲的引入	146
6.4.2 单缓冲	147
6.4.3 双缓冲	148
6.4.4 循环缓冲	148
6.4.5 缓冲池	148
6.5 设备分配	149
6.5.1 设备分配中的数据结构	149
6.5.2 设备分配策略	150
6.5.3 设备分配程序	152
6.5.4 Spooling 系统	153
6.6 I/O 软件的层次结构	154
6.6.1 中断处理程序	155
6.6.2 设备驱动程序	155
6.6.3 与设备无关的软件	156
6.6.4 用户空间的软件	157
6.7 小结	159
练习题 6	159
第 7 章 文件管理	162
7.1 文件系统的概念	162
7.1.1 文件和文件系统	162
7.1.2 文件分类	163
7.2 文件结构与存储设备	164
7.2.1 文件的逻辑结构	164
7.2.2 文件的物理结构	165
7.2.3 文件的存取方法	165
7.2.4 文件的存储设备	166
7.3 文件存储空间的分配与管理	170
7.3.1 文件存储空间的分配	170
7.3.2 空闲存储空间的管理	173

7.4 文件目录管理	175
7.4.1 文件目录	175
7.4.2 单级目录结构	176
7.4.3 二级目录结构	177
7.4.4 多级目录结构	178
7.4.5 目录的查询方法	179
7.5 文件共享及文件管理的安全性	180
7.5.1 文件共享	180
7.5.2 文件保护	184
7.5.3 文件的转储和恢复	186
7.6 文件的使用	187
7.7 小结	188
练习题 7	189

第 8 章 UNIX 操作系统

193

8.1 UNIX 操作系统概述	193
8.1.1 UNIX 系统的历史	193
8.1.2 UNIX 系统的特点	194
8.1.3 UNIX 系统核心体系结构	194
8.2 UNIX 的进程描述	196
8.2.1 进程描述的数据结构	196
8.2.2 进程状态及其转换	198
8.2.3 进程上下文	199
8.3 进程控制与调度	201
8.3.1 进程控制	201
8.3.2 进程调度	203
8.3.3 进程切换	205
8.4 进程的同步与通信	206
8.4.1 软中断信号	206
8.4.2 管道	208
8.4.3 消息	210
8.4.4 共享存储区	213
8.4.5 信号量	215
8.5 存储管理	217
8.5.1 请求调页管理的数据结构	217
8.5.2 偷页进程	219
8.5.3 请求调页	220
8.6 设备管理	222

8.6.1 设备缓冲管理.....	222	8.9 文件系统的系统调用	236
8.6.2 核心与设备驱动程序的接口	225	8.9.1 系统调用 open	236
8.6.3 磁盘驱动程序.....	227	8.9.2 系统调用 creat	237
8.7 文件管理.....	228	8.9.3 系统调用 close.....	238
8.7.1 文件存储空间的管理	229	8.9.4 系统调用 link 和 unlink.....	238
8.7.2 文件的物理结构.....	230	8.9.5 系统调用 read 和 write	239
8.7.3 用户文件描述符表和文件表	232	8.10 小结	240
8.8 目录管理.....	233	练习题8	241
8.8.1 索引节点的管理.....	233	附录 各章习题参考答案.....	244
8.8.2 构造目录和删除目录.....	235	参考文献	270
8.8.3 检索目录.....	235		

第 1 章

操作系统引论

CHAPTER 01

操作系统是现代计算机系统中不可缺少的系统软件，是其他所有系统软件和应用软件的运行基础。操作系统控制和管理整个计算机系统软硬件资源，并为用户提供计算机提供一个方便灵活、安全可靠的工作环境。本章主要介绍操作系统的定义、操作系统的基本类型、操作系统的特征、操作系统的功能。

1.1 计算机系统概述

一个完整的计算机系统，不论是大型机、小型机还是微型机，都由两大部分组成：计算机硬件和计算机软件。计算机硬件是指计算机系统中由电子、机械、电气、光学和磁学等元器件构成的各种部件和设备，这些部件和设备依据计算机系统结构的要求组成一个有机整体。计算机软件是指由计算机硬件执行以完成一定任务的程序及其数据。计算机软件包括系统软件和应用软件，系统软件包括操作系统、编译程序、编辑程序、数据库管理系统等；应用软件是为各种应用目的而编制的程序。

计算机硬件主要由运算器、控制器、存储器、输入设备和输出设备组成，如图1.1所示。

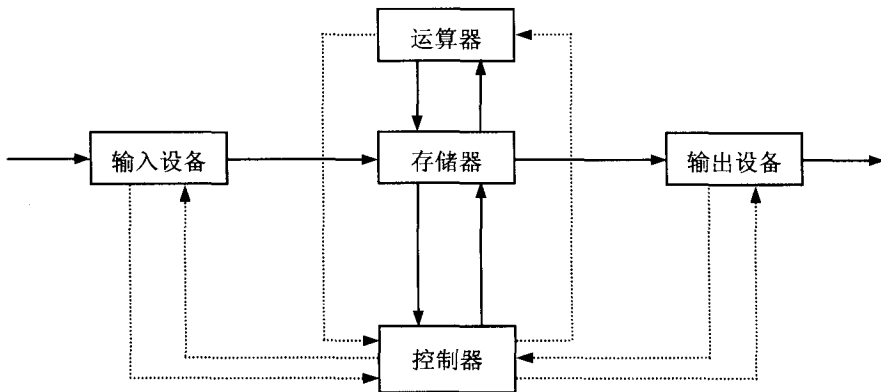


图 1.1 计算机的基本组成

运算器的主要功能是对数据进行算术运算和逻辑运算；存储器的主要功能是存储二进

制信息；控制器的主要功能是按照机器代码程序的要求，控制计算机各功能部件协调一致地工作，即从存储器中取出程序中的指令，对该指令进行分析和解释，并向其他功能部件发出执行该指令所需要的各种时序控制信号，然后再从存储器中取出下一条指令执行，如此连续运行下去，直到程序执行完为止。通常将控制器与运算器集成在一起，称为中央处理机。输入设备的主要功能是将用户信息（数据、程序等）变换为计算机能识别和处理的信息形式；输出设备的功能特点与输入设备正好相反，主要是将计算机中二进制信息变换为用户所需要并能识别的信息形式。

没有配置软件的计算机称为裸机（即计算机硬件），它仅仅构成了计算机系统的物质基础，而实际呈现在用户面前的计算机系统是经过若干层软件改造的计算机，如图1.2所示。

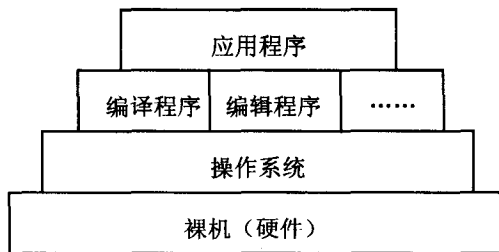


图 1.2 计算机系统的层次结构

从图1.2中可以看出，计算机的硬件和软件以及软件的各部分之间形成了一种层次结构的关系。计算机硬件在最下层，它的上面是操作系统，经过操作系统提供的资源管理功能和方便用户的各种服务功能把硬件改造成为功能更强、使用更方便的机器，通常将覆盖了软件的机器称为虚拟机或扩展机，而各种实用程序和应用程序运行在操作系统之上，它们以操作系统为支撑环境，同时向用户提供完成其工作所需的各种服务。

操作系统是计算机硬件之上的第一层软件，是对硬件功能的首次扩充。引入操作系统的目的是：提供一个计算机用户与计算机硬件系统之间的接口，使计算机系统更易于使用；有效地控制和管理计算机系统中的各种硬件和软件资源，使之得到更有效的利用；合理地组织计算机系统的工作流程，以改善系统性能。

1.2 操作系统的发展历史

如同任何其他事物一样，操作系统也有它的诞生、成长和发展历史。为了更清楚地把握操作系统的实质，了解操作系统的发展历史很有必要，因为操作系统的许多基本概念都是在操作系统的发展过程中出现并逐步得到发展和成熟的。下面让我们回顾一下操作系统的发展历史。

1.2.1 手工操作阶段

从1946年诞生第一台计算机起到20世纪50年代末，计算机处于第一代时期。此时构成计算机的主要元器件是电子管，计算机运算速度慢（只有几千次/秒），没有操作系统，甚

至没有任何软件,人们采用手工操作方式使用计算机。在手工操作方式下,用户一个接一个地轮流使用计算机,每个用户的使用过程大致如下:先将程序纸带(或卡片)装到输入机上,然后启动输入机把程序和数据送入计算机,接着通过控制台开关启动程序运行,当程序运行结束时,由用户取走纸带和计算结果。

从上述操作过程可以看出,程序运行期间计算机系统中的所有资源为单个用户独占,并且在程序运行过程中需要人工干预,以完成装卸纸带、拨动开关等操作。由此可见,手工操作方式具有用户独占计算机资源、资源利用率低及CPU等待人工操作的特点。

随着CPU速度的大幅提高,人工操作的慢速与CPU运算的快速之间出现了矛盾,这就是所谓的人机矛盾。例如,一个用户程序在速度为1万次/秒的计算机上运行需要1小时,人工操作时间需要3分钟,这种情况下人工操作时间与机器运行时间之比为1:20;若机器速度提高到60万次/秒,则该用户程序的运行时间降低为1分钟,而人工操作的速度不会有多大的提高,仍假定为3分钟,此时人工操作时间和机器运行时间之比为3:1。这就是说,人工操作时间远远超过了机器运行时间。由此可见,缩短人工操作时间就显得非常必要了。另一方面,CPU与I/O设备之间速度不匹配的矛盾也日益突出,为了缓和这些矛盾,引入了批处理技术及脱机输入/输出技术。

1.2.2 早期批处理

为了解决程序运行过程中的人工干预问题,需要缩短建立作业和人工操作的时间,人们提出了从一个作业到下一个作业的自动过渡方式,从而出现了批处理技术。完成作业自动过渡的程序称为监督程序,监督程序是一个常驻内存的程序,它管理作业的运行,负责装入和运行各种系统程序来完成作业的自动过渡。监督程序是最早的操作系统雏形。

批处理技术是指计算机系统对一批作业自动进行处理的一种技术。早期的批处理分为联机批处理和脱机批处理两种类型。

1. 联机批处理

在早期联机批处理系统中,操作员把用户提交的若干个作业集中成为一批,由监督程序先把它们输入到磁带上,当该批作业输入完成之后,监督程序就开始执行。它自动地把磁带上该批作业的第一个作业调入内存,并对该用户程序进行汇编或编译,然后由装配程序把汇编或编译结果装入内存,再启动执行。计算机完成该作业的全部计算或处理后,输出计算或处理结果。第一个作业完成之后,监督程序又自动地调入该批作业中的第二个作业,并重复上述执行过程,直到该批作业全部完成为止。在完成了一批作业之后,监督程序又控制输入另一批作业到磁带上,并按上述步骤重复处理。

2. 脱机批处理

在联机批处理中,作业的输入/输出都是联机的,也就是说将作业信息先送到磁带上,再从磁带上调入内存,然后将计算结果在打印机上输出,这些工作都在CPU的直接控制下进行。由于设备的输入/输出速度比CPU的速度慢得多,这种联机输入/输出操作影响了CPU速度的发挥,为此引入了脱机批处理系统。

在脱机批处理系统中,采用脱机输入/输出技术,即系统中除主机之外另设了一台外围

机(又称卫星机), 外围机只与外部设备打交道, 不与主机直接连接, 如图1.3所示。用户作业通过外围机输入到磁带上, 而主机只负责从磁带上把作业调入内存, 并予以执行。作业完成后, 主机负责把结果输出到磁带上, 然后再由外围机把磁带上的信息在打印机上输出。

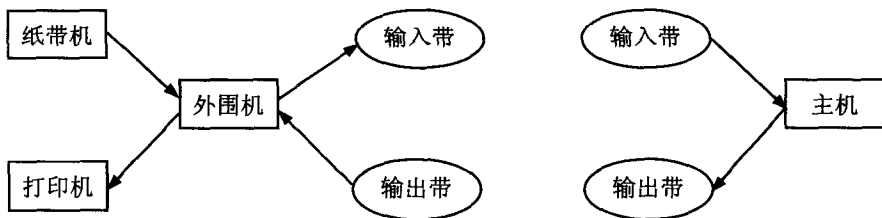


图 1.3 脱机输入/输出方式

脱机输入是指用户程序和数据在外围机的控制下, 预先从低速输入设备输入到磁带上, 当CPU需要这些程序和数据时, 再直接从磁带机高速输入到内存。脱机输出是指当程序运行完毕或告一段落, CPU需要输出时, 直接把计算结果高速地送到磁带上, 然后在外围机的控制下, 把磁带上的计算结果由相应的输出设备输出。

若输入/输出操作在主机控制下进行, 则称之为联机输入/输出。

脱机批处理系统中采用脱机输入/输出技术后, 低速I/O设备上数据的输入/输出都在外围机的控制下进行, 而CPU只与高速的磁带机打交道, 从而有效地减少了CPU等待慢速设备输入/输出的时间。

1.2.3 多道程序设计技术

在早期批处理系统中, 内存中仅有一道用户程序运行, 这种程序运行方式称为单道程序运行方式, 图1.4给出了单道程序运行的工作情况。

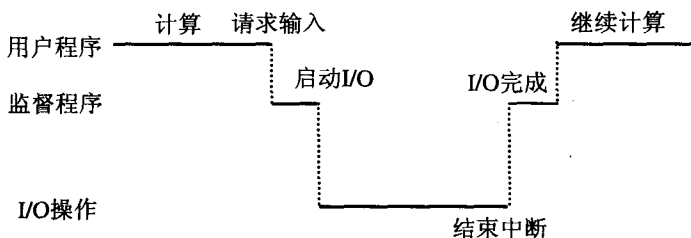


图 1.4 单道程序运行情况

从图1.4中可以看出, 当程序发出I/O请求时, CPU便处于等待状态, 此时CPU空闲, 当I/O完成后才继续运行程序, 因此单道程序系统中的CPU利用率不高。为进一步提高CPU的利用率, 引入了多道程序设计技术。

多道程序设计的基本思想是在内存中同时存放多道程序, 这些程序在管理程序的控制下交替运行, 共享处理机及系统中的其他资源。现代计算机系统一般都基于多道程序设计技术。图1.5给出了两道程序运行的工作情况。

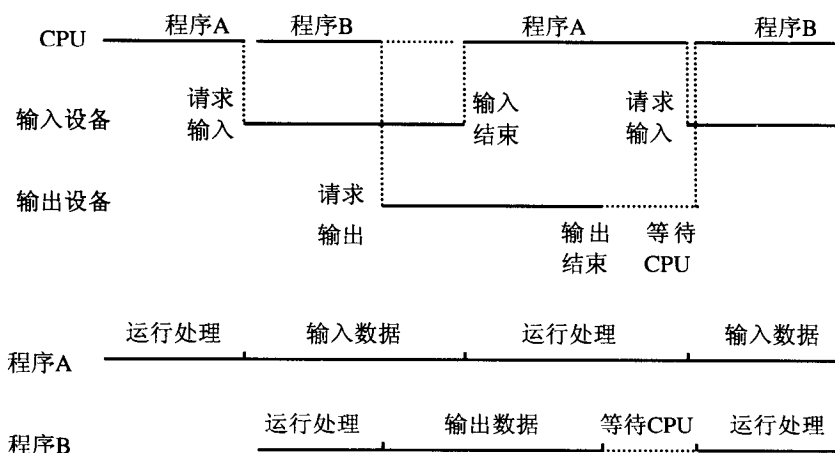


图 1.5 两道程序运行情况

从图1.5中可以看出，计算机系统中有A、B两道程序运行，它们的运行过程如下：

(1) 程序A先在CPU上运行，当程序A请求输入时，程序A停止运行；系统管理程序启动设备做输入工作，并将CPU分配给程序B。此时，程序A利用输入设备进行输入，而程序B正在CPU上执行。

(2) 当程序B请求输出时，程序B停止运行；系统管理程序启动输出设备做输出工作。此时，输入设备和输出设备都在工作，而CPU处于空闲状态。

(3) 当程序A请求的输入工作完成时，向系统发出I/O结束中断的请求，系统管理程序进行中断处理，然后调度程序A运行。此时，程序A在CPU上执行，程序B利用输出设备进行输出。

(4) 当程序B请求的输出工作完成时，向系统发出I/O结束中断的请求，系统管理程序进行中断处理，由于此时程序A在CPU上运行，故程序B处于等待CPU的状态。

(5) 当程序A再次请求输入时，程序A让出CPU；系统管理程序启动设备做输入工作，并再次调度程序B运行……

从A、B程序的执行过程可以看出，两个程序可以交替运行，若安排合适就会使CPU保持忙碌状态，而I/O设备也可满负荷工作。与单道程序运行情况相比，两道程序运行时系统资源利用率提高了，在一段给定的时间内计算机所能完成的总工作量也增加了。

综上所述，在单处理机计算机系统中多道程序运行的特点如下：

(1) 多道。计算机内存中同时存放多道相互独立的程序。

(2) 宏观上并行。同时进入系统的多道程序都处于运行过程中，即它们先后开始了各自的运行，但都未运行完毕。

(3) 微观上串行。内存中的多道程序轮流占有CPU，交替执行。

多道程序设计技术能有效提高系统的吞吐量和改善资源利用率，但实现多道程序系统时，由于内存中同时存在多道作业，因而还需要妥善解决下述一系列问题：

(1) 处理机管理问题。应如何在多道程序之间分配处理机,以使处理机既能满足各程序运行的需要又有较高的利用率;将处理机分配给某程序后,应何时收回等。

(2) 存储器管理问题。如何为每道程序分配必要的内存空间,使它们各自获得需要的存储空间又不致因相互重叠而丢失信息;应如何防止因某道程序出现异常而破坏其他程序等。

(3) 设备管理问题。多道程序共享系统中的多类I/O设备,应如何分配这些I/O设备,如何做到既方便用户使用设备又能提高设备的利用率等。

(4) 文件管理问题。现代计算机系统通常都存放有大量的文件,应如何组织这些文件才能既方便用户使用又能保证文件的安全性和一致性等。

1.2.4 操作系统的形成及发展

针对多道程序系统中存在的问题,人们研制了一组软件,利用这组软件来妥善有效地处理上述问题,这样便形成了操作系统。

操作系统是一组控制和管理计算机硬件和软件资源,合理地组织计算机工作流程,以及方便用户的程序的集合。

虽然操作系统已存在多年,但至今仍然没有一个统一的定义,对操作系统定义的不同说法从不同角度反映了操作系统的特征。值得说明的是:操作系统是一个系统软件,它由一组程序组成;操作系统的基本职能是控制和管理计算机系统内部的各种资源,有效地组织多道程序的运行;同时操作系统还提供众多服务功能,以方便用户使用计算机,并扩充硬件功能。

批处理系统缺少人机交互能力,因此用户使用不方便。为了解决这个问题,人们开发出分时系统。在分时系统中,一台主机可以连接几台乃至上百台终端,每个用户可以通过终端与主机交互,方便地编辑和调试自己的程序,向系统发出各种控制命令,请求完成某项工作;系统完成用户提出的要求,输出计算结果及出错、警告或提示等必要的信息。

为了满足某些应用领域对实时处理的需求,人们又开发出实时系统。实时系统具有专用性,不同的实时系统用于不同的应用领域。

近些年来,人们又开发出个人计算机操作系统、网络操作系统、分布式操作系统以及嵌入式操作系统等。伴随着硬件技术的飞速发展和应用领域的急剧扩大,操作系统不仅种类越来越多,而且功能更加强大,给广大用户提供了更为舒适的应用环境。

1.2.5 推动操作系统发展的动力

从操作系统形成至今的几十年间,其性能、规模、应用等方面都取得了飞速发展。推动操作系统发展的因素很多,但主要可归结为硬件技术更新和应用需求扩大两大方面。

1. 硬件技术更新

伴随计算机器件的更新换代——从电子管到晶体管、集成电路、大规模集成电路,直至当今的超大规模集成电路,计算机系统的性能得到快速提高,也促使操作系统的性能和结构有了显著提高。从没有软件,到早期的监督程序,发展成多道批处理系统、分时系统、

实时系统等。计算机体系结构的发展——从单处理器系统到多处理器系统，从指令串行结构到流水线结构、超级标量结构，从单总线到多总线应用等，这些发展有力地推动了操作系统的更大发展，如从单CPU操作系统发展到对称多处理器系统（SMP），从主机系统发展到个人机系统，从单自治系统发展到网络系统以及分布式系统。此外，硬件成本的下降也极大地推动了计算机技术的应用推广和普及。

2. 应用需求扩大

应用需求促进了计算机技术的发展，也促进了操作系统的不断更新升级。为了充分利用计算机系统内的各种宝贵资源，形成了早期的批处理系统；为了方便多个用户同时上机、实现友好的人机交互，形成了分时系统；为了实时地对特定任务进行可靠的处理，形成了实时系统；为了实现远程信息交换和资源共享，形成了网络系统及分布式系统等。在当今信息时代，信息处理离不开计算机，也就离不开操作系统这个软件平台。可以预见，操作系统将会以更快的速度更新换代。

1.3 操作系统的类型

根据操作系统具备的功能、特征、规模和所提供应用环境等方面的差异，可以将操作系统划分为不同类型。最基本的操作系统类型有三种，即批处理操作系统、分时操作系统和实时操作系统，分别简称为批处理系统、分时系统和实时系统。后来随着计算机体系结构的发展及应用需求的扩大，又出现了许多新型操作系统，主要有嵌入式操作系统、个人计算机操作系统、多处理机操作系统、网络操作系统和分布式操作系统。这些系统各有特点，适用于不同的应用领域。

1.3.1 批处理系统

描述任何一种操作系统都要用到作业的概念。所谓作业就是用户在一次解题或一个事务处理过程中要求计算机系统所做工作的集合，包括用户程序、所需的数据及命令等。

单道批处理系统是早期计算机系统中配置的一种操作系统类型，其工作流程大致如下：用户将作业交给系统操作员，系统操作员将多个用户作业组成一批输入并传送到外存储器；然后批处理系统按一定的原则选择其中的一个作业调入内存并使之运行；作业运行完成或出现错误而无法再进行下去时，由系统输出有关信息并调入下一个作业运行。重复上述过程，直至这批作业全部处理完毕为止。

单道批处理系统大大减少了人工操作的时间，提高了机器的利用率。但是对于某些作业来说，当它发出输入/输出请求后，CPU必须等待I/O的完成，这就意味着CPU空闲，特别是当I/O设备的速度较低时，将使CPU的利用率很低。为了提高CPU的利用率，引入了多道程序设计技术。

在批处理系统中引入多道程序设计技术就形成了多道批处理系统。在多道批处理系统中，不仅内存中可以同时有多道作业运行，而且作业可随时（不一定集中成批）被接受进