

计算机科学与技术专业规划教材



计算机操作系统

郑鹏 曾平 李蓉蓉 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

计算机操作系统/郑鹏,曾平,李蓉蓉编著. —武汉:武汉大学出版社,
2009. 12

计算机科学与技术专业规划教材

ISBN 978-7-307-07451-4

I. 计… II. ①郑… ②曾… ③李… 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(2009)第 218878 号

责任编辑:林 莉

责任校对:黄添生

版式设计:支 笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北金海印务公司

开本:787×1092 1/16 印张:17.75 字数:451千字 插页:1

版次:2009年12月第1版 2009年12月第1次印刷

ISBN 978-7-307-07451-4/TP·346 定价:30.00元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。



前言

操作系统是现代计算机系统中必不可少的基本系统软件,是其他所有系统软件和应用软件的运行基础,也是计算机专业教学的重要内容。操作系统主要用来管理和控制计算机系统的软、硬件资源,提高资源利用率,且为用户提供一个方便、灵活、安全和可靠地使用计算机的工作环境。

操作系统课程是计算机专业的一门重要的专业基础课,并从 2009 年开始作为计算机专业硕士研究生的统考课程之一。其特点是概念多、内容抽象、灵活性和综合性较强。本书结合现代操作系统的设计并考虑操作系统的发展方向,着重介绍了操作系统的基本概念、基本原理和基本技术,并按照计算机专业研究生入学考试大纲的要求,对教材内容进行了取舍和组织。

本书共分 11 章。第 1 章简要介绍操作系统的基本概念、功能、分类以及发展历史等。第 2 章主要讨论操作系统进程和线程管理的有关概念和技术,如进程的引入,进程与程序的区别,进程的特征,进程的控制,线程的引入,线程与进程的区别等。第 3 章讨论进程的同步和通信问题,包括实现同步和互斥的方法,进程通信的方式等。第 4 章主要介绍处理机的调度策略和死锁问题,包括调度原则,调度时机,调度算法,死锁的概念,死锁的原因,死锁的必要条件,解决死锁的方法等。第 5 章介绍存储管理技术,包括存储管理的基本概念,单一连续分配,分区分配,伙伴系统,覆盖和交换技术,分页管理,分段管理和段页式管理等。第 6 章讨论虚拟存储技术,包括虚拟存储器的概念,请求分页管理方法和请求分段管理方法。第 7 章讨论对输入/输出设备的控制和管理,包括输入/输出体系结构,输入/输出控制方式,中断技术,缓冲区管理等。第 8 章介绍文件系统,对文件逻辑组织,文件物理结构,文件目录,外存空间管理进行了讲解,讨论了磁盘的调度和控制。第 9 章简单讨论操作系统的安全问题,包括计算机系统安全的要求,操作系统安全评测标准,操作系统安全模型,面临的安全威胁,以及增强操作系统安全的方法。第 10 章简要介绍网络操作系统和分布式操作系统。第 11 章介绍两个操作系统实例 Windows 和 Linux 系统。

本书由郑鹏、曾平、李蓉蓉共同编写,由郑鹏对全书进行了整理和统编。

尽管本书几易其稿,但由于编著者水平有限,书中难免有错误和不妥之处,恳请广大读者批评指正。

在本教材的编写过程中,得到武汉大学出版社的大力支持和帮助,在此表示衷心感谢。

作者

2009 年 10 月



目 录



第 1 章 操作系统概论	1
1.1 操作系统的概念	1
1.1.1 用户观点	2
1.1.2 系统观点	3
1.2 操作系统的形成与发展	3
1.2.1 手工操作阶段	3
1.2.2 早期批处理	4
1.2.3 多道程序设计技术	5
1.2.4 操作系统的发展	6
1.3 操作系统的类型	7
1.3.1 批处理系统	7
1.3.2 分时系统	8
1.3.3 实时系统	9
1.3.4 其他操作系统类型	10
1.4 操作系统的特征	11
1.5 操作系统的作用与功能	12
1.5.1 操作系统的作用	12
1.5.2 操作系统的功能	13
1.6 操作系统的接口	14
1.6.1 命令接口	15
1.6.2 程序接口	16
1.6.3 图形用户接口	17
1.7 操作系统的运行环境和内核结构	18
1.7.1 操作系统的运行环境	18
1.7.2 操作系统的内核结构	18
1.8 小结	20
习题 1	21
第 2 章 进程与线程	23
2.1 进程的引入	23
2.1.1 前趋图	23
2.1.2 程序的顺序执行	24
2.1.3 程序的并发执行及特点	24

2.1.4	程序并发执行的条件	25
2.2	进程的定义及描述	26
2.2.1	进程的定义	26
2.2.2	进程的特征	26
2.2.3	进程和程序的关系	26
2.2.4	进程控制块	27
2.3	进程的状态和转换	28
2.3.1	进程的三种基本状态	28
2.3.2	进程的创建状态和退出状态	29
2.3.3	进程的挂起状态	29
2.4	进程的控制和管理	30
2.4.1	进程创建	31
2.4.2	进程撤销	32
2.4.3	进程阻塞与唤醒	32
2.4.4	进程的挂起与激活	33
2.5	进程的组织	34
2.6	线程	35
2.6.1	线程的概念	35
2.6.2	线程与进程的比较	37
2.7	小结	38
01	练习题 2	39
	第 3 章 进程同步与通信	43
3.1	同步与互斥的概念	43
3.1.1	临界资源与临界区	43
3.1.2	同步与互斥	45
3.2	互斥的实现方法	45
3.2.1	互斥算法	45
3.2.2	硬件方法	48
3.2.3	锁机制	50
3.3	信号量	51
3.3.1	信号量的定义	51
3.3.2	利用信号量实现进程互斥	52
3.3.3	利用信号量实现前趋关系	53
3.3.4	经典进程同步问题	55
3.3.5	信号量集机制	61
3.4	管程	62
3.4.1	管程的定义	62
3.4.2	用管程实现生产者—消费者问题	63
3.4.3	用管程实现哲学家进餐问题	64



3.5 进程通信	66
3.5.1 进程通信的类型	66
3.5.2 消息传递系统	66
3.6 小结	69
练习题 3	70
第 4 章 调度与死锁	74
4.1 调度的层次	74
4.1.1 作业调度	75
4.1.2 进程调度	75
4.1.3 中级调度	75
4.1.4 调度性能的评价	75
4.2 作业调度	77
4.2.1 作业的状态及转换	77
4.2.2 作业调度	78
4.3 进程调度	79
4.3.1 进程调度的功能	79
4.3.2 进程调度的方式	80
4.4 调度算法	80
4.4.1 先来先服务调度算法	80
4.4.2 短作业优先调度算法	81
4.4.3 优先级调度算法	82
4.4.4 时间片轮转调度算法	83
4.4.5 高响应比优先调度算法	85
4.4.6 多级队列调度算法	85
4.4.7 多级反馈队列调度算法	86
4.5 死锁	87
4.5.1 死锁的概念	87
4.5.2 死锁产生的原因和必要条件	88
4.5.3 处理死锁的基本方法	90
4.5.4 死锁的预防	90
4.5.5 死锁的避免	91
4.5.6 死锁的检测和解除	95
4.6 小结	97
练习题 4	98
第 5 章 存储器管理	103
5.1 存储管理的基本概念	103
5.1.1 程序的装入	103
5.1.2 程序的链接	106

5.1.3	内存保护	107
5.2	单一连续分配	108
5.3	分区分配	108
5.3.1	固定分区	109
5.3.2	动态分区分配	109
5.3.3	可重定位分区分配	112
5.4	伙伴系统	114
5.5	覆盖与交换技术	115
5.5.1	覆盖技术	115
5.5.2	交换技术	116
5.6	分页存储管理	117
5.6.1	分页实现思想	117
5.6.2	页表	118
5.6.3	基本地址变换机构	118
5.6.4	具有快表的地址变换机构	119
5.7	分段存储管理	122
5.7.1	分段实现思想	122
5.7.2	段表及地址变换	123
5.7.3	分段与分页的区别	123
5.8	段页式存储管理	124
5.9	小结	125
	练习题 5	127
第 6 章 虚拟存储器		130
6.1	虚拟存储器概念	130
6.2	请求分页存储管理	131
6.2.1	请求分页存储管理的实现思想	131
6.2.2	页表	132
6.2.3	缺页中断与地址变换	132
6.2.4	页面分配和置换策略	134
6.2.5	页面置换算法	135
6.2.6	页面大小的选择	140
6.2.7	工作集和抖动	140
6.2.8	页的共享与保护	142
6.3	请求分段存储管理	143
6.3.1	请求分段存储管理的实现思想	143
6.3.2	段的共享与保	144
6.3.3	虚拟段页式存储管理	145
6.4	小结	145
	练习题 6	146



第 7 章 设备管理	149
7.1 设备管理概述	149
7.1.1 设备分类	149
7.1.2 设备管理的任务和功能	150
7.1.3 设备控制器与 I/O 通道	150
7.1.4 I/O 系统结构	152
7.2 输入/输出控制方式	153
7.2.1 程序直接控制方式	154
7.2.2 中断控制方式	154
7.2.3 DMA 控制方式	154
7.2.4 通道控制方式	155
7.3 中断技术	156
7.3.1 中断的基本概念	156
7.3.2 中断的分类与优先级	156
7.3.3 中断处理过程	157
7.4 缓冲技术	157
7.4.1 缓冲的引入	157
7.4.2 单缓冲	158
7.4.3 双缓冲	158
7.4.4 循环缓冲	159
7.4.5 缓冲池	159
7.5 设备分配	160
7.5.1 设备分配中的数据结构	160
7.5.2 设备分配策略	161
7.5.3 设备分配程序	163
7.5.4 Spooling 系统	164
7.6 I/O 软件的层次结构	165
7.6.1 中断处理程序	165
7.6.2 设备驱动程序	165
7.6.3 与设备无关的软件	166
7.6.4 用户空间的软件	167
7.7 小结	169
练习题 7	169
第 8 章 文件管理	172
8.1 文件系统的概念	172
8.1.1 文件和文件系统	172
8.1.2 文件分类	173
8.2 文件结构与存储设备	174
8.2.1 文件的逻辑结构	174



8.2.2	文件的物理结构	175
8.2.3	文件的存取方法	176
8.2.4	文件的存储设备	176
8.3	文件存储空间的分配与管理	181
8.3.1	文件存储空间的分配	181
8.3.2	空闲存储空间的管理	184
8.4	文件目录管理	186
8.4.1	文件目录	186
8.4.2	单级目录结构	187
8.4.3	二级目录结构	188
8.4.4	多级目录结构	188
8.5	文件共享及文件管理的安全性	190
8.5.1	文件共享	190
8.5.2	文件保护	193
8.5.3	文件的转储和恢复	196
8.6	文件的使用	197
8.7	小结	197
	练习题 8	198
第 9 章 操作系统安全		202
9.1	安全性概述	202
9.1.1	基本概念	203
9.1.2	漏洞扫描	204
9.1.3	安全评测	204
9.1.4	评测标准	205
9.2	安全模型	207
9.3	安全威胁	209
9.3.1	入侵检测	210
9.3.2	计算机病毒	211
9.4	保护	212
9.4.1	存储器的保护	212
9.4.2	文件的保护	214
9.5	操作系统安全增强的实现方法	215
9.5.1	安全威胁的来源	215
9.5.2	安全操作系统的实现方法	216
9.5.3	安全操作系统的一般开发过程	217
9.5.4	操作系统近年来受到重视的安全增强技术	217
9.5.5	安全操作系统的设计原则	219
9.6	访问控制	220



9.6.1 访问控制的概念	220
9.6.2 自主访问控制	223
9.6.3 强制访问控制	224
9.6.4 角色访问控制	225
9.7 小结	226
练习题 9	226

第 10 章 多处理机操作系统 228

10.1 网络操作系统	228
10.1.1 网络的基本概念	228
10.1.2 网络操作系统的基本概念	229
10.1.3 基本通信技术	230
10.1.4 网络文件系统	232
10.1.5 数据和文件资源的共享	232
10.2 分布式操作系统	233
10.2.1 分布式系统概述	233
10.2.2 分布式进程通信	234
10.2.3 分布式资源管理	236
10.2.4 分布式进程同步	238
10.2.5 分布式系统中的死锁	240
10.2.6 分布式文件系统	242
10.2.7 分布式进程迁移	243
10.2.8 计算机集群	244
10.3 小结	245
练习题 10	245

第 11 章 操作系统实例简介 248

11.1 Windows 操作系统	248
11.1.1 Windows 发展历程	248
11.1.2 Windows 2000/XP 的构成	249
11.1.3 Windows 进程管理	251
11.1.4 Windows 内存管理	255
11.1.5 Windows 设备管理	258
11.1.6 Windows 文件系统	259
11.2 Linux 操作系统	262
11.2.1 Linux 系统发展历史	262
11.2.2 Linux 进程管理	263
11.2.3 Linux 内存管理	265
11.2.4 Linux 设备管理	268
11.2.5 Linux 文件系统	268



11.3 小结	271
练习题 11	271

参考文献	273
------	-----

第10章 操作系统的发展与未来	275
10.1 操作系统的发展	275
10.2 操作系统的未来	275
10.3 操作系统的未来	275
10.4 操作系统的未来	275
10.5 操作系统的未来	275
10.6 操作系统的未来	275
10.7 操作系统的未来	275
10.8 操作系统的未来	275
10.9 操作系统的未来	275
10.10 操作系统的未来	275
10.11 操作系统的未来	275
10.12 操作系统的未来	275
10.13 操作系统的未来	275
10.14 操作系统的未来	275
10.15 操作系统的未来	275
10.16 操作系统的未来	275
10.17 操作系统的未来	275
10.18 操作系统的未来	275
10.19 操作系统的未来	275
10.20 操作系统的未来	275

第11章 操作系统的安全	275
11.1 操作系统的安全	275
11.2 操作系统的安全	275
11.3 操作系统的安全	275
11.4 操作系统的安全	275
11.5 操作系统的安全	275
11.6 操作系统的安全	275
11.7 操作系统的安全	275
11.8 操作系统的安全	275
11.9 操作系统的安全	275
11.10 操作系统的安全	275
11.11 操作系统的安全	275
11.12 操作系统的安全	275
11.13 操作系统的安全	275
11.14 操作系统的安全	275
11.15 操作系统的安全	275
11.16 操作系统的安全	275
11.17 操作系统的安全	275
11.18 操作系统的安全	275
11.19 操作系统的安全	275
11.20 操作系统的安全	275



第1章 操作系统概论

操作系统是伴随着计算机系统的发展,逐步形成、发展和成熟起来的,它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中的软硬件资源,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。本章主要介绍操作系统的基本概念,包括操作系统的特征、操作系统的作用与功能、操作系统的发展过程与分类、操作系统的运行环境等。

1.1 操作系统的概念

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。

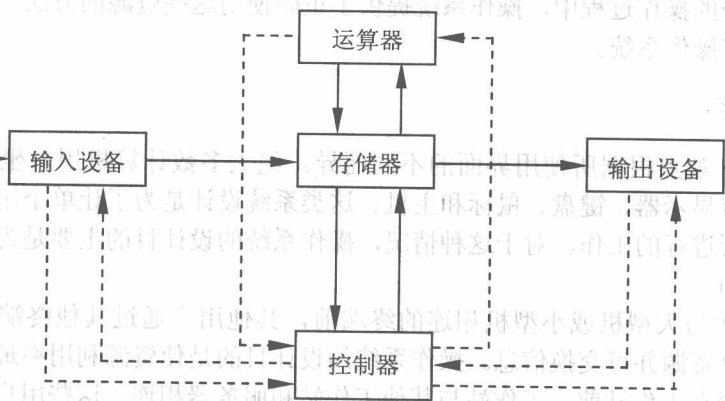


图1.1 计算机的基本组成

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，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。

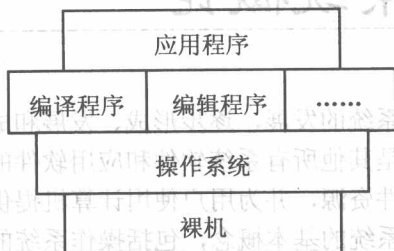


图 1.2 计算机系统的层次关系

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。裸机在最下层，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裸机改造成为功能更强、使用更方便的机器，通常将裸机之上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。

操作系统是裸机上的第一层软件，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。

在计算机系统的操作过程中，操作系统提供了正确使用这些资源的方法。可通过用户或系统的观点来研究操作系统。

1.1.1 用户观点

计算机的用户观点根据所使用界面的不同而异。绝大多数计算机用户坐在个人计算机前，个人计算机有显示器、键盘、鼠标和主机。这类系统设计是为了让单个用户单独使用其资源，优化用户所进行的工作。对于这种情况，操作系统的设计目的主要是为了用户使用方便，性能是次要的。

有些用户坐在与大型机或小型机相连的终端前，其他用户通过其他终端访问同一计算机。这些用户共享资源并可交换信息。操作系统的设计目的是使资源利用率最大化。

另一些用户坐在工作站前，工作站与其他工作站和服务器相连。这些用户不仅可以使用的资源，而且还可以使用共享资源。这类操作系统的设计目的是个人可用性和资源利用率的折中。

近年来，出现了许多类型的手持计算机，它们大多为单个用户所独立使用。有的也通过有线或无线与网络相连。由于受电源和接口限制，它们只能执行相对少的远程操作。这类操作系统的设计目的主要是个人可用性和电源管理。有的计算机几乎没有或根本没有用户观点。如家电和汽车中所使用的嵌入式计算机，这些设备及其操作系统通常设计成无需用户干预就能执行。



1.1.2 系统观点

从计算机系统的角度看,操作系统是与计算机硬件最为密切的程序。可以将操作系统看做资源分配器,它是资源管理者。

在计算机系统中有两类资源:硬件资源和软件资源,按其作用又可以将它们分为四大类资源:处理机、存储器、外部设备和文件(程序和数据)。这四类资源构成了操作系统本身和用户作业赖以活动的物质基础和工作环境。它们的使用方法和策略决定了整个操作系统的规模、类型、功能和实现。例如面对许多甚至冲突的资源请求,操作系统必须决定如何为各个程序和用户分配资源,以便计算机系统能正常有效地运行。

现代计算机系统通常是多道程序系统,所谓资源管理,主要指的是在多道程序之间合理地分配和回收各种资源,使资源得到充分有效的使用,使多道程序得以有条不紊地运行。资源管理包括用以下两种方式实现多路复用资源:时间复用和空间复用。当一种资源在时间上复用时,不同的用户或程序可以轮流使用它。空间复用是指每个客户同时得到资源的一部分,而不用轮流排队。例如系统中通常只有一个CPU,对CPU的管理,可采用时间复用的方式,对内存等资源,可采用空间复用的方式。

1.2 操作系统的形成与发展

要理解什么是操作系统,必须首先弄清楚操作系统是如何发展的。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。

操作系统的发展过程是一个从无到有、从简单到复杂的过程。下面我们将从最初的系统开始,经过批处理、多道程序系统、分时系统和实时系统,到现代操作系统,以此来追寻操作系统的发展。

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。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(只有几千次每秒),硬件价格昂贵,没有操作系统,甚至没有任何软件,人们采用手工操作方式操作计算机。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该用户独占,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,这种情况下操作时间和运行时间的比为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运行时间的比为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，一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，作业信息先送到磁带，再由磁带调入内存，以及计算结果在打印机上输出，这些都是由CPU来处理的，这种联机输入/输出的缺点是速度慢。为此，在批处理技术中引进了脱机输入/输出技术。

在脱机批处理系统中，除主机之外另设了一台外围机（又称卫星机），该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。

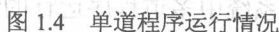
脱机输入是指将用户程序和数据在外围机的控制下，预先从低速输入设备输入到磁带上，当CPU需要这些程序和数据时，再直接从磁带机高速输入到内存。脱机输出是指当程序运行完毕或告一段落，CPU需要输出时，无需直接把计算结果送至低速输出设备，而是高速地把结果送到磁带上，然后在外围机的控制下，把磁带上的计算结果由相应的输出设备输出。

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。



图 1.3 脱机输入/输出方式



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



(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。系统管理程序向

(2) 半程序化技术给山叶、积层、倍十、三行、系统管理程序、方寸给山叶与做给十、

(3) 当程序 A 请求的输入工作完成时, 向系统发 I/O 结束中断, 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 I/O 结束中断, 系统管理程序进行中断处理, 由于此时程序 A 在 CPU 上运行, 则程序 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。