

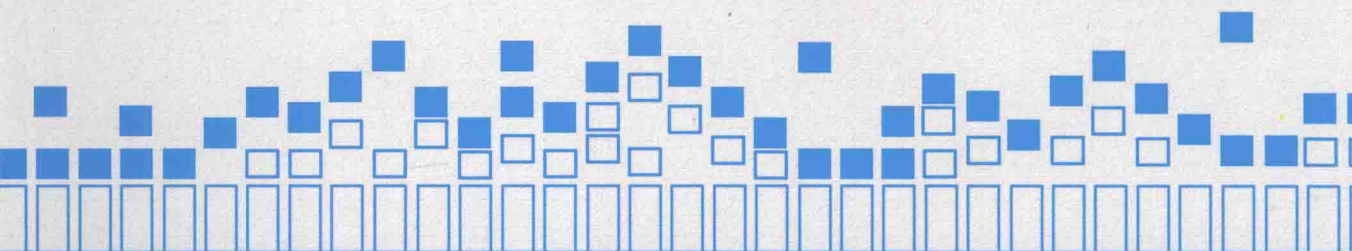


普通高等教育“十二五”规划教材
全国高职高专规划教材·计算机系列

CAOZUO XITONG JICHU YU SHIJIAN

操作系统基础与实践

汤 敏 刘 均◎主编



北京大学出版社
PEKING UNIVERSITY PRESS



普通高等教育“十二五”规划教材
全国高职高专规划教材·计算机系列

操作系统基础与实践

主 编 汤 敏 刘 均

副主编 廖仕东 聂 敏 何 静



北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书共有7章,第1章为操作系统概述;第2~6章,分别介绍操作系统的五大功能,它们是:第2章作业管理,第3章处理机管理,第4章存储管理,第5章文件系统,第6章设备管理;第7章为操作系统实践,包括基础篇和拓展篇两部分,以求加深对操作系统五大功能的理解。

本书可作为高职高专院校计算机专业相关课程的教学用书,也可以作为计算机爱好者学习操作系统的入门参考用书。

图书在版编目(CIP)数据

操作系统基础与实践/汤敏,刘均主编. —北京:北京大学出版社,2012.9

(全国高职高专规划教材·计算机系列)

ISBN 978-7-301-20693-5

I. ①操… II. ①汤…②刘… III. ①操作系统—高等职业教育—教材 IV. ①TP316

中国版本图书馆 CIP 数据核字(2012)第 104181 号

书 名: 操作系统基础与实践

著作责任者: 汤 敏 刘 均 主编

策 划 编 辑: 桂 春

责 任 编 辑: 桂 春

标 准 书 号: ISBN 978-7-301-20693-5/TP·1224

出 版 发 行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn>

电 子 信 箱: zyjy@pup.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62765126 出版部 62754962

印 刷 者: 三河市北燕印装有限公司

经 销 者: 新华书店

787 毫米×1092 毫米 16 开本 12.25 印张 313 千字

2012 年 9 月第 1 版 2012 年 9 月第 1 次印刷

定 价: 26.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话: (010)62752024 电子信箱: fd@pup.pku.edu.cn

前 言

随着科学技术的飞速发展和计算机应用的日益普及,人类已经进入计算机广泛应用的信息时代,学习和掌握操作系统的基础知识和基本技能已经成为信息社会对高职高专计算机专业学生的必然要求。

本书根据高职高专教育“理论够用,注重实践”、理论与实践相结合的原则编写而成,反映了高职高专计算机专业课程教学改革的最新成果。

本书层次分明、重点突出,简要介绍了操作系统概念、功能以及实现技术;通过 Windows 2000 操作系统的使用,介绍了操作系统的实现技术和具体应用,使得学生通过学习掌握操作系统的基本知识;每章都配备有适当的习题,帮助学生消化并掌握操作系统知识;还设置有针对性的实训项目,作为课程实验的参考。

本书的第 1、2、3 章及习题由汤敏编写,第 4、5、6 章由聂敏、刘均编写,第 7 章由何静、廖仕东编写。全书由汤敏统稿。

由于编者水平有限,错误与不妥之处在所难免,敬请广大读者批评指正,以便我们改进、完善计算机操作系统的教学体系,谢谢!

编 者

2012 年 8 月

目 录

第 1 章 操作系统概述	(1)
1.1 计算机系统	(1)
1.2 操作系统	(2)
1.2.1 实用操作系统	(2)
1.2.2 操作系统的定义	(3)
1.3 操作系统的功能	(4)
1.3.1 操作系统的功能	(4)
1.3.2 操作系统设计原则	(5)
1.3.3 操作系统的发展	(5)
1.4 操作系统的分类	(8)
1.4.1 批处理系统	(9)
1.4.2 分时系统	(9)
1.4.3 实时系统	(11)
1.4.4 分布式操作系统	(12)
1.4.5 网络操作系统	(14)
1.5 操作系统的结构	(16)
1.5.1 操作系统的观点	(16)
1.5.2 操作系统的层次结构	(19)
1.6 小结	(20)
习题一	(20)
第 2 章 作业管理	(23)
2.1 用户界面	(23)
2.1.1 作业控制语言	(23)
2.1.2 作业控制命令	(23)
2.1.3 菜单控制	(24)
2.1.4 窗口和图标	(25)
2.1.5 系统调用	(26)
2.2 作业状态与作业管理	(26)
2.2.1 作业的状态	(26)
2.2.2 作业控制块	(27)
2.2.3 作业调度程序	(28)
2.3 作业调度	(28)

2.3.1	调度算法设计原则	(28)
2.3.2	作业调度算法衡量指标	(29)
2.3.3	常用作业调度算法	(29)
2.4	终端作业的管理	(32)
2.4.1	命令语言	(32)
2.4.2	终端作业的控制	(32)
2.5	小结	(33)
	习题二	(34)
第3章	处理机管理	(37)
3.1	概述	(37)
3.1.1	多用户	(37)
3.1.2	程序的顺序执行	(37)
3.1.3	程序的并发执行	(38)
3.2	进程及其状态转换	(39)
3.2.1	进程的定义	(39)
3.2.2	进程的状态及其转换	(40)
3.2.3	进程控制块	(42)
3.2.4	线程	(44)
3.3	进程的控制与调度	(44)
3.3.1	进程控制原语	(44)
3.3.2	进程调度	(46)
3.3.3	Linux 中的进程控制	(47)
3.4	进程的同步与互斥	(48)
3.4.1	同步与互斥的概念	(48)
3.4.2	临界资源与临界区	(49)
3.4.3	互斥与同步的实现	(51)
3.5	进程通信	(54)
3.5.1	send(B,M)原语	(55)
3.5.2	receive(B,X)原语	(55)
3.6	死锁	(57)
3.6.1	死锁的形成	(57)
3.6.2	死锁的定义	(58)
3.6.3	死锁的防止	(59)
3.6.4	死锁的避免	(60)
3.6.5	死锁的检测	(61)
3.7	进程、程序、作业与任务	(61)
3.8	小结	(62)
	习题三	(62)

第4章 存储器管理	(67)
4.1 存储管理概述	(67)
4.2 地址映射	(68)
4.2.1 逻辑地址	(68)
4.2.2 物理地址	(68)
4.2.3 地址映射方式	(69)
4.3 连续分配存储管理	(69)
4.3.1 单一连续分区存储管理	(69)
4.3.2 固定分区存储管理	(71)
4.4 可变分区存储管理	(72)
4.4.1 可变分区概念	(72)
4.4.2 可变分区分配	(72)
4.4.3 空闲分区的分配算法	(73)
4.4.4 地址转换与存储保护	(74)
4.4.5 可变分区存储管理的特点	(75)
4.5 覆盖与交换	(76)
4.5.1 覆盖(overlay)	(76)
4.5.2 交换(swapping)	(76)
4.6 分页存储管理	(76)
4.6.1 实现原理	(77)
4.6.2 存储分配	(78)
4.6.3 碎片概念	(78)
4.7 分段存储管理	(79)
4.7.1 分段引入	(79)
4.7.2 实现原理	(79)
4.7.3 分段与可变分区的区别	(80)
4.7.4 分段与分页的区别	(80)
4.7.5 分段存储管理的特点	(81)
4.8 段页式存储管理	(81)
4.8.1 实现原理	(81)
4.8.2 数据结构	(81)
4.8.3 地址映射	(82)
4.9 虚拟存储器管理	(82)
4.9.1 基本概念	(82)
4.9.2 请求分页存储管理	(83)
4.10 常用操作系统的存储方案	(89)
4.10.1 DOS 操作系统存储方案	(89)
4.10.2 Windows 操作系统的存储管理	(90)

4.10.3	Linux 操作系统的存储管理	(90)
4.11	小结	(90)
	习题四	(92)
第 5 章	文件系统	(95)
5.1	文件和文件系统的基本概念	(95)
5.1.1	文件	(95)
5.1.2	文件系统	(96)
5.1.3	文件系统结构和存取方法	(97)
5.1.4	文件的物理结构和存储介质	(98)
5.1.5	UNIX 系统的文件物理结构	(102)
5.2	文件目录管理	(102)
5.2.1	基本概念	(102)
5.2.2	一级目录结构	(103)
5.2.3	二级目录结构	(103)
5.2.4	多级目录结构(树型目录)	(104)
5.3	文件存储空间管理	(105)
5.3.1	位示图	(105)
5.3.2	空闲区表	(105)
5.3.3	空闲块链	(106)
5.3.4	常用操作系统对磁盘存储空间的管理	(107)
5.4	文件的操作	(109)
5.4.1	创建文件	(109)
5.4.2	打开文件	(109)
5.4.3	关闭文件	(109)
5.4.4	删除文件	(110)
5.4.5	读写文件	(110)
5.5	文件的共享和保护	(110)
5.5.1	文件的共享	(110)
5.5.2	文件的保护	(110)
	习题五	(111)
第 6 章	设备管理	(114)
6.1	概述	(114)
6.1.1	计算机外部设备的分类	(114)
6.1.2	设备管理的功能与目标	(116)
6.2	外部设备输入/输出控制方式	(117)
6.2.1	程序直接控制方式	(117)
6.2.2	中断控制输入/输出方式	(118)
6.2.3	直接存储器存取方式	(120)

6.2.4 通道方式	(122)
6.3 设备分配	(123)
6.3.1 设备分配中的数据结构	(123)
6.3.2 设备分配思想	(125)
6.3.3 SPOOLing 技术	(127)
6.4 中断技术	(128)
6.4.1 中断及与中断相关的基础知识	(128)
6.4.2 硬中断、内中断和软中断	(129)
6.4.3 中断优先级	(129)
6.4.4 中断处理过程	(130)
6.5 缓冲技术	(130)
6.5.1 缓冲技术的引入	(130)
6.5.2 缓冲技术的实现方法	(131)
6.5.3 缓冲的种类及工作过程	(131)
6.6 设备驱动程序	(134)
6.6.1 设备驱动程序的引入	(134)
6.6.2 设备控制器	(134)
6.6.3 设备驱动程序的工作	(134)
6.7 小结	(135)
习题六	(135)
第 7 章 操作系统实践	(138)
基础篇	(138)
实验一 了解 Windows XP 的系统信息及注册表	(138)
实验二 进程管理	(139)
实验三 存储管理	(141)
实验四 磁盘调度算法	(158)
实验五 银行家算法	(160)
拓展篇	(162)
实验一 Linux 及其使用环境	(162)
实验二 shell 编程	(165)
实验三 进程互斥	(167)
实验四 进程通信——信号	(170)
实验五 proc 文件系统	(174)
实验六 模拟进程调度方法	(175)
实验七 Linux 内核编译	(176)
实验八 添加系统调用	(177)
实验九 Linux 字符设备驱动程序	(180)
参考文献	(186)

第 1 章 操作系统概述

【本章导读】 计算机系统是由硬件和软件两部分组成,操作系统是与计算机关系最为密切的系统软件,它是硬件的第一层扩充,是其他软件运行的基础,是系统的控制中心,管理着系统的所有资源,具有作业管理、处理机管理、存储管理、文件管理和设备管理的功能。

操作系统是在人们使用计算机的过程中,为了满足提高资源利用率和增强计算机系统性能的需求,伴随着计算机技术本身及其应用的日益发展,而逐步地形成和完善起来的。在其发展过程中,多道程序设计技术起了关键性的作用。研究操作系统有不同的观点,包括资源管理观点、用户管理观点、进程管理观点等。

1.1 计算机系统

计算机系统是能按人的要求接收和存储信息,自动进行数据处理和计算,并输出结果信息的机器系统。计算机系统由硬件系统和软件系统组成。硬件系统是借助电、磁、光、机械等原理构成的各种物理部件的有机组合,是系统赖以工作的实体。软件系统是由各种程序和数据组成,用于指挥全系统按指定的要求进行工作。如图 1.1 所示。



图 1.1 计算机系统的层次结构

最内层是硬件系统,主要由中央处理器(CPU)、主存储器、输入/输出控制系统和各种外围设备组成。中央处理器是对信息进行高速运算和处理的部件;主存储器用于存放程序和数据,它可被中央处理器直接访问;输入/输出(I/O)控制系统管理外围设备(如键盘、显示器、打印机、磁带机、磁盘机等)与主存储器之间的信息传送。最外层是使用计算机的人,人与硬件系统之间的接口界面是软件系统,软件系统为人们使用计算机提供方便,软件系统包括系统软件、支援软件和应用软件三部分。系统软件有操作系统、编译系统等,操作系统的功能是实现资源的管理和控制程序的执行;编译系统的功能是把高级语言(如 Fortran、Pascal 等)或汇编语言所编写的源程序翻译成机器可执行的由机器语言(指令)表示的目标程序。各种接口软件、软件开发工具等都是支援软件,它支援其他软件的编制和维护。应用软

件是按某种需要而编写的专用程序。系统软件、支援软件和应用软件既有分工又有结合,并不能截然分开。例如,操作系统既可看做系统软件,又可看做支援软件。有时在一个系统中是系统软件,而在另一个系统中就成为支援软件。

1.2 操作系统

1.2.1 实用操作系统

从市场上购买回来的电脑其实不单纯是一堆组合了的计算机的硬件,除那些摸得着的物体外,计算机硬盘上还安装了大量的软件,其中最重要的是计算机操作系统。

对于个人计算机(PC)用户来说,最常见的情况是,打开计算机的电源后,等待显示屏上闪烁的文字、图像逐渐稳定下来。稳定以后我们看到了 Windows XP 所展示的任务桌面(见图1.2),上面有不同的图标分别代表着不同的功能,还有打开的窗口代表用户正在运行的任务。最下面一行是一些按钮和状态显示,由“开始”按钮可以引出各种各样的可执行任务,任务栏中的按钮代表着正在执行的任务,按下其中的按钮能在桌面上弹出对应的窗口;还有一些状态显示:时间,输入法,计划任务,声音等。



图 1.2 Windows XP 的桌面

用户可以通过双击桌面图标或者单击任务栏“开始”按钮选择要执行的程序来实现计算机的使用功能。双击图标会立即执行任务,例如,单击“开始”→“程序”→“附件”→“游戏”→“红心大战”即可打开游戏;双击“我的电脑”可以查看计算机安装了哪些软件;双击“控制面板”可以查看设备的设置情况。

使用 Windows XP 的桌面系统,一切都很方便,只要有一个鼠标,就可以做任何操作了。各种硬件和软件在后台有序地完成着各自的工作,而这一切并不需要用户去操心。那么,到底是谁在提供这种方便,谁在背后进行操作呢?这就是操作系统。

另外一个值得一提的系统是 Linux。许多计算机用户可以通过该系统连接在一起,共享计算机的资源,还能够相互打电话:通话的双方同时使用指令“TALK USERNAME”来

要求系统接通电话,直到看到系统连接成功的提示,通过各自的键盘和显示屏就可以相互对话。

如果不需要直接对话,还可以采用写信的方式,这就是电子邮件系统。只要在适当的位置输入信件的内容和信件接收者的名称,然后单击“发送”按钮,就可以成功地发送信件。

Linux 系统还可以实现许多其他的功能:它可以通过监视系统来了解每个用户的工作,还可以通过管理系统来确定给用户的权利,甚至可以控制用户行为等。这种监视、管理和控制是谁来实现的呢?还是操作系统,只不过这时的操作系统称为多用户操作系统。

因此,计算机操作系统是一个幕后管理和控制系统,它管理着计算机上的所有资源,包括硬件、软件、数据;提供某种方法让用户方便地使用计算机;对计算机及用户的行为进行控制。如果买回来的计算机不带有操作系统,就好像人没有大脑一样是无法指挥各个部件进行工作的。因此,操作系统是计算机软件中最核心的部分,没有它,普通用户基本上无法使用计算机。那么,操作系统到底应该具有哪些功能才能满足设计人员及普通用户的需要呢?这是下面要讨论的问题。

1.2.2 操作系统的定义

归纳起来,操作系统有如下几个特点。

(1) 操作系统是程序的集合。从形式上讲,操作系统只不过是存放在计算机中的程序。这些程序一部分存放在内存中,一部分存放在硬盘上,中央处理机在适当的时候调用这些程序,以实现所需要的功能。

(2) 操作系统管理和控制系统资源。计算机的硬件、软件、数据等都需要操作系统的管理。操作系统通过许多数据结构,对系统的信息进行记录,根据不同的系统要求,对系统数据进行修改,达到对资源进行控制的目的。

(3) 操作系统提供了方便用户使用计算机的用户界面。在介绍操作系统的时候我们已经看到,用户只需要通过鼠标双击相应的图标就可以做相应的操作,桌面以及其上的图标就是操作系统提供给用户使用的界面,有了这种用户界面,对计算机的操作就比较容易了。用户界面又称为操作系统的前台表现形式,Windows XP 采用的是窗口和图标,DOS 系统采用的是命令,Linux 系统既采用命令形式也配备有窗口形式。不管是何种形式的用户界面,其目的只有一个,那就是方便用户的使用。操作系统的发展方向是简单,直观,方便使用。

(4) 操作系统优化系统功能的实现。由于系统中配备了大量硬件、软件,因而它们可以实现各种各样的功能,这些功能之间必然免不了发生冲突,导致系统性能下降。操作系统要使计算机的资源得到最大利用,使系统处于良好的运行状态,还要采用最优的实现功能的方式。

(5) 操作系统协调计算机的各种动作。计算机的运行实际上是各种硬件的同时动作,是许多动态过程的组合,通过操作系统的介入,使各种动作和动态过程达到完美的配合和协调,以最终对用户提出的要求反馈满意的结果。如果没有操作系统的协调和指挥,计算机就会处于瘫痪状态,更谈不上完成用户所提出的任务。

因此,操作系统可以定义为如下 3 个方面的程序集合:

(1) 控制和管理计算机系统的硬件和软件资源;

(2) 合理地组织计算机的工作流程;

(3) 方便用户使用。

综上所述,操作系统可以定义为:对计算机系统资源进行直接控制和管理,协调计算机的各种动作,为用户提供便于操作的人机交互界面,存在于计算机软件系统最底层核心位置的程序的集合。

1.3 操作系统的功能

1.3.1 操作系统的功能

可以根据计算机系统资源的分类来对操作系统的功能进行划分。一般说来,计算机系统资源包括硬件和软件两大部分,硬件指处理机、存储器、标准输入/输出设备和其他外围设备;软件指各种文件和数据、各种类型的程序。由于操作系统是对计算机系统进行管理、控制、协调的程序的集合,因此,按这些程序所要管理的资源来确定操作系统的功能,将其分为5个部分。

(1) 作业管理。当用户开始与计算机打交道时,第一个接触的就是作业管理部分,作业是用户交给计算机执行的具有独立功能的任务,用户通过作业管理所提供的界面对计算机进行操作,因此作业管理担负着两方面的工作:向计算机通知用户的到来,对用户要求计算机完成的任务进行记录和安排;向用户提供操作计算机的界面和对应的提示信息,接受用户输入的程序、数据及要求,同时将计算机运行的结果反馈给用户。更具体地说,作业管理要提供:安全的用户登录方法,方便的用户使用界面,直观的用户信息记录形式,公平的作业调度策略等。

(2) 处理机管理。处理机是计算机中的核心资源,所有程序的运行都要靠它来实现。如何协调不同程序之间的运行关系,如何及时反映不同用户的不同要求,如何让众多用户能够公平地得到资源等都是处理机管理要关心的问题。具体地说,处理机管理是操作系统设计者的设计理念。

(3) 存储器管理。存储器用来存放用户的程序和数据,存储器容量越大,存放的数据越多,尽管硬件制造者不断地扩大存储器的容量,还是无法跟上用户对存储器容量的需求,而且存储器容量也不可能无限制地增长,但用户需求的增长却是无限的。在众多用户或者程序共用一个存储器的时候,自然会带来许多管理上的要求,这就是存储器管理要做的,存储器管理要进行如下工作:以最合适的方案为不同的用户和不同的任务划分出分离的存储区域,保障各存储器区域不受别的程序干扰;在主存储器区域不够大的情况下,使用硬盘等其他辅助存储器来替代主存储器的空间,自行对存储器空间进行整理等。

(4) 信息管理。计算机中存放的、处理的、流动的都是信息。信息有不同的表现形态,如数据项、记录、文件、文件的集合等也有不同的存储方式,既可以连续存放也可以分开存放;还有不同的存储位置,如可以放在主存储器上,也可以存放在辅助存储器上,甚至可以停留在某些设备上。不同用户的不同信息共存于有限的媒体上,如何对这些文件进行分类,如何保障不同信息之间的安全,如何将各种信息与用户进行联系,如何使信息不同的逻辑结构

与辅助存储器上的存储结构进行对应,这些都是信息管理要做的事情。

(5) 设备管理。计算机主机连接着许多设备,有专门用于输入/输出数据的设备,也有用于存储数据的设备,还有用于某些特殊要求的设备。而这些设备又来自于不同的生产厂家,型号更是五花八门,如果没有设备管理,用户一定会茫然不知所措。设备管理的任务是:为用户提供设备的独立性,用户不管通过程序还是命令来操作设备,都不需要了解设备的具体参数和工作方式,用户只需要简单地使用一个设备名就可以了;在幕后实现对设备的具体操作,设备管理接到用户的要求以后,将用户提供的设备名与具体的物理设备进行连接,再将用户要处理的数据送到物理设备上;对各种设备信息的记录、修改;对设备行为的控制。

此外,为了方便使用,操作系统还提供了两类接口:

- (1) 命令接口,提供一组命令供用户直接或间接操作;
- (2) 程序接口,提供一组系统调用命令供用户程序使用。

1.3.2 操作系统设计原则

对于操作系统设计者来说,操作系统是架构在底层硬件上的软件系统,因此,硬件的原始功能是靠操作系统来实现的,在实现的过程中,就必须考虑各种硬件的使用效率。而对于用户来说,操作系统是使用计算机的手段,这种手段就必须能够满足用户的需求,要求清晰、明确、快速地对用户的动作作出反应,特别是在多用户使用同一个计算机系统的情况下,系统对用户的反应能力显得尤为重要。以上都是操作系统的设计者设计时应该考虑的问题。

操作系统的设计原则如下。

(1) 尽可能高的系统效率。这里指的效率包括:处理机时间的最大利用,存储器空间的合理安排,输入/输出设备的均衡使用。

(2) 尽可能大的系统吞吐能力。在多用户情况下,虽然许多用户同时使用计算机,但每个用户并不考虑别人的工作状况,每个用户都可能进行大量的数据传输,这对于系统的负荷能力是一种考验,因此,系统吞吐量是操作系统设计的一个质量标志。吞吐量的好坏直接影响系统的稳定性,大的吞吐量使系统能流畅地工作,小的吞吐量可导致系统在高负载下瘫痪。

(3) 尽可能快的系统响应时间。响应时间指系统对用户的输入作出反应的时间。通常情况是,用户数量越多,需要的响应时间越快,并且对每一个用户来说响应时间应该是平均的,因此,系统必须提供一个用户能够承受的系统响应时间的下限。

以上是操作系统设计的三个原则,一般情况下,要想获得高的系统利用率就应该尽量避免用户的参与,因此,响应时间就不可能很快;要想获得最佳的用户效果,难免牺牲对系统资源的利用率。这使操作系统设计者处于进退两难的境地。目前,还没有哪个系统能同时完全满足上面三个设计原则,任何一个系统都具有倾向性,通常都是在以某一个设计原则为主的情况下,兼顾另外的设计原则。那么到底以哪个设计原则为主呢?这要看计算机系统的使用目的,在操作系统的发展过程中,这些设计原则交替起着主导作用。

1.3.3 操作系统的发展

操作系统是随着计算机的发展而发展的,从早期的无操作系统的计算机发展到今天,操

作系统已经成为计算机的灵魂,离开了操作系统计算机将无法运行。

1. 计算机系统发展初期

1946 年所产生的计算机系统是没有操作系统的,当时的计算机由硬件的几大部分构成:运算器、控制器、存储器、输入设备、输出设备(见图 1.3)。操作员通过控制台的各种开关来指挥各个部分的运行,它通知输入设备接受用户准备好的装有程序和数据的输入卡片,将输入的程序数据安排到存储器的某个具体位置,通知运算器运行程序并处理数据,通知输出结果打印成纸带。若发现系统在运行过程中有问题,则操作员可通过控制台的开关对各种参数进行设置,将系统调整为正常状态。这个时期的操作员是非常专业的,只有他们才能实现对计算机系统的控制。因此,操作员的能力和反应速度直接影响到计算机的工作效率。但不管多么高级的操作员,其手动速度永远无法和机器速度相比较,机器的运行速度必然受到人工速度的极大制约。又因为早期的计算机硬件价格非常昂贵,人们希望计算机尽可能多地处于运行状态,希望处理器运行尽可能地饱满,这样才不至于造成资源的浪费。其解决办法是尽可能地减少人的干预,让机器来做更多的事情,这就是早期的批处理系统。

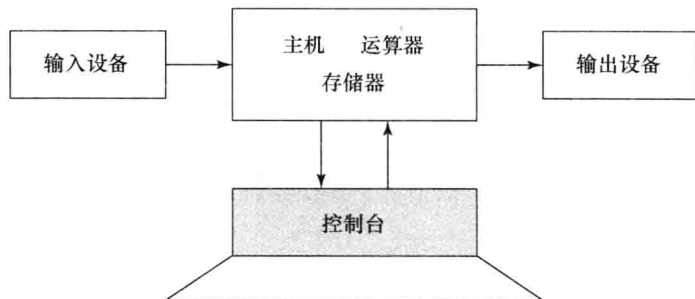


图 1.3 早期计算机的构成

为了减少人的参与,操作员对要送到计算机上运行的程序进行组织,通常是按程序的执行步骤进行分类。凡是运行步骤大致相同的程序组织成一批,由操作员通过输入机输入到磁带机上,再将磁带机连接到计算机主机上准备运行,余下的控制工作交由称为监督程序的程序来控制完成。完成后操作员将存有输出结果的磁带机取下,再连接到输出设备上逐一地输出不同程度的输出结果,最后交给用户。这时的计算机系统称为脱机系统,输入/输出设备与主机之间不再有直接的联系,主机只与磁带机打交道。

监督程序模拟操作员的工作:将磁带机上的程序调入存储器,安排程序运行,将运行结果输出到磁带机上,然后安排一个程序的运行,如此周而复始直到这一批程序全部处理完毕。例如:有一个用高级语言编写的程序需要运行,监督程序将磁带机上的源代码调入主存储器,再调用编译程序对源代码进行编译形成目标代码,然后安排目标代码运行,直到产生结果,最后将结果送到存放结果的磁带机上。当一个程序运行完毕以后,监督程序又将下一段源代码调入主存储器,然后重复上面的过程,直到运行完磁带机上的所有的程序。整个过程都是由监督程序来控制的。监督程序是事实上的管理者,管理者的出现意味着操作系统有了产生的基础。

因为监督程序的参与,人的干预减少到最低,计算机主机只与输入/输出设备打交道,避免了由人引起的计算机资源的等待。但新的问题又出现了:由于输入/输出设备是纯机械

设备或者机械加磁设备,而计算机主机是电子器件,计算机主机还是不可避免地要等待输入/输出设备的运行,主机的利用率不可能很高。那么如何解决电子速度与机械速度严重不匹配的问题呢?采用的办法不是提高输入/输出设备的速度,而是让计算机主机同时连接多台机械设备,以增加主机的工作量。多道批处理系统由此而产生。图 1.4 指出了多道批处理系统的控制。

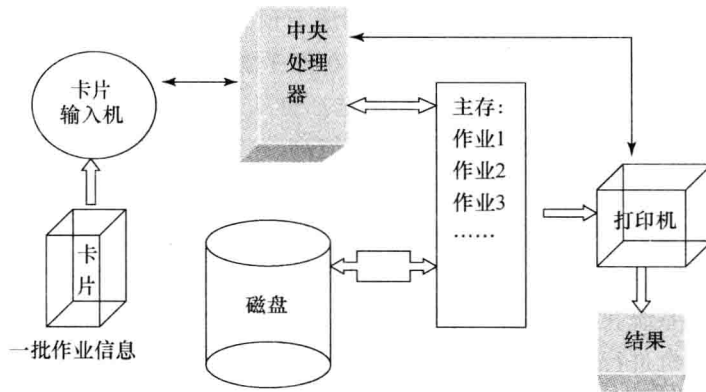


图 1.4 作业成批处理的控制

2. 多道批处理系统

同样是将用户作业组织成批,但主存储器中存放着不止一批的作业,处理机在调用一批作业运行时,如发现输入/输出所产生的等待,监督程序就引导处理机去执行另外的程序,这样就使处理机总是处于工作状态。图 1.5 描述了多道批处理系统处理机的时间分配:程序 A 首先获得处理机,进行了一段时间后,它需要完成输入/输出工作,这时监督程序运行,一

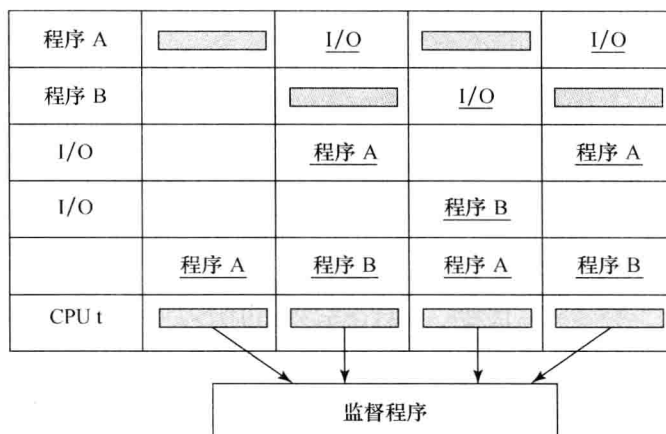


图 1.5 多道批处理 CPU 的时间分配

方面安排程序 A 进行输入/输出处理,另一方面安排程序 B 到处理机上去运行;当程序 B 运行一段时间后也需要完成输入/输出工作时,又由监督程序来安排程序 B 进行输入/输出处理,帮助程序 A 结束输入/输出工作,安排程序 A 再到处理机上运行。所以,从处理机的时间轴上可以看到:程序 A 和程序 B 是交替运行的,如果在这个时间内只有一个程序运行,处理机将有一半时间在等待输入/输出设备完成工作。当然监督程序也要占用一定的处理机

时间,但它与程序运行所需要的处理机时间相比是微不足道的。

这时的监督程序就变得更为复杂,它不但要管理某一批程序的运行与中断,还要对不同批次的程序进行处理机时间的分配。从理论上讲,存储器上存放的程序批次越多,处理机的利用率就越高。如果存储器上存放的程序无限多,则处理机的利用率可以达到100%。这样看来,多道批处理系统可以使计算机资源的利用率达到最大。

多道批处理系统不允许用户干预。用户无法干预并不等于用户不想干预,也许在程序刚被送入主存储器,用户就希望重新修改;也许在程序刚开始运行时,用户就发现了错误;也许在程序的运行过程中,用户希望参与自己的选择意见。总之用户希望干预计算机的运行,这就给管理程序提出了更高的要求:既要尽可能高地提高主机的利用率,又要使用户能够方便地干预程序的运行。于是用户与主机之间不再通过磁带机相互隔开,而是通过输入/输出设备直接相连,新一轮的联机系统出现了。

3. 联机多道程序系统

联机多道程序系统在现实生活中随处可见,典型的组成如图1.6所示,每一个用户从自己的终端上和计算机进行交互,存储器上的不同区域中保存着不同用户的程序,处理机按一定的规则对不同用户的程序进行反应,共享的输入/输出设备按用户的要求在工作。联机多道系统靠程序来控制计算机设备和用户终端,它要面对多个用户,要进行处理机时间的安排,进行内存空间的划分,安排用户分享的输入/输出设备,协调用户在运行程序时发生的各种冲突等,这种程序有一个新的名字,称为操作系统。

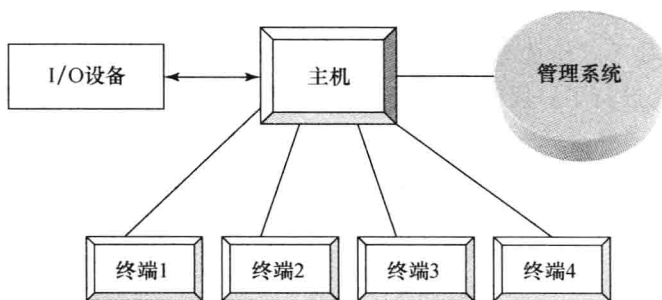


图 1.6 联机多道系统工作图

1.4 操作系统的分类

迄今为止,各操作系统均属于下列操作系统之一或它们的组合:

- (1) 单用户(微机)操作系统;
- (2) 批处理系统;
- (3) 分时系统;
- (4) 实时系统;
- (5) 分布式操作系统;
- (6) 网络操作系统;
- (7) 多处理机操作系统。