

高等学校计算机科学与技术教材

操作系统原理与应用

胡立栓摇王育平摇夏明萍摇编著

清华大学出版社

北京交通大学出版社

·北京·

内 容 简 介

本书是一本面向计算机应用和信息管理类专业的教材,深入浅出地介绍了计算机操作系统的基本原理与相关技术及其在 ~~Windows~~ 系统中的具体应用,并结合 ~~Windows~~ 操作系统案例,从实践出发,引导学生学习操作系统的核心技术。

全书共分 ~~10~~ 章。第 ~~1~~ 章到第 ~~10~~ 章,全面系统地介绍了操作系统的基本原理、处理机管理、存储管理、设备管理和文件管理等;第 ~~7~~ 章到第 ~~10~~ 章,讨论了 ~~Windows~~ 操作系统的产生、发展、基本概念、算法和内核的实现等。书后附有上机实验指导。

本书是操作系统理论与实践结合得较好的教材,适合于广大在校学生学习,也可供计算机爱好者阅读、参考。

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,侵权必究。侵权举报电话: ~~010-62786544~~ 摇 ~~010-62776969~~ 摇 ~~010-62786544~~ 摇 ~~010-62776969~~ 摇

摇图书在版编目(CIP)数据

摇操作系统原理与应用 ~~鞠立栓,王育平,夏明萍编著~~—北京:清华大学出版社;北京交通大学出版社, ~~2013~~

摇(高等学校计算机科学与技术教材)

摇 ~~陈昊,张恩,范原,陈原,陈原,陈原~~

摇 I 操...摇 II 鞠...摇 ②王...摇 ③夏...摇 III 操作系统 原高等学校 原教材摇 IV 摇 ~~431.4~~

摇中国版本图书馆 CIP 数据核字(~~2013~~)第 ~~161616~~ 号

责任编辑:谭文芳

特邀编辑:宋林静

出版发行:清华大学出版社 摇 邮 编: ~~100084~~ 摇 电 话: ~~010-62770175~~ 摇 邮 表: ~~010-62770175~~ 摇 邮 表: ~~010-62770175~~

北京交通大学出版社 摇 邮 编: ~~100044~~ 摇 电 话: ~~010-51686000~~ 摇 邮 表: ~~010-51686000~~ 摇 邮 表: ~~010-51686000~~

印 刷 者:北京交大印刷厂

经 销:全国新华书店

开 摇 本: ~~185mm×260mm~~ 摇 印 张: ~~16.5~~ 摇 字 数: ~~350~~ 千 字

版 摇 次: ~~2013~~ 年 ~~10~~ 月 第 ~~1~~ 版 摇 ~~2013~~ 年 ~~10~~ 月 第 ~~1~~ 次 印 刷

书 摇 号: ~~陈昊,张恩,范原,陈原,陈原,陈原,陈原~~ 摇 号

印 摇 数: ~~1-10000~~ 册 摇 定 价: ~~39.00~~ 元

本书如有质量问题,请向北京交通大学出版社质监组反映。对您的意见和批评,我们表示欢迎和感谢。

投诉电话: ~~010-51686000~~ 摇 ~~010-51686000~~; 传 真: ~~010-51686000~~ 摇 邮 表 责 任 部 摇 邮 表 责 任 部

前 摇 摇 言

操作系统作为一门计算机的基础课程，无论是对计算机信息技术专业的学生或研究人员，还是对一般计算机应用人员而言，都是非常有益和重要的。目前计算机应用和信息管理类专业在操作系统教学安排上比较重视基础理论学习，而相对于应用实践教育来说比较欠缺。本书力求做到把理论与实践进行有机地结合。

本书由浅入深地介绍操作系统的基本原理与相关技术，并结合 ~~系统~~操作系统案例，从实践出发，引导学生学习操作系统的核心技术：处理机管理、存储管理、设备管理和文件管理等。

全书可分为两大部分：第一部分从第 ~~一~~章到第 ~~九~~章，全面系统地讨论了操作系统的基本原理和实现技术；第二部分为第 ~~十~~章到第 ~~十五~~章，介绍了 ~~系统~~操作系统的产生、发展、基本概念、算法和内核的实现等。全书各章的安排分别为：第 ~~一~~章为操作系统概述，其主要目的是让读者先对操作系统有一个概括的了解，并初步建立起操作系统的整体概念；第 ~~二~~章讲述了处理机管理，介绍了进程的基本概念、实现方法、控制技术，各种调度模型和一些重要的调度算法；第 ~~三~~章讲述了存储管理，介绍了实存和虚存管理的有关概念和主要的存储管理方法；第 ~~四~~章讲述了文件管理，介绍了文件系统的有关概念，文件的组织、存取、共享、保护方法，以及文件目录、文件存储空间的管理方法；第 ~~五~~章讲述了设备管理，介绍了 ~~设备~~设备的管理技术和各种磁盘调度算法；第 ~~六~~章讲述了并发进程，介绍了进程互斥、同步、通信的概念与方法，死锁的概念和各种死锁对策；第 ~~十~~章到第 ~~十五~~章把 ~~系统~~作为操作系统的案例介绍了 ~~系统~~操作系统的产生过程、发展趋势；讨论了 ~~系统~~的进程管理、存储器管理、文件系统和设备管理中的核心技术。书中每一章的开头部分都给出了该章的介绍说明，对读者学习和理解该内容起到一定的指导作用；每章末附有习题。附录部分给出了 ~~十~~个实验。前 ~~五~~个实验主要针对第一部分前 ~~九~~章中的内容，是对一些典型算法的具体实现；后 ~~五~~个实验针对第二部分内容，以 ~~系统~~操作系统作为实验平台，掌握 ~~系统~~平台下的系统操作。

本书由胡立栓负责全书的大纲制定和统编。参加编写工作的有：王育平、夏明萍、陆顺、董南萍、周冠玲、项彩虹；由陆顺和项彩虹同志提供了实验素材，并由项彩虹负责全书的校对工作。本书在编写过程中参考了国内外大量的相关著作和教材，还参考了 ~~附录~~提供的技术资料。在此，谨向这些教材的作者和 ~~附录~~资料的提供者深表谢意。在本书的编写、出版过程中，得到了北京交通大学出版社的大力支持，谨向为本书的出版而辛勤工作的同志们表示诚挚的谢意。由于作者水平所限，书中疏漏与错误之处在所难免，恳请各位专家和读者批评指正。

编 者
二〇〇八年 远月

第 1 章 操作系统概论

本章要点：

1.1 操作系统的定义与功能

- 1.1.1 操作系统的形成与发展
- 1.1.2 操作系统的分类
- 1.1.3 操作系统的运行环境

计算机是人类社会 20 世纪最伟大的创造之一，自 1946 年诞生第一台计算机至今的短短 70 多年中，其技术得到了突飞猛进的发展。目前它不仅被广泛应用于科学计算、过程控制、数据处理及军事技术等领域，而且也渗透到办公、教育和家庭等方方面面，已成为社会信息化的重要支柱和人类文明高度发展的象征。

1.1 操作系统的定义与功能

1.1.1 硬件与软件

计算机由硬件系统和软件系统两个部分组成，它们构成了一个完整的计算机系统。

计算机硬件是各种物理设备的总称，是完成工作任务的物质基础。按功能可以把硬件划分成五个部分：运算器、控制器、存储器、输入设备和输出设备，其中运算器和控制器通常被称为中央处理机（CPU），这是冯·诺依曼式计算机的结构，如图 1-1 所示。其中的实线代表控制信号，虚线代表数据传输。

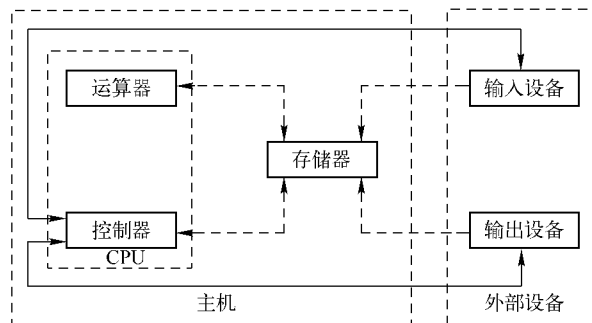


图 1-1 计算机硬件的结构

计算机软件是指程序和与程序相关的文档的集合，是计算机系统的重要组成部分。按功能划分软件可分为系统软件、支撑软件和应用软件。

系统软件是指由计算机生产厂家提供、具有通用功能、与硬件共同构成其他软件运行的基础(要求高效率),比如操作系统、语言处理程序(如悦语言编译程序)、数据库管理系统及各种完成服务功能的程序。支撑软件是与系统软件和硬件一起构成一个应用程序的开发平台,比如办公软件的支撑系统、各种图片处理软件等。应用软件是领域知识与计算机知识结合,为解决实际问题而研制的那些软件,它涉及计算机应用的各个领域,完成特定功能,如各种数据管理软件、用于工程计算的软件包、辅助设计软件及过程控制软件等。

其中操作系统(简称操作系统)是计算机系统中最重要的系统软件,如果把计算机系统看作是一个工厂,操作系统就是这个工厂的厂长,厂长管理的好坏直接决定工厂的命运。因此,操作系统管理的如何直接决定整个计算机系统的运行效率。操作系统是整个系统的控制中心,它控制和管理计算机系统的各类资源,并为其他系统程序和应用程序提供基本的服务。硬件常称为裸机,它的功能即使很强,但也往往不方便用户使用,功能上相对来说也是有局限性的。而软件可以在硬件基础之上对硬件的性能加以扩充和完善,因而一个裸机在每加上去一层软件后,就变成了一个功能更强的机器,通常把这“新的更强功能的机器”称之为“虚拟机”。

综上所述,操作系统的主要作用是为用户提供使用(操作系统是用户与硬件之间的接口程序)、有效地管理和使用计算机系统资源,提高资源的利用率。

显然以上操作系统的两个作用有时是矛盾的(一个软件为重、一个硬件为重)。在计算机技术发展早期,由于资源十分昂贵,从效率上考虑多,因而操作系统的许多理论也主要是讨论如何优化地使用这些资源。在计算机高度普及的今天,强调方便用户则是软件设计更为重要的目标,以至于“用户友好”(或用户界面)已成为计算机术语中的常用词。随着计算机硬件的发展,系统结构及操作系统都在不断完善和发展,二者相辅相成,互相促进,今后的发展仍然如此。

2.1 操作系统的定义

操作系统是在裸机上加载的第一层软件,是对计算机硬件系统功能的首次扩充。从用户的角度看,计算机系统配置了操作系统后,由于操作系统隐蔽了硬件的复杂细节,用户会感到机器使用起来更简单、更容易了。这就是上面所说的操作系统为用户提供了一台功能经过扩展的虚拟机,因为现实生活中并不存在具有这种功能的真实机器,它只是用户的一种感觉而已。从计算机系统的角度看,由于操作系统的组织与管理,系统中的各种硬、软件资源得到了更有效的利用,机器的工作流程更为合理与协调。因此,操作系统是现今计算机系统中不可缺少的一个系统软件。

综上所述,可以把操作系统定义为:“操作系统是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及方便用户使用计算机的一个大型程序”。

2.2 操作系统的特征

2.2.1 并行性

并行性是指在计算机系统中,同时存在多个程序,从宏观上看程序是同时向前推进的,在微观上,程序是顺序执行的,在单处理器上是轮流执行的,即两个或两个以上的进程的执行在时

间上有重叠。

（源）共享性

共享性是指操作系统程序和多个用户程序共同使用系统中各种硬件和软件资源。通常有互斥共享和同时共享两种方式。

（猿）随机性

随机性是指因为多个用户程序共享系统中各种资源,造成了系统中很多不确定性因素,因此操作系统是在一种随机的环境下运行的,操作系统不能对所运行的程序的行为及硬件设备的情况作出任何事先的假定,程序运行时的推进速度和运行结果都具有随机性。

（缘）虚拟性

虚拟性的本质含义是把物理上的一个变成逻辑上的多个。操作系统可以对计算机硬件系统的功能进行扩充,对用户隐蔽硬件的复杂细节,操作系统为用户提供了一台功能经过扩展的机器,但现实生活中并不存在具有这种功能的真实机器,因此操作系统具有虚拟性。

总之,共享是操作系统的目的,并发是实现共享的,由于并发带来了系统的随机性,为提高系统效率,通过操作系统扩充了硬件的功能。

猿源 猿 操作系统的功能

把操作系统看作是计算机中各类资源的管理系统,为用户提供一种简便、有效地使用资源的,以充分发挥各种资源的利用率。

计算机中的资源包括硬件资源和软件资源。其中硬件资源包括处理机(悦载)、存储器(包括内存和外存)、各种输入输出设备,软件资源包括程序和数据(也称信息资源)。

要管理好上述资源,首先要将资源状态记录下来;其次要合理分配资源,防止无限等待的发生;再者要进行资源分配;最后要回收资源。

从资源管理的角度看,操作系统应该具有四个方面的功能,即处理机管理、存储管理、设备管理、文件管理。这四大部分功能相互配合,协同工作,实现对计算机系统的资源管理和控制程序的执行。下面就分四个方面对操作系统的功能做一个简略的说明。

（员）处理机管理

处理机管理的主要工作是给多个作业分配悦载,其功能又分为进程管理与作业管理,关于处理机管理方法的有关内容在第 猿章详细介绍。

（圆）存储管理

存储管理的任务是分配主存和收回主存空间,保护主存内信息安全,关于存储管理方法的有关内容在第 猿章详细介绍。

（猿）文件管理

文件管理涉及文件存储空间的分配与回收、文件目录管理、文件读写与保护等。关于文件管理方法的有关内容在第 源章详细介绍。

（源）设备管理

设备管理主要是对设备进行分配、回收与控制。进行设备优化调度,以提高设备使用效率。关于设备管理方法的有关内容在第 缘章详细介绍。

如图 猿源所示是操作系统的四个功能的层次关系,其中下层为上层提供服务。

作业管理
文件管理
设备管理
存储管理
进程管理

图 1-1 操作系统的功能层次关系

1.1 操作系统的形成与发展

通常把未配置任何软件的计算机称为“裸机”。如果让用户直接面对裸机,事事都深入到计算机的硬件里面去,那么他们的精力就绝对不可能集中在如何用计算机解决自己的实际问题上,计算机本身的效率也不可能充分发挥出来。

举例说,要在一台裸机上进行硬盘读操作,使用者至少应该把磁盘地址、内存地址、字节数和操作类型(读/写)等具体值装入到特定的硬件寄存器中,否则根本谈不上完成预定的任务。实际上,对许多设备而言,除此以外往往会要求比这更多的操作参数。在任务结束后,还需要对设备返回的诸多状态加以判别。

又例如,某计算机内存储器可供用户使用的容量为 1MB。若现在装入的用户程序占用其中的 500KB,那么余下的 500KB 就被闲置了。如果能够在内存中装入多个程序,比如在 1MB 中再装一道需要存储量 500KB 的程序进去,当第一个程序等待 I/O 的完成而暂时不用 CPU 时,能让第二道程序投入运行,那整个计算机系统的利用率就会比原来大为提高。由此可见,内存中装入多道程序的好处是:

- ① 内存浪费得少了,原来浪费 500KB,现在只浪费 500KB;
- ② CPU 比原来更加忙碌了,在第一个程序等待 I/O 的完成时,原来 CPU 只能采取空转的方式来等待,现在可以让它去执行第二个程序;
- ③ 在 CPU 执行第二个程序时,它与第一个程序启动的 I/O 设备并行工作。

可见,为了从复杂的硬件控制中脱身出来,为了能合理有效地使用计算机系统,为了能给用户使用计算机提供必要的方便,最好的解决办法就是要开发一种软件,通过它来管理整个系统,发挥系统的潜在能力,达到扩展系统功能、方便用户使用的目的。实际应用的需要,就是“操作系统”这一软件呼之欲出的根本原因。

只有深刻地了解操作系统的过去和现在,才能更好地迎接未来,下面先简单地回顾一下操作系统的形成和发展过程。

1.1.1 手工操作阶段

20 世纪 40 年代后期是手工操作阶段。第一台电子管计算机出现后,计算机系统基本上仅由硬件组成,计算机上并没有名为“操作系统”的软件。那时计算机的运行速度慢,外部设备少,整个系统是由用户直接控制使用,因此程序的装入、调试及控制程序的运行等工作,全部由上机的人员自己通过按动控制台上的一排排开关和按钮来实现。这一时代的特点是人工完成上、下机的操作,一台计算机被一个用户所独占。通常称计算机发展初期的这种操作方式为“手工操作”阶段。早期计算机系统的工作过程如图 1-2 所示。

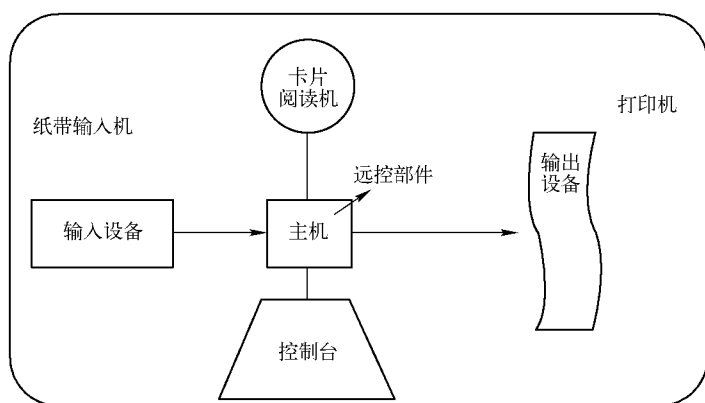


图 5-1-1 早期计算机系统的工作过程

“手工操作”的特点是用户上机时一人独占全机资源,程序运行前的准备时间过长,人机存在矛盾,即人的操作速度与机器运行速度相比,仍存在速度极不匹配的矛盾(操作员等待人工操作)。

随着计算机速度的提高,人机矛盾加剧。例如某程序在 1 万次每秒的机器上运行需 1 小时,人工操作需 1 分钟,人机之比为 1:60,这种比例一般还可以接受。若机器运行速度提到 10 万次每秒,此时程序运行只需 1 分钟,人工不变,人机之比变成 1:600,这种比例是难以让人接受的,这时产生了早期批处理系统。

5.1.1 早期批处理系统——联机批处理

20 世纪 50 年代初期,为了缓和早期使用计算机时存在的人—机速度严重不匹配的矛盾,提高资源利用率,人们开始利用计算机系统软件来代替操作员的部分工作,从而产生了早期批处理系统。

由于人机存在的矛盾,软件设计人员提出了“让计算机自动控制用户作业的运行,废除上、下机手工交接”的要求。为了达到这个目的,需要用户在编写自己的程序时,还要编写“作业说明书”,详细规定程序运行的步骤,并将其与程序、数据一起提交给系统;而系统需要设计一个常驻内存的“管理程序”(也称监督程序),它的功能是从输入设备上读入第一个作业的作业说明书,按照它的规定控制该作业执行。这个作业运行结束后,它又从输入设备上读入第二个作业的作业说明书,继而执行之。这一过程一直进行到提交给系统的一批作业全部执行完毕时为止。操作员有选择地把若干作业合成一批,安装到输入设备上,并启动监督程序,然后由监督程序自动控制这批作业运行,从而减少部分人工干预,有效地缩短了作业运行前的准备时间,相对地提高机器的利用率,如图 5-1-2 所示。

由于这种系统一次集中处理一批用户作业,故被称为批处理系统。这个时代的特点是对一批作业自动进行处理,没有人工交接,在一个用户作业运行时,仍独占计算机所有资源,用户交互性差,一个作业运行完后,还要等待从纸带机或卡片机上读入下一个作业,机器仍不能充分利用。

由于主机直接控制与外部设备传输数据,所以这种工作方式称为早期联机批处理系统。

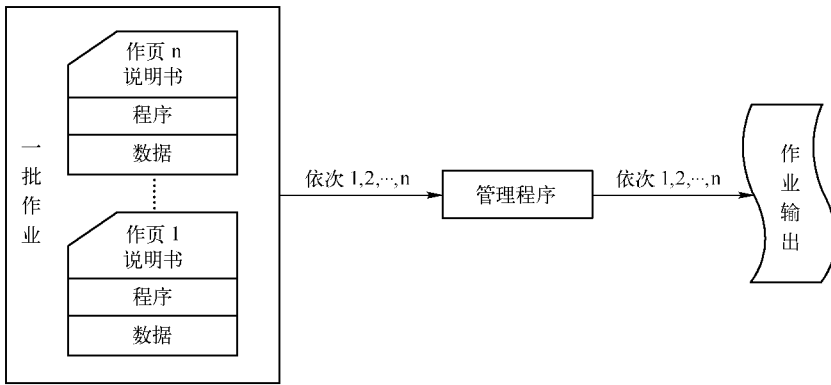


图 1.1 早期联机批处理系统的工作过程

1.1.2 早期批处理系统——脱机批处理

20 世纪 50 年代中期出现了磁带机，磁带机的传输速度比卡片机、光电机和打印机的速度快，用磁带机代替这类低速的外部设备，可进一步缩短系统与外部设备间速度上的差异，提高系统的利用率。这时出现了早期脱机批处理系统，即主机不直接与外部设备传输数据，数据的输入输出由卫星机完成，主机主要完成运算，主机与低速的卡片机、光电机和打印机等外部设备脱离，所以这种工作方式称为早期脱机批处理系统。其工作过程如图 1.2 所示。

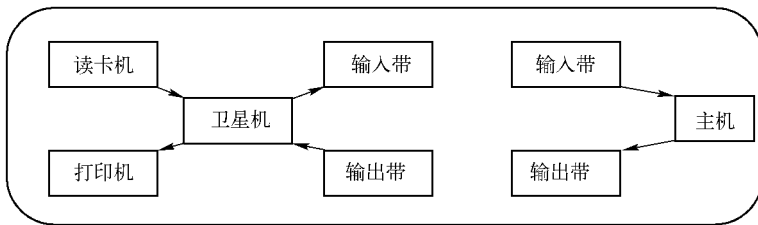


图 1.2 早期脱机批处理系统的工作过程

虽然这种工作方式实现了主机与卫星机的并行操作，但因系统中作业之间仍以串行方式被处理，所以无法继续提高系统内存的利用率，为从根本上解决这一问题，人们提出了多道程序设计技术。

1.1.3 操作系统

20 世纪 50 年代末期，硬件技术取得了两个方面的重大进展，即通道技术的引进和中断技术的发展，使得通道具有中断主机工作的能力，主机与通道之间借助中断相互通信，通信受主机直接控制。另外，外部设备出现了磁盘机，磁盘是一种比磁带更快并能随机存取的外部存储设备（磁带机只能顺序存储）。

为了提高资源利用率，人们开始使用磁盘、输入输出缓冲、脱机输入输出、虚拟设备（也称杂项设备系统）等技术，尤其是引入了多道程序设计，使简单批处理系统发展为高级批处理系统。监督程序与中断处理程序合称为执行系统，监督程序长住内存，负责程序切换。我们

称这种操作方式为“执行系统”阶段。图 1-10 所示。

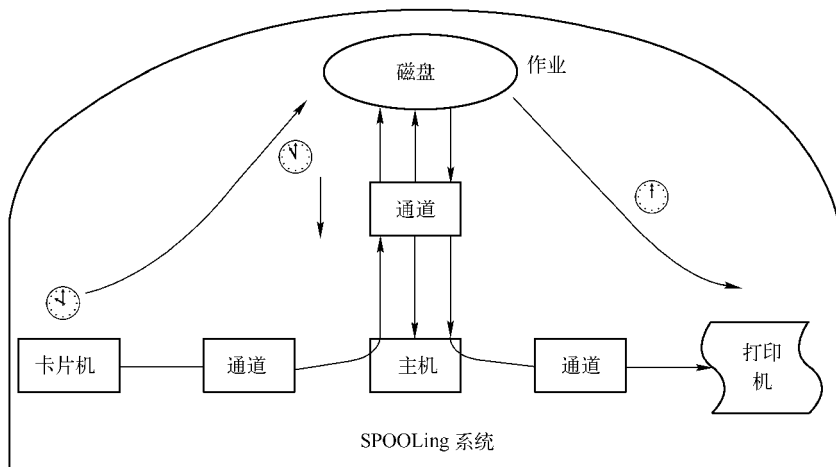


图 1-10 SPOOLing 系统的工作过程

执行系统的优点是实现了主机与外部设备的并行工作。其缺点是系统中作业之间仍以串行方式被处理,所以效率不高。

在执行系统的基础上,引入了多道程序设计后,单道批处理系统发展为多道批处理系统。多道批处理系统的出现,标志着操作系统正式形成。

操作系统的分类

单道批处理操作系统

单道批处理系统

在单道批处理操作系统的控制和管理下,用户为自己的作业编写程序和准备数据,同时编写控制作业运行的作业说明书。然后将它们一并交给操作员。计算机系统的工作过程如下:

① 操作员将收到的一批作业信息存入辅助存储器中等待处理;

② 单道批处理操作系统从辅助存储器中依次选择作业,按其作业说明书的规定自动控制它的运行,并将运行结果存入辅助存储器;

③ 操作员将该批作业的运行结果打印输出,并分发给用户。

单道批处理操作系统有如下特点。

☒ 单路性:每次只允许一个用户程序进入内存。

☒ 独占性:整个系统资源被进入内存的一个程序独占使用,因此资源利用率不高。

☒ 自动性:作业一个一个地自动接受处理,期间任何用户不得对系统的工作进行干预。由于没有了作业上、下机时用户手工操作耗费的时间,因此提高了系统的吞吐量。

☒ 封闭性:在一批作业处理过程中,用户不得干预系统的工作。即便是某个程序执行中出现一个很小的错误,也只能等到这一批作业全部处理完毕后,才能进行修改,这给用户带来不便。

图 10-1 多道批处理系统

在单道批处理的基础上,引入多道程序设计技术,就产生了多道批处理操作系统。配置多道批处理操作系统的本质仍然是批处理。不同的是由于采用了多道程序设计技术,允许若干个作业同时装入内存,造成对系统资源共享与竞争的态势。用户为自己的作业编写程序和准备数据,同时编写控制作业运行的作业说明书,然后将它们一并交给操作员。其工作过程如下:

① 操作员将收到的一批作业信息存入辅助存储器中等待处理。

② 多道批处理操作系统中的作业调度程序从辅助存储器里的该批作业中选出若干合适的作业装入内存,使它们不断地轮流占用 **悦载** 来执行,并同时使用各自所需的外部设备。若内存中有作业运行结束,则又可从辅助存储器的后备作业中选择对象装入内存执行。

③ 操作员将该批作业的运行结果打印输出,分发给用户。

多道批处理操作系统有如下特点。

☞ **多路性**:每次允许多个用户程序进入内存,它们轮流交替地使用 **悦载**,提高了内存储器 and **悦载** 的利用率。

☞ **共享性**:整个系统资源被进入内存的多个程序共享使用,因此整个系统资源的利用率较高。

☞ **自动性**:作业处理期间任何用户不得对系统的工作进行干预。由于没有了作业上、下机时用户手工操作耗费的时间,因此提高了系统的吞吐量。

☞ **封闭性**:在一批作业处理过程中,用户不得干预系统的工作。即便是某个程序执行中出现一个很小的错误,也只能等到这一批作业全部处理完毕后,才能进行修改,这给用户带来不便。

图 10-2 分时操作系统

将多道程序设计技术与分时技术结合在一起,就产生了分时操作系统。配有分时操作系统的计算机系统称为分时系统。

所谓分时系统,即一台计算机与多个终端设备连接,最简单的终端可以由一个显示器和一个键盘组成。每个用户通过终端向系统发出命令,请求系统为其完成某项工作。系统根据用户的请求完成指定的任务,并把执行结果返回。这样用户可以根据运行结果,再次通过终端向系统提出下一步请求。重复这种交互会话过程,直至每个用户实现自己的预定目标。图 10-2 所示为分时系统工作过程的示意图。

分时系统之所以能在较短的时间内响应用户的请求,同时为多个终端用户提供服务,主要是因为分时操作系统中采用了“时间片轮转”的处理机调度策略。这种调度策略是把处理机时间划分成一个个很短的“时间片”,对提出请求的每个联机用户终端,系统轮流分配一个时间片给其使用。若在一个时间片内,用户所请求的工作未能全部做完,就会被暂时中断执行,等待下一轮循环再继续做,让出的 **悦载** 被分配给另一个终端使用。

由于计算机的处理速度很快,只要时间片的间隔取得适当,那么用户就不会感觉到从一个时间片跨越到另一个时间片之间的“停顿”,就好像整个系统全由他“独占”使用似的。例如,若时间片为 **几** 毫秒,系统中有 **几** 个用户终端分享 **悦载**,并假定忽略操作系统为实现用户终端之间的切换所需耗费的时间,那么每个用户平均响应时间(即从用完一个时间片到获得下

一个时间片所需的时间间隔)为 1 毫秒。这 1 毫秒的“停顿”,用户是完全感觉不出来的。

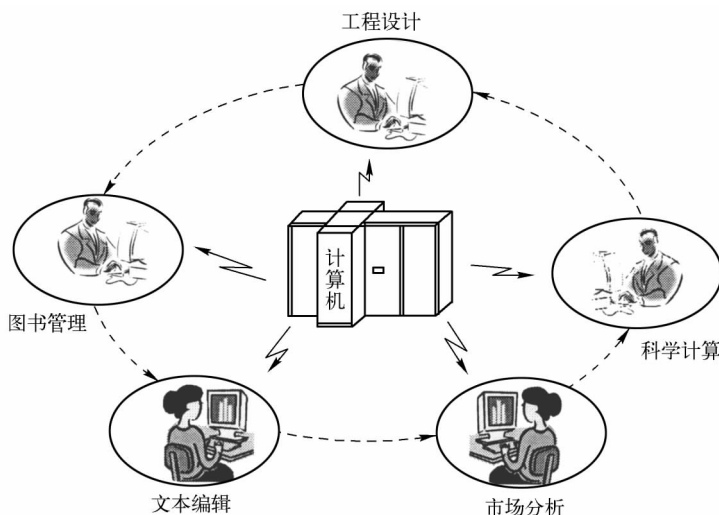


图 1-1 分时系统工作过程的示意图

分时系统的目标是对用户请求快速反应,多道批处理的目标是提高机器效率。

分时系统具有多路性、交互性、独占性和及时性的特点。

1.2 实时操作系统

计算机应用范围日益扩大,比如在控制飞机飞行、导弹发射及冶炼轧钢等生产过程中采用了实时控制系统,在飞机订票、银行业务中采用了实时信息处理系统,它们都打破了只把计算机用于科学计算和数据处理等方面的格局。

所谓“实时”是指能够及时响应随机发生的外部事件,并对事件做出快速处理的一种能力。而“外部事件”是指与计算机相连接的设备向计算机发出的各种服务请求。因此实时操作系统的工作方式是使计算机在规定的时间内及时响应外部事件的请求,同时完成对该事件的处理,并能控制所有实时设备和实时协调一致地工作。实时操作系统根据控制的对象不同又分为实时控制系统和实时信息处理系统两类。

如图 1-2 所示是一个用计算机系统控制化学生产反应的例子。两种原料通过阀门进入反应堆。反应堆中的各种传感装置周期性地把所测得的温度、压力、浓度等测量信号传送给计算机系统。计算机中的实时操作系统及时接收这些信号,并调用指定的处理程序对这些数据进行分析,然后给出反馈信号,控制两种原料的流量,确保反应堆中的各原料参数维持在正常范围之内。若参数超过极限允许值,就立即发出报警,甚至关闭反应堆,以免发生事故。

要求实时操作系统能在严格的时间内对外部请求做出反应,不能出现差错,所以其特点如下。

☞ 及时性:对外部请求在严格时间内作出响应。

☞ 高度可靠性:系统保证不出错,为了提高可靠性,一般采用双工方式。

☞ 专机专用。

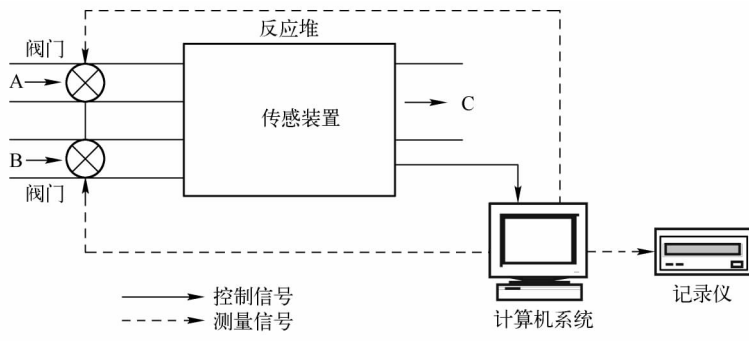


图 1-1-1 一个用计算机系统控制化学生产反应的例子

1.1.2 网络操作系统

所谓计算机网络,是指把地理上分散的、具有独立功能的多个计算机和终端设备,通过通信线路加以连接,以达到数据通信和资源共享目的的一种计算机系统。计算机网络是计算机技术和通信技术相结合的产物。

分时系统提供的资源共享有两个限制,一是限于计算机系统内部,二是限于同一地点(或地理位置很近)。计算机网络在分时系统的基础上,又大大前进了一步。

网络操作系统是基于计算机网络的,是在各种计算机操作系统之上按网络体系结构协议标准设计开发的软件,它包括网络管理、通信、安全、资源共享和各种网络应用。在网络范围内,网络操作系统用于管理网络通信和共享资源,协调各计算机上任务的运行,并向用户提供统一的、有效方便的网络接口的程序集合。要说明的是,在网络中各独立计算机仍有自己的操作系统,由它管理着自身的资源。只有在它们要进行相互间的信息传递、要使用网络中的可共享资源时,才会涉及网络操作系统。

网络操作系统的特点是具有自治性、互联性、统一性和资源共享及信息交换功能等。

1.1.3 分布式操作系统

分布式计算机系统是由多台计算机组成的一种特殊的计算机网络,为分布式计算机系统配置的操作系统称为分布式操作系统。分布式操作系统是网络操作系统的更高级形式,除了有网络操作系统的功能之外,其特征是系统中所有主机使用同一个操作系统,可以实现资源的深度共享,系统具有透明性和自治性。分布式操作系统处理能力增强,运算速度加快,系统可靠性增强。

网络操作系统和分布式操作系统的主要区别如下。

- ☞ 分布式操作系统的各个计算机间相互通信,无主从关系;网络操作系统的计算机有主从关系。
 - ☞ 分布式操作系统的资源为所有用户共享;而网络系统为有限制地共享。
 - ☞ 分布式操作系统中若干个计算机可以相互协作,共同完成一项(多机合作)。
 - ☞ 分布式操作系统比网络操作系统健壮。
- 目前,还没有实用的分布式操作系统。

另外,除上面介绍的操作系统外,还有微机操作系统、通用操作系统和嵌入式操作系统等系统,在此不再赘述。

11.1 操作系统运行环境

因为操作系统是作用在硬件上的第一层软件,所以本节简要介绍一些与操作系统运行有关的硬件结构。

11.1.1 中央处理器(CPU)

本小节首先介绍中央处理器的构成与基本工作方式,然后介绍特权指令、非特权指令和处理器工作状态,最后介绍程序状态字(PSW)。

操作系统作为一个程序要在处理器上执行,如果一个计算机系统只有一个处理器,一般称之为单机系统,如果有多个处理器则称之为多机系统,后面章节中讲述操作系统功能时,使用的是单机系统。

每个计算机处理器都有自己的指令系统。早期的中央处理器的指令系统的功能相对来说比较简单。随着大规模集成电路技术的飞速发展,当代的处理器结构和指令系统已经变得丰富和复杂起来。

11.1.1.1 CPU的构成与基本工作方式

处理器由运算器和控制器组成。运算器实现指令中的算术和逻辑运算,是计算机计算的核心。控制器负责控制程序运行的流程,包括取指令、维护CPU状态、CPU与内存的交互等。

寄存器是一种暂时存储器件,用于CPU执行指令的过程中暂存数据、地址及指令信息。

在计算机的存储系统中,寄存器的访问速度最快。寄存器为处理器本身提供了一定的存储能力,其速度比主存储器快得多,但因为寄存器集成在微处理器芯片中,所以它的造价较高,存储容量一般也比较小。

(1) 处理器中的寄存器

处理器中的寄存器分为用户可见寄存器及控制和状态寄存器两类。

用户可见寄存器,对于高级语言来说,编译器通过一定的算法分配并使用这些寄存器,以最大限度地减少程序运行时访问主存储器的次数,加快程序的运行速度。这种寄存器由机器语言直接引用,对所有程序都是可用的。用户可见寄存器又分为以下三类。

☞ 数据寄存器:也称为通用寄存器,主要用于各种算术运算和逻辑运算。

☞ 地址寄存器:用于存储数据及指令的物理地址或者逻辑地址,用于某种特定方式的寻址。

☞ 条件码寄存器:用于保存CPU操作结果的各种标记位,例如,算术运算产生的溢出、符号等,以控制程序指令的流向。

控制和状态寄存器用于控制处理器的操作,一般由具有特权的操作系统代码使用,以控制其他程序的执行。这些寄存器的大部分对于用户是不可见的,控制和状态寄存器又分为以下三类。

☞ 程序计数器:记录了将要取出的指令的地址。

☞ 指令寄存器 :包含了最近取出的指令。

☞ 程序状态字(PSW) :记录了处理器的运行模式信息。

(四) 指令执行的基本过程

处理器处理指令时,首先,处理器每次从存储器中读取一条指令,并在读取指令完成后,根据指令类别自动将程序计数器的值变成下一条指令的地址,通常是自增1。其次,将读取到的指令放在处理器的指令寄存器中,于是处理器解释并执行这条指令。

一个单条指令处理过程称为一个指令周期。程序的执行就是由不断读取指令和执行指令的指令周期组成的。

指令大致可以分成以下五类。

☞ 访问存储器指令 :负责处理器和存储器之间的数据传送。

☞ 传送指令 :负责处理器和I/O模块之间的数据传送和命令发送。

☞ 算术逻辑指令 :也称数据处理指令,用以执行有关数据的算术和逻辑操作。

☞ 控制转移指令 :可以指定一个新的指令的执行起点。

☞ 处理器控制指令 :用于修改处理器状态,改变处理器工作方式。

(五) 特权指令和非特权指令

对于单用户、单任务下使用的微型计算机系统,普通用户通常都可使用计算机指令系统中的全部指令。但是在多用户或多任务的多道程序设计环境中,指令系统中的指令必须区分成两部分,即特权指令和非特权指令。

特权指令是指指令系统中仅由操作系统使用的指令,不允许一般用户使用,否则可能使系统陷入混乱。特权指令可用于设置程序状态字、启动外部设备、设置时钟、清内存、设置中断屏蔽、建立存储保护等。

非特权指令是指指令系统中用户能使用的指令。

操作系统能使用所有的指令(包括特权指令和非特权指令)。如果在某种计算机的指令系统中不能区分特权指令和非特权指令,在这样的硬件环境下,操作系统无法管理资源,也不能设计出一个具有多道程序运行的操作系统。

(六) 处理器的状态

(1) 管态和目态

处理器有时执行用户程序,有时执行操作系统的程序。在执行不同程序时,根据运行程序对资源和机器指令的使用权限而将此时的处理器设置为不同状态。

处理器工作状态划分为管态和目态。管态一般指操作系统管理程序运行的状态,具有较高的特权级别,又称为特权态、系统态或核心态,可使用特权指令和非特权指令。目态一般指用户程序运行时的状态,具有较低的特权级别,又称为普通态、用户态。

例如,英特尔公司出品的Pentium系列处理器(包括Pentium、Pentium Pro、Pentium 4系列处理器),提供四个特权级别(0、1、2、3),较大的数字表示较低的特权。0运行那些最关键的代码,比如操作系统的内核代码,较外部的特权运行其他一些相对来说不是关键的代码。

(2) 管态状态的转换

从目态进入管态的唯一途径是中断,中断的相关知识将在第9章介绍。从管态进入目态可以直接修改程序状态字PSW的值。

(狗) 限制用户程序执行特权指令

如果用户程序在执行时,企图访问操作系统所在的区域或想使用某个特权指令(如改变指令计数器的内容),就立即被捕获,而被迫中止执行,然后由操作系统处理这一事件,这样保证了操作系统的权利并使其程序不会被破坏。当处理机处于目态中时,不可使用特权指令。

灑 程序状态字(孕宰)

在处理器中最常见的控制和状态寄存器是程序状态字(孕宰)。程序状态字的作用是指示处理器状态。在处理器的状态字中通常包括以下状态代码。

- ☞ 悦裁的工作状态代码:指明当前悦裁的工作状态是管态还是目态,用来说明当前在悦裁上执行的是操作系统程序还是一般用户程序,从而决定其是否可以使用特权指令或拥有其他的特殊权力。
- ☞ 条件码:反映指令执行后的结果特征。
- ☞ 中断码和中断屏蔽码:指出当前中断号和是否允许中断。

不同机器的程序状态字的格式及其包含的信息不同。现以孕藻罐皂系列微处理器的对应程序状态字寄存器(耘德佩尔)中包含的若干标志位为例,来介绍程序状态字孕宰,如图员怨所示。

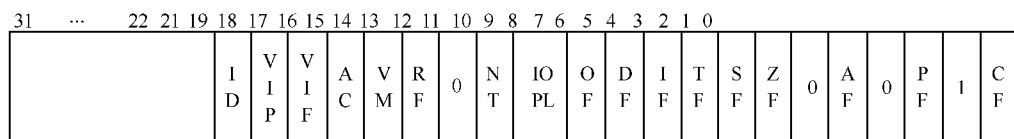


图 员怨 孕藻罐皂系列微处理器程序状态字结构

悦表示进位标志位,左表示结果为零标志位,左表示符号标志位,能表示溢出标志位。几乎所有的计算机的孕宰中都有与这四个标志位类似的标志位。有时称这四种标志位为标准条件位。

耘表示陷阱标志位,左表示中断屏蔽标志位,左表示虚拟中断标志位,左表示虚拟中断待决标志位。左和左用于支持可屏蔽的硬件中断处理。左表示左的特权级别,它的作用在前面已有叙述。

在图员怨中,还有其他许多标志位,从这些众多的标志位的设计,也可看出孕藻罐皂系列微处理器的功能是非常强的,有兴趣的读者可以参考其他有关资料。

员怨 圆 中断机制

计算机必须能够对处理器外面发生的事情作出响应。例如,当按动键盘上一个按键或时钟的报时信号来到或软盘驱动器工作完毕发出中断信号时,均将引起处理器的注意并处理相应事件,这就是中断。

中断能充分发挥处理机的使用效率。因为输入输出设备可以用中断的方式同悦裁通信,报告其完成悦裁所要求的数据传输的情况和问题,这样可以免除悦裁不断地查询和等待,从而大大提高处理机的效率。

中断还可以提高系统的实时处理能力。因为具有较高实时处理要求的设备,可以通过中断方式请求及时处理,从而使处理机立即运行该设备的中断处理程序。

关于中断的知识将在第 4 章详细介绍。

2.1.2 设备控制的技术

在一台计算机系统中,有大量的外部设备,为了控制设备的操作,提高处理器和外部设备的运行效率,出现了各种不同的设备的硬件结构。

每台设备都配有各自的设备控制器,由设备控制器分别控制各台外部设备的运行。

在早期的计算机系统中,设备控制器通过设备的硬件结构与中央处理器连接。对设备控制器的操作是由处理器直接发出的控制指令来实现的。这种程序直接控制设备的主要缺陷是,处理器为了关注设备的控制器的状态,必须耗费大量的时间轮询设备,严重地降低了整个系统的性能。这种早期设备的结构,由于效率太低,已经被淘汰。

现代计算机中使用通道技术和直接内存存取(DMA)技术实现外部设备与主存设备与设备的并行工作。缓冲区是外部设备在进行数据传输期间专门用来暂存这批数据的主存区域,在通道和 DMA 工作中使用缓冲技术,使设备利用率大为提高,同时也提高了主存的利用率。

2.1.3 系统时钟

2.1.3.1 系统时钟的作用

在计算机系统中,设置时钟是十分必要的。这是由于时钟可以为计算机完成以下的必不可少的工作:

- ① 在多道程序运行的环境中,时钟可以为系统发现一个陷入死循环(由编程错误引起)的作业,从而防止机时的浪费;
- ② 在分时系统中,用时钟间隔来实现各个作业按时间片轮转运行;
- ③ 在实时系统中,按要求的時間间隔输出正确的时间信号给相关的实时控制设备;
- ④ 定时唤醒那些要求按照事先给定的时间执行的各个外部事件(如定时为各进程计算优先数,银行系统中定时运行某类结账程序等);
- ⑤ 记录用户使用各种设备的时间和记录某外部事件发生的时间间隔;
- ⑥ 记录用户和系统所需要的绝对时间,即年、月、日。

由上述时钟的作用可以看到,时钟是操作系统运行时必不可少的硬件设施,所以现在的微型机系统中均有时钟。在微型机系统中,通常只有一个间隔时钟(也用作绝对时钟),在大型机中时钟类型通常要多些。

2.1.3.2 系统时钟的分类

- ① 硬件时钟(绝对时钟)的工作原理:在电路中的晶体振荡器,每隔一定时间间隔产生固定的脉冲频率,时钟电路中的时钟寄存器依据时钟电路所产生的脉冲数,对时钟寄存器进行加 1 的工作。
- ② 软件时钟(相对时钟)的工作原理:常用作相对时钟,利用内存单元模拟时钟寄存器,并采用一段程序来计算相应的脉冲数,对内存时钟寄存器进行加 1 或减 1 的工作,从而模拟了时钟的功能。

由于硬件提供的时钟总是比较少的,往往不能满足操作系统和应用程序对时钟的要求,因