

高职高专**计算机**系列教材

JISUANJI

计算机操作系统

Jisuanji Caozuo Xitong

主 编 张 京

副主编 李成大



重庆大学出版社

计算机操作系统

主 编 张 京
副主编 李成大

重 庆 大 学 出 版 社

内 容 提 要

本教材介绍了计算机系统中的一个重要的系统软件——操作系统,以介绍操作系统的基本原理为主,依据资源管理的观点论述操作系统的结构、原理及实现方法。并在此基础上较完整地介绍 UNIX 操作系统的内核结构和 UNIX 系统的使用。本教材在内容上突出理论与实践相结合和能力的培养,在编写上力求循序渐进、通俗易懂,每章都附有习题,便于教学和自学。

本书可作为高职高专、成人高等院校计算机专业和相关专业的操作系统课程的教材,也可供有关计算机应用人员阅读和参考。

图书在版编目(CIP)数据

计算机操作系统/张京主编. —重庆:重庆大学出版社,2004.6

(高职高专计算机系列教材)

ISBN 7-5624-3142-6

I. 计... II. 张... III. 操作系统—高等学校:技术学校—教材 IV. TP316

中国版本图书馆 CIP 数据核字(2004)第 043020 号

计算机操作系统

主 编 张 京

副主编 李成大

责任编辑:曾令维 穆安民 版式设计:曾令维

责任校对:蓝安梅 责任印制:张立全

*

重庆大学出版社出版发行

出版人:张鹤盛

社址:重庆市沙坪坝正街174号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn (市场营销部)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787×1092 1/16 印张:13.25 字数:330千

2004年6月第1版 2004年6月第1次印刷

印数:1—5 000

ISBN 7-5624-3142-6/TP·484 定价:18.00元

本书如有印刷、装订等质量问题,本社负责调换

版权所有 翻印必究

前言

操作系统是现代计算机系统中最重要系统软件。它控制和管理计算机系统的所有硬件和软件资源,并为用户提供友好工作环境。

计算机操作系统不仅是计算机专业必修课程,也是从事计算机应用人员必不可少专业知识。本教材以介绍操作系统基本原理为主,依据资源管理观点论述操作系统基本原理、基本思想和基本方法,并在此基础上较完整地介绍 UNIX 操作系统的内核结构和使用方法,从而系统地、完整地讲述操作系统从基本原理到应用实践主要内容。本教材在内容上突出理论与实践相结合和能力的培养,并力求反映操作系统发展最新技术。本教材编写力求循序渐进、通俗易懂,每章都附有习题,便于教学和自学。

全书共分 8 章,第 1 章介绍操作系统发展过程、基本特性和主要功能;第 2 章详细阐述了进程基本概念、进程控制、进程同步、进程通信、进程调度和死锁,并简要介绍了能进一步提高程序并发执行程度线程概念;第 3 章介绍存储器管理方式,包括连续分配方式、离散分配方式和虚拟存储器。第 4、5、6 章分别介绍了设备管理、文件管理和作业管理;第 7 章介绍了一个操作系统的实例——UNIX 操作系统的内核结构;第 8 章介绍了 UNIX 系统的使用方法,主要内容为常用命令、文本编辑器 vi、Shell 程序设计、系统管理等。

本书以 UNIX System V 作为全书的范例,较完整地讲述它的内核结构和使用方法,便于读者从系统整体角度认识和理解操作系统,有助于读者领会、理解和加深各章所述内容。建议教学中讲完各章原理后或读者阅读各章后,可参考 UNIX System V 中的相关内容。

通过本书学习,可使学生从理论到实践较全面地理解现代操作系统基本概念、原理和方法,并在此基础上熟悉 UNIX 系统的具体实现和使用方法,为今后从事相关工作打下坚实基础。

目 录

第 1 章 操作系统概论	1
1.1 操作系统的概念	1
1.2 操作系统的发展和分类	4
1.3 操作系统的特性和功能	8
1.4 操作系统的进一步发展	12
习题	13
第 2 章 进程管理	14
2.1 进程的基本概念	14
2.2 进程控制	21
2.3 进程互斥和同步	24
2.4 进程通信	30
2.5 进程调度	34
2.6 死锁	39
2.7 线程	45
习题	47
第 3 章 存储器管理	50
3.1 存储器管理概述	50
3.2 连续分配存储管理方式	53
3.3 覆盖与对换技术	58
3.4 分页存储管理方式	61
3.5 分段存储管理方式	64
3.6 虚拟存储器的基本概念	70
3.7 请求分页存储管理方式	73
习题	78
第 4 章 设备管理	80
4.1 设备管理概述	80
4.2 I/O 控制方式	86

4.3	设备分配	90
4.4	虚拟设备技术	94
4.5	磁盘存储器管理	96
	习题	100
第5章	文件管理	101
5.1	文件和文件系统	101
5.2	文件的逻辑结构	103
5.3	外存分配方式	104
5.4	目录管理	108
	习题	112
第6章	作业管理	114
6.1	作业管理概述	114
6.2	作业调度	117
	习题	120
第7章	UNIX 系统内核结构	122
7.1	UNIX 系统概述	122
7.2	进程的描述和控制	125
7.3	进程的同步和通信	131
7.4	存储器管理	138
7.5	设备管理	141
7.6	文件管理	148
	习题	154
第8章	UNIX 系统的使用	156
8.1	UNIX 常用命令	156
8.2	文本编辑器 vi	170
8.3	Shell 程序设计	182
8.4	UNIX 系统管理	192
	习题	205
	主要参考文献	206

第 1 章

操作系统概论

计算机发展到今天,从个人计算机到巨型计算机系统,毫无例外地都要配置一种或多种计算机操作系统。如果让用户去使用一台没有操作系统的计算机,那将是难以想象的。操作系统 OS(Operating System)是现代计算机系统的重要组成部分,其他所有软件,例如,汇编程序、编译程序、数据库管理系统等系统软件以及各种应用软件,都要依赖于操作系统的支持,并取得它的服务。那么究竟什么是操作系统,操作系统在计算机系统中的地位等内容,将在这一章做简要阐述。为了阐明上述问题,回顾一下操作系统的形成和发展的过程以及理解各类操作系统的特性是必要的。为便于今后的学习,还要介绍操作系统的特性和功能。

1.1 操作系统的概念

1.1.1 计算机系统的层次结构

虽然任何一种计算机系统都要配置操作系统,并且计算机用户大多具有使用操作系统的经验,但要给操作系统下一个精确的定义并非轻而易举。

众所周知,计算机系统由两部分组成:计算机硬件和计算机软件。计算机硬件通常由 CPU(中央处理机)、存储器、输入/输出设备等部件组成,它构成了系统本身和用户作业赖以活动的物质基础和工作系统。计算机软件则包括系统软件和应用软件。系统软件如操作系统、多种语言处理程序(汇编和编译程序等)、连接装配程序、工具软件等;应用软件是为特定应用目的而编制的程序。

没有任何软件支持的计算机称为裸机(Bare Machine),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。图1.1展示了这种情形。

由图 1.1 可见,裸机处于最底层,裸机的外面是操作系统。操作系统是运行在计算机基本硬件系统上的最基本的系统软件。操作系统通过系统核心程序对计算机系统中主



图 1.1 计算机系统的层次结构

要几类资源进行管理,如:处理机、存储器、输入/输出设备、数据与文档资源、用户作业等,并为其他系统软件和应用软件提供运行支持和向用户提供若干服务。操作系统所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功能更强、使用更为方便的机器,并把所有对硬件的复杂操作隐藏起来,为用户提供一个友好、透明的操作环境。而位于操作系统之上的是另外一些系统软件,最外层是应用软件层。

1.1.2 什么是操作系统

从以上的分析可以看出,引入操作系统的目的可从两方面来考察:一方面,从系统管理人员的观点来看,引入操作系统是为了合理地去组织计算机工作流程;另一方面,从用户的观点来看,引入操作系统是为了给用户使用计算机提供一个良好的界面,用户无须了解有关硬件和系统软件的细节,就能方便灵活地使用计算机。

综上所述,可以对操作系统做如下的定义:操作系统是一组控制和管理计算机硬件和软件资源,合理地对各类作业进行调度,以及方便用户的程序的集合。

1.1.3 操作系统的目标和作用

(1) 操作系统的目标

在计算机系统上配置操作系统的主要目标,与计算机系统的规模和操作系统的运行环境有关。通常,对于配置在大、中型计算机系统上的操作系统,都有着较高的要求,相应地,其操作系统就具有较强的功能。又如,对应用于实时工业控制和武器控制环境下的操作系统,则要求其操作系统具有实时性和高度可靠性。

操作系统的目标通常有以下几点:

1) 方便性

配置操作系统后可使计算机系统更容易使用。一个未配置操作系统的计算机系统是极难使用的,因为计算机硬件只能识别0和1这样的机器代码。因此,用户要在计算机上运行自己所编写的程序,就必须用机器语言书写程序;用户要想输入数据或打印数据,也都必须自己用机器语言书写相应的输入程序或打印程序。如果在计算机硬件上配置了操作系统,用户便可通过操作系统所提供的各种命令来使用计算机系统。比如,编译命令可方便地把用户的用高级语言书写的程序翻译成机器代码,这样大大地方便了用户,从而使计算机变得易学易用。

2) 有效性

在未配置操作系统的计算机系统中,诸如CPU、I/O设备等各类资源,都会因经常处于空闲状态而得不到充分利用;内存及外存中所存放的数据可能由于无序而浪费存储空间。在配置了操作系统后,可使CPU和I/O设备由于能保持忙碌状态而得到有效的利用,并且由于可使内存和外存中存放的数据有序而节省存储空间。此外,操作系统还可以通过合理地组织计算机的工作流程,进一步改善资源的利用率及提高系统的吞吐量。

方便性和有效性是设计操作系统时最重要的两个目标。在过去的很长一段时间内,由于计算机系统非常昂贵,因而非常重视其有效性。正因如此,现在的大多数操作系统其理论上都着重于如何提高计算机系统的资源利用率和系统的吞吐量问题。然而,近年来在微机上所配置的操作系统的,则更重视其方便性。

3) 可扩充性

随着超大规模集成电路技术和计算机技术的迅速发展,计算机硬件和体系结构也随之得到迅速的发展,相应地,它们也对操作系统提出了更高的功能和性能要求。此外,计算机网络的发展,特别是 Internet 的发展,也对操作系统提出了一系列更新的要求。因此,操作系统必须具有很好的可扩充性,方能适应发展的要求。这就是说,操作系统在软件结构上应采用分层结构及微内核结构,以提高操作系统软件的可维护性(指便于增加新的功能层次和模块,并能方便地修改老的功能层次和模块)。

4) 开放性

20 世纪 80 年代以来,由于计算机网络的发展,尤其是局域网的迅速发展,使计算机操作系统的应用环境,已逐步由单机环境转向网络环境。为使来自不同厂家的计算机和设备能通过网络加以集成化,并能正确、有效地协同工作,实现应用的可移植性和互操作性,必须具有统一的开放环境,进而要求操作系统具有开放性。开放性是指系统能遵循世界标准规范,特别是遵循开放系统互连 OSI 国际标准。遵循国际标准所开发的硬件和软件,能彼此兼容,可方便地实现互连。

(2) 操作系统的作用

可以从不同的观点(角度)来观察操作系统的作用。从一般用户的观点,可把操作系统看做用户与计算机硬件系统之间的接口;从资源管理观点上看,则可把操作系统视为计算机系统资源的管理者。

1) 操作系统作为用户与计算机硬件系统之间的接口

操作系统作为用户与计算机硬件系统之间接口的含义是:操作系统处于用户与计算机硬件系统之间,用户通过操作系统来使用计算机系统。或者说,用户在操作系统帮助下,能够方便、快捷、安全、可靠地操纵计算机硬件和运行自己的程序。应注意,操作系统是一个系统软件,因而这种接口是软件接口。图 1.2 是操作系统作为接口的示意图。

由图 1.2 可以看出,用户可通过以下三种方式使用计算机:

①命令方式。这是指由操作系统提供了一组联机命令(语言),用户可通过键盘输入有关命令来直接操纵计算机系统。

②系统调用方式。操作系统提供了一组系统调用,用户可在自己的应用程序中通过相应的系统调用来操纵计算机。

③图形、窗口方式。用户通过屏幕上的窗口和图标来操纵计算机系统和运行自己的程序。

2) 操作系统作为计算机系统资源的管理者

操作系统的主要功能是实现对系统中四类资源进行有效的管理,即:处理机管理,用于分配和控制处理机;存储器管理,主要负责内存的分配与回收;I/O 设备管理,负责 I/O 设备的分配与操纵;文件管理,负责文件的存取、共享和保护。可见,操作系统是计算机系统资源的管理者。

3) 操作系统用作扩充机器

对于裸机来说,即使其功能再强大,也必定是难于使用的。如果在裸机上覆盖上一层 I/O 设备管理软件,用户便可利用它所提供的 I/O 命令,来进行数据输入和打印输出。此时用户所

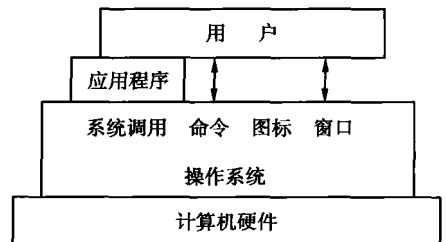


图 1.2 操作系统作为接口的示意图

看到的机器,将是一台比裸机功能更强、使用更方便的机器。通常把覆盖了软件的机器称为扩充机器或虚机器。如果又在第一层软件上再覆盖上一层文件管理软件,则用户可利用该软件提供的文件存取命令,来进行文件的存取。如果在文件管理软件上又再覆盖一层面向用户的窗口软件,则用户便可在窗口环境下方便地使用计算机,这样就形成了一台功能更强的虚机器。由此可知,每当人们在计算机系统上覆盖上一层软件后,系统功能便增强一级。由于操作系统自身包含了若干个层次,因此当在裸机上覆盖上操作系统后,便可获得一台功能显著增强、使用极为方便的多层扩充机器或多层虚机器。

1.2 操作系统的发展和分类

操作系统是由于客观的需要而产生的,它伴随着计算机技术本身及其应用的日益发展而逐渐发展和不断完善。它的功能由弱到强,在计算机系统中的地位不断提高。至今,它已成为计算机系统的核心,任何计算机系统都要配置操作系统。

为了更好地理解操作系统的基本概念、特性和功能,回顾一下操作系统形成和发展的历史过程是很有意义的。

1.2.1 无操作系统的计算机系统

(1) 人工操作方式

在 20 世纪 50 年代以前的第一代计算机时期,构成计算机的主要元器件是电子管,计算机运算速度慢,没有操作系统,甚至没有任何软件。用户直接用机器语言编制程序,并在上机时独占全部计算机资源。用户既是程序员,又是操作员。上机完全是手工操作:先把程序纸带(或卡片)装上输入机,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。计算完毕,用户取走并卸下纸带(或卡片);第二个用户上机,照此过程办理。

这种人工操作方式有以下两个方面的缺点:

1) 用户独占全机。计算机及其全部资源只能由上机用户独占。

2) CPU 等待人工操作。当用户进行装带(卡)、卸带(卡)等人工操作时,CPU 及内存等资源是空闲的。

可见,人工操作严重降低了计算机资源的利用率,此即所谓的人机矛盾。随着 CPU 速度的提高和系统规模的扩大,人机矛盾变得日趋严重。此外,随着 CPU 速度的迅速提高,CPU 和 I/O 设备之间速度不匹配的矛盾也变得更加突出。后来引入的脱机输入输出技术,较好地缓和了这些矛盾。

(2) 脱机输入输出(Off-Line I/O)方式

为了解决人机矛盾及 CPU 和 I/O 设备之间速度不匹配的矛盾,20 世纪 50 年代末出现了脱机输入/输出技术。该技术是指事先将装有用户程序和数据的纸带(或卡片)装入纸带输入机(或卡片机),在一台外围机的控制下,把纸带(卡片)上的数据(程序)输入到磁带上。当 CPU 需要这些程序和数据时,再从磁带上高速地调入内存。类似地,当 CPU 需要输出时,可由 CPU 直接高速地把数据从内存送到磁带上,然后再在另一台外围机的控制下,将磁带上的结果通过相应的输出设备输出。图 1.3 示出了脱机输入/输出过程。

由于程序和数据的输入和输出都是在外围机的控制下完成的,或者说,它们是在脱离主机的情况下进行的,故称为脱机输入/输出方式。与之相反,在主机的直接控制下进行输入/输出的方式称为联机输入/输出(On-Line I/O)方式。这种脱机输入/输出方式的主要优点如下:

1) 减少了 CPU 的空闲时间。装带(卡)、卸带(卡)以及将数据从低速 I/O 设备送到高速磁带(或磁盘)上,都是在脱机情况下进行的,都不占用主机时间,从而有效减少了 CPU 的空闲时间,缓和了人机矛盾。

2) 提高 I/O 速度。当 CPU 在运行中需要数据时,是直接从高速的磁带(或磁盘)上将数据调入内存的,不再是从低速 I/O 设备上输入,从而大大缓和了 CPU 和 I/O 设备速度不匹配的矛盾,进一步减少了 CPU 的空闲时间。

1.2.2 批处理系统

由于早期的计算机系统非常昂贵,为了能充分利用它,应尽量让系统连续运行,以减少系统的空闲时间,这样,就产生了最早的操作系统:批处理操作系统。批处理系统是将用户作业成批地提交给系统,由系统进行作业运行的控制,让作业能一个接一个地连续处理。所谓作业是指用户要求计算机系统所做的工作的集合。这些工作可能是一次计算过程、一次数据处理、一次信息查询等。批处理操作系统一般分为单道批处理系统和多道批处理系统。

(1) 单道批处理系统

单道批处理系统是早期批处理系统的类型。在单道批处理系统中,通常是把一批作业以脱机输入方式输入到磁带上,并在系统配置的监控程序的控制下使这批作业能一个接一个地连续处理。其自动处理过程是:首先由监控程序将磁带上的第一个作业装入内存,并把运行控制权交给该作业,当该作业的处理完成后又把控制权交还给监控程序,再由监控程序把磁带上的第二个作业调入内存。计算机系统就这样自动地一个作业接一个作业地进行处理,直至磁带上的所有作业都全部完成为止。由于系统对作业的处理是成批进行的,且在内存中始终只保持一道作业,故称此系统为单道批处理系统。

不难看出,单道批处理系统是在解决人机矛盾和 CPU 与 I/O 设备速度不匹配的矛盾过程中形成的。虽然单道批处理系统旨在提高系统资源利用率和系统吞吐量,但仍然不能很好利用系统资源。系统吞吐量指系统单位时间内完成的作业数。

单道批处理系统具有如下特征:

- 1) 自动性。在顺利情况下,磁带上的一批作业能自动地逐个依次运行,而无须人工干预。
- 2) 顺序性。磁带上的各道作业顺序地进入内存,正常情况下,各道作业完成的顺序与它们进入内存的顺序是一样的。
- 3) 单道性。内存中某一时刻只有一道作业在运行,仅在当前运行的一道作业运行完毕或发生异常情况时,才调入下一道作业进入内存后运行。

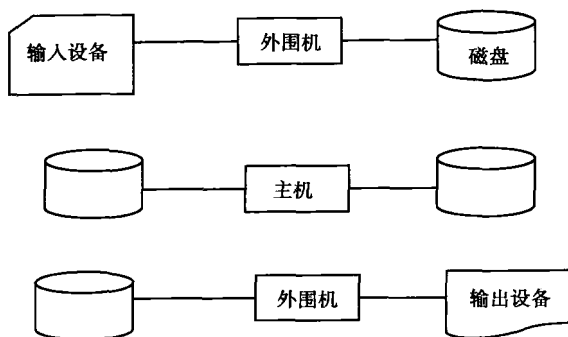


图 1.3 脱机 I/O 示意图

(2) 多道批处理系统

多道批处理是基于多道程序设计技术的应用,在内存中可以同时存放多道作业,并允许这些作业在系统中交替地运行,由此而形成了多道批处理系统。在多道批处理系统中,用户通过系统提供的各种功能,如作业控制语言、命令、程序等,将用户程序、数据等提交给系统,所提交的作业都存放在外存上形成作业的“后备队列”;然后,系统根据一定的调度原则从后备队列中选取若干作业调入内存,系统按照多道方式组织在内存中的这些作业的运行。某一道作业运行完毕后,系统又将调入新的作业,内存中始终存放多个作业,这些作业交替地运行。这样,作业不断进入系统,又不断退出系统,形成源源不断的作业流,从而大大提高了系统的资源利用率和系统的吞吐量。

多道批处理系统具有如下特征:

1) 多道性。在内存中可以同时驻留多道程序,并允许它们并发执行,从而有效地提高了资源利用率和系统吞吐量。系统吞吐量是指系统在单位时间内完成的总工作量。

2) 无序性。多道作业完成的先后顺序与它们进入内存的顺序之间没有严格的对应关系。即先进入内存的作业可能后完成,后进入内存的作业可能先完成。

3) 调度性。作业从提交给系统开始直至最后完成,需要经过两种调度:一是作业调度,即按一定的作业调度算法,从外存的后备队列中选择若干个作业调入内存;二是进程调度,这是指按一定的进程调度算法,从已在内存的作业中选择一个作业,将处理机分配给它,使之执行。在本书后面章节中将对作业调度、进程调度进行详细讨论。

多道批处理系统的主要优点如下:

1) 资源利用率高。由于在内存中装入了多道作业,它们共享系统资源,使得系统中的各种资源得到了更充分的利用。

2) 系统吞吐量大。能提高系统吞吐量的原因主要有两个方面:第一,CPU 和其他资源保持忙碌状态;第二,仅当作业完成或无法运行时才进行切换,系统开销小。

多道批处理系统的主要缺点是:

1) 平均周转时间长。作业周转时间是指从作业进入系统开始,直至完成并退出系统为止所经历的时间。而平均周转时间是指一批作业中所有作业周转时间的平均值。在多道批处理系统中,由于作业要依次进行处理,因而作业的平均周转时间较长。

2) 无交互能力。用户一旦把作业提交给系统后就失去了对自己作业的控制,而由系统根据用户提供的作业说明书来控制作业的执行,这对用户修改和调试自己的程序是非常不方便的。

要实现多道批处理,系统需要解决这样一些问题:

1) 并行运行的程序要共享计算机系统的硬件和软件资源,既有对资源的竞争,但又须相互同步。因此同步与互斥机制成为操作系统设计中的重要问题。

2) 多道程序同时驻留在内存中,出现了内存不够用的问题,提高内存的使用效率的需要变得非常迫切。因此出现了诸如覆盖技术、对换技术和虚拟存储技术等内存管理技术。

3) 多道程序共同存在于内存,这就需要保证系统程序存储区和各用户程序存储区的安全可靠,由此提出了内存保护的要求。

多道程序系统标志着操作系统进入渐趋成熟的阶段。在操作系统中先后出现了处理机管理、存储器管理、设备管理、文件管理、作业管理等功能。

1.2.3 分时系统

在批处理方式下,用户以脱机操作方式使用计算机,用户在提交作业以后,在作业运行过程中,不管出现什么情况都不能加以干预。只有当该批作业处理结束后,用户才能根据得到的计算结果做下一步的处理。这样既能保证计算机的效率,又能方便用户使用计算机,成为一种操作系统新的追求目标。20世纪60年代中期,计算机技术和软件技术的发展使得这种追求成为可能。

随着CPU速度不断提高和采用分时技术,一台计算机可同时连接多个用户终端,而每个用户都可在自己的终端上联机使用计算机,就像自己独占机器一样。所谓分时技术,就是把处理机的运行时间分成很短的时间片,按照时间片轮流把处理机分配给各联机作业使用。若某个作业在分配给它的时间片内不能完成其计算,处理机会分配给另一作业使用,该作业则暂时中断,等待下一轮分配到时间片后再继续其运行。由于计算机速度很快,作业运行轮转得很快,给每个用户的感受是他独占了一台计算机。而每个用户都可以通过自己的终端向系统发出各种操作控制命令,完成作业的运行。虽然物理上只有一台计算机,但是每一个用户都可以得到及时的服务响应,每一个用户都感觉到是一台计算机在专门为它服务,这就是分时系统。多用户分时操作系统是当今计算机操作系统中最普遍使用的一类操作系统。

分时系统具有如下特征:

1)同时性(多路性)。若干用户通过各自终端同时使用一台计算机,从宏观上看,所有用户在同一时间并行工作,从微观上看各个用户是轮流使用计算机。

2)独立性。每个用户各占一个终端,彼此独立操作,互不干扰。用户感觉就像他自己独占主机。

3)及时性。系统能及时响应每一个用户的请求。

4)交互性。用户通过终端与计算机系统进行交互会话,用户通过终端发出命令和服务请求,系统通过终端向用户反馈信息。交互性是分时系统的一个重要特征,因此,分时系统也被称为交互系统。

如何确定分时时间片的长短问题是分时系统中的一个重要问题,它将直接影响到终端用户能否得到及时响应,即影响响应时间。顺便指出,分时系统的响应时间是指用户发出终端命令到系统进行响应做出应答所需的时间,它是衡量分时系统性能的指标。

1.2.4 实时系统

实时系统是为了满足实时问题对计算机应用所提出的特殊要求而出现的一种操作系统,它主要应用于需要对外部事件进行及时响应并处理的领域。所谓实时,是指“立即”、“马上”的意思。实时系统则是指系统对特定输入做出反应的速度足以控制发出实时信号的对象,或者说计算机能够及时响应外部事件的请求,在规定的短时间内完成对该事件的处理,并控制所有实时设备和实时任务协调一致地运行。

实时系统是另外一类联机的操作系统,它是随着计算机应用于实时控制和实时信息处理领域中而发展起来的。实时系统的主要特点是提供即时响应和高可靠性,也就是说,实时系统必须保证对实时信息的分析和处理的速度比其进入系统的速度要快,而且系统本身有很好的安全性和可靠性。因为像生产过程的实时控制、武器系统的实时控制、航空订票、银行业务等

实时事务系统,信息处理的延误或丢失往往会带来不堪设想的后果。实时系统往往具有一定的专用性。与批处理系统、分时系统相比,实时系统的资源利用率可能较低。

实时系统可以分为实时控制系统和实时信息系统,两者的主要区别一是服务对象,二是对响应时间的要求。实时控制系统通常指以计算机为中心的过程控制系统,也称为计算机控制系统。它既用于生产过程中的自动控制,也可用于监测指导性控制。实时信息系统通常指实时信息处理系统,它可以是主机型多终端的联机系统,也可以是远程在线式的信息服务系统,还可以是网络互连式的信息系统。作为信息处理的计算机要及时响应终端用户或远程终端用户的服务请求,进行信息的检索、存储、修改、更新、删除和处理,并及时对用户做出应答。

实时系统具有如下特征:

1)多路性。对实时控制系统,多路性主要表现在对多路的现场信息进行采集、处理和多个对象或多个执行机构进行控制;对实时信息系统,则指按分时原则为多个终端用户服务。

2)独立性。实时控制系统中信息的采集和对象的控制是彼此独立、互不干扰的。在实时信息系统中,每个终端用户在向实时系统提出服务请求时,也是彼此独立操作、互不干扰的。

3)及时性。实时系统的及时性是非常关键的,主要反映在对用户的响应时间要求。实时系统对响应时间的要求较之分时系统更为严格。分时系统的响应时间是以人们通常能够接受的等待时间来确定的,例如2~3 s。而实时系统的响应时间是以被控对象或信息处理服务请求所能接受的延迟来确定的,它可能是秒级,也可能是毫秒、微秒级。

4)交互性。由于实时系统绝大多数都是专用系统,因此对用户能进行的干预赋予了不同权限,例如,实时控制系统在某些情况下不允许用户干预,而实时信息系统只允许用户在其授权范围内访问有关的计算机资源。

5)可靠性。实时系统要求系统高度可靠,这也是实时系统最重要的设计目标之一。对实时控制系统,尤其是重大控制项目,如航空航天、核反应、武器控制等,任何差错都可能导致灾难性后果;对实时信息系统,则要求数据与信息的完整性。因此,在实时系统中,必须考虑容错机制,以保证系统和数据的安全。

1.3 操作系统的特性和功能

1.3.1 操作系统的特性

操作系统本身是一种复杂的系统软件,这种特殊的软件与其他的系统软件和应用软件有所不同,有它自己的特殊性,这些特殊性构成了操作系统的特性。上节介绍的批处理系统、分时系统、实时系统三种基本操作系统分别具有各自的特征,但它们又都有四个共同的基本特性:并发、共享、虚拟和异步性。理解和掌握操作系统内部的这些抽象结构和概念,对理解和掌握操作系统的核心是至关重要的。另外,随着操作系统的发展,这些概念还在不断地深化和演变,应予以充分注意。

(1) 并发(Concurrence)

所谓并发,是指两个或多个事件在同一时间间隔内发生。与并发对应的另外一个概念是并行,所谓并行,是指两个或多个事件在同一时刻发生。在单处理机系统中采用多道程序设计

技术后,可以实现硬件之间的并行操作和程序之间的并发执行。硬件之间的并行操作是一个微观概念,例如,当 I/O 设备在进行 I/O 操作时,CPU 可以进行计算工作。而程序之间的并发执行则是一个宏观上的概念,从宏观上看,一段时间内有多道程序在同时运行,但从微观上看,任意时刻却最多只能有一道程序真正在处理机上执行。即宏观上多道程序在并发运行,而微观上这些程序却是在交替执行的。通常把能够并发执行的程序称为并发程序,相应的系统称为并发系统。

现代计算机系统是一种并发系统,操作系统作为这种并发系统的管理者,它必须支持程序的并发执行。应当指出,通常意义上的程序只是一个静态实体,它们是不能并发执行的。为使程序能够并发执行,系统必须为每个程序建立进程。所谓进程,可简单地理解为程序的执行,它是一个活动实体,并作为系统运行和资源分配的基本单位。有关进程的详细内容将在第 2 章中讨论。

操作系统的并发性有效地提高了系统资源的利用率和系统的吞吐量,但也导致了操作系统的复杂化,使得操作系统必须具有控制和管理各种并发活动的能力。

(2) 共享(Sharing)

所谓共享是指系统中的资源可供多个用户(用户程序)共同使用而不是被某个用户(用户程序)独占。操作系统是一种资源管理程序,要负责对多个用户提出的系统资源共享请求进行协调和分配。由于资源属性的不同,对资源的共享方式也有所不同,一般分为互斥共享和同时共享。

1) 互斥共享

系统中的某些资源,虽然它们可以提供给多个用户作业使用,但在一段时间内只允许一个用户作业使用,其他欲访问资源的作业必须等待,仅当该用户作业访问完并释放该资源后,才允许另一用户作业对该资源进行访问。把在一段时间内只允许一个用户作业访问的资源称为临界资源,如打印机、磁带机等物理设备都属于临界资源,临界资源只能被互斥共享。

2) 同时共享

同时共享是指在一段时间内,多个用户作业可以同时使用系统中的某个资源。这里,“同时”的概念仍然是宏观的,微观上则是多个作业交替地对资源进行访问。例如,对磁盘存储设备的访问就是采用同时共享的方式。

可以看到,互斥共享和同时共享都是指在“一段时间内”的共享,只不过前者是在一段时间内独占,而后者是在一段时间内由多个作业同时使用。

并发和共享是操作系统的两个最基本的特征,它们又互为存在条件。一方面,资源的共享是以程序(进程)的并发执行为条件,没有程序(进程)的并发执行,就没有资源的共享;另一方面,若系统不能对资源实施有效的管理,协调好程序(进程)对共享资源的访问,也必将影响到程序的并发执行。

(3) 虚拟(Virtual)

在操作系统中的所谓“虚拟”,是指通过某种技术把一个物理上的实体映射为若干个逻辑上的对应物。这里的前者是实际存在的,而后者是虚的,是感觉性的存在。虚拟主要通过分时使用的方法实现。例如:在多道分时系统中,虽然物理上只有一个 CPU,但每个终端用户却都认为有一个单独的 CPU 在为自己服务,亦即,利用多道程序技术把一台物理上的 CPU 虚拟为了多台逻辑上的 CPU,也称为虚处理机。这种虚拟性还可在系统其他地方出现,如虚拟存储

器、虚拟设备等。以上提到的各种虚拟技术的具体实现将在以后章节中详细介绍。

(4) 异步性 (Asynchronism)

所谓异步性是指内存中的多个进程均按照各自独立的、不可预知的速度向前推进。在多道程序环境下,尽管允许多个进程并发执行,但由于资源等因素的限制,进程通常不能一直不停地执行,而是以走走停停的方式运行的。内存中的每个进程何时执行、何时暂停,以怎样的速度向前推进,每道程序总共需要多少时间才能完成等,都是不可预知的。很可能是先进入内存的作业后完成,而后进入内存的作业先完成。一句话,在多道程序环境下,进程是以异步方式运行的。

尽管进程的运行具有异步性,但只要运行环境相同,作业经多次运行,都会获得完全相同的结果。因此,异步运行方式是允许的,它是操作系统的一个重要特征。

值得注意的是,操作系统的异步性是并发与共享的必然结果。

1.3.2 操作系统的功能

在多道程序环境下,各个用户作业所需资源的总和一般要大于系统所拥有的资源量,这就必然引起多个作业之间对系统资源的相互竞争。为使多道程序能有条不紊地运行,操作系统必须提供对处理机、存储器、设备和文件的管理功能。此外,为了便于用户对操作系统的使用,还必须提供一个使用方便的用户接口。

(1) 处理机管理

在单道作业或单用户的情况下,处理机为一个作业或一个用户所独占,所以对处理机的管理非常简单。但在多道程序或多用户的情况下,要组织多个作业并发运行,就必须解决处理机的调度策略、分配实施和资源回收等问题,这就是操作系统的处理机管理功能。正是由于操作系统对处理机管理策略的不同,其提供的作业处理方式也就不同,例如批处理方式、分时处理方式和实时处理方式,从而呈现在用户面前的是具有不同性质、功能的操作系统。

在传统的多道程序系统中,处理机的分配和运行,都是以进程(或线程)为基本单位,因而对处理机的管理,可归结为对进程(或线程)的管理。进程管理包括进程控制、进程同步、进程通信和进程调度几个方面,详细的内容将在第2章中介绍。

(2) 存储器管理

在计算机中,存储器(也称为内存或主存)是程序运行、中间数据和系统数据存放的地方,因受硬件限制,存储器的存储容量是有限的。存储器管理的主要任务是为用户作业或进程分配并回收存储空间,并进行存储空间的优化管理,以提高存储器的利用率。存储器管理应具有内存分配、内存保护、内存扩充等功能。

1) 内存分配。在内存中除了操作系统和其他系统软件外,还有用户程序,特别当有多个用户程序共享存储器时,如何分配内存空间,以保证系统程序和各用户程序对存储区的需要,这就是内存分配的主要任务。

2) 内存保护。内存保护的主要任务是确保每道用户程序都在自己的内存空间中运行,互不干扰。进一步说,绝不允许用户程序访问操作系统的程序和数据,也不允许转移到非共享的其他用户程序中去执行。

3) 内存扩充。由于物理内存的容量有限,当用户作业所需要的内存容量超过计算机系统所提供的内存容量时,借助于虚拟存储技术,把内部存储器和外部存储器结合起来管理,从逻

辑上去扩充内存容量,为用户提供一个容量比实际内存容量大得多的虚拟存储器,而用户使用这个虚拟存储器和使用实际物理内存一样方便,这就是内存扩充所要完成的任务。

(3) 设备管理

现代计算机系统外部设备的发展异常迅速,如何响应用户的 I/O 请求、有效地分配和使用设备,如何协调处理机与设备操作的时间差异、提高设备利用率和系统总体性能,这就是设备管理的主要任务。设备管理具体应包含缓冲管理、设备分配和设备处理等方面。

(4) 文件管理

上面提到的三种管理都是针对计算机的硬件资源的管理,而文件管理则是对系统的软件资源的管理。

把程序和数据统称为信息或文件。当一个文件暂时不被使用时,就把它放到外部存储器(如磁盘、磁带、光盘等)上保存起来。这样,在外存上就会保存大量的文件。对这些文件如不能很好地管理,就会引起混乱,甚至会使文件遭受破坏。文件管理功能就是要解决外存上文件的组织和管理问题。

另外,文件管理还要解决信息的共享、保密和保护等问题。如果系统允许多个用户协同工作,那么就on应该允许用户共享某些信息文件。但这种共享应该是受控制的,所以就应该有相应的授权和保密机制。另外,还应有一定的保护机制,以免文件被非授权用户调用和修改,即使在意外情况下,如系统失效、文件使用不当时,也能尽量保护信息免遭破坏。

(5) 用户接口

操作系统与用户的接口也简称为用户接口。在早期的操作系统中,用户接口通常仅有命令和系统调用两种形式,前者供用户在终端键盘上使用,后者供用户在编写程序时使用。而在现代操作系统中,除上述两种接口外,还向用户提供图形接口。

1) 命令接口。在命令接口方式下,用户可以通过该接口向作业发出命令,以控制作业的运行。命令接口又可进一步分为联机命令接口和脱机命令接口两种。联机命令接口是为联机用户提供的,它由一组键盘命令和命令解释程序组成。这样,用户可通过输入不同的命令,来实现对作业的不同控制,直至作业完成。而脱机命令接口是为批处理作业的用户提供的,它由一组作业控制语言 JCL(Job Control Language)组成。批处理作业的用户可以使用 JCL 语言,把需要对作业进行的控制和干预事先写在作业说明书上来实现对作业的各种控制。

2) 程序接口。该接口是为用户程序在执行过程中访问系统资源而设定的,是用户程序取得操作系统服务的惟一途径。程序接口由一组系统调用组成,每个系统调用都是一个能完成特定功能的子程序,每当应用程序要求操作系统提供某种服务(功能)时,便调用具有相应功能的系统调用。

3) 图形接口。该接口采用图形化的操作界面,用非常容易识别的图标将系统的各种操作直观、逼真地表示出来。用户可通过鼠标、菜单和对话框来完成对应用程序和文件的操作。这样,用户就不需要去记忆操作系统命令以及它们的格式。另外,图形用户接口还可以方便地将文字、图形和图像集成在一个文件中:可在文字型文件中加入彩色图像,也可在图像中写入必要的文字,甚至可以把图像、文字和声音集成在一起。

20 世纪 90 年代推出的操作系统一般都提供了图形用户接口,如大家熟悉的 Windows 操作系统。