

全国计算机等级 考试四级教程



教育部考试中心

——操作系统原理 (2016年版)

高等教育出版社



全国计算机等级考试四级教程

——操作系统原理

(2016年版)

Quanguo Jisuanji Dengji Kaoshi Siji Jiaocheng
——Caozuo Xitong Yuanli

教育部考试中心

主编 陈向群

编者 向勇 王雷 马洪兵

高等教育出版社·北京

内容提要

本书根据教育部考试中心最新颁布的《全国计算机等级考试四级操作系统原理考试大纲(2013年版)》的要求编写而成。主要内容包括:操作系统的基本概念、基本结构及运行机制,进程线程模型,进程线程同步机制,内存管理方案,文件系统的设计、实现及性能提升,I/O设备管理,死锁的概念及解决方案。在本书的最后附有最新版的考试大纲、样题及参考答案。本书的编写目标是使考生具备操作系统基础知识,掌握操作系统中各部分功能的工作原理及关联性,熟悉各种典型实现技术,从而能够在应用开发中运用所掌握的知识。

本书可供报考四级操作系统原理的全国计算机等级考试考生使用,也可作为普通高等学校计算机操作系统课程的参考书。

图书在版编目(CIP)数据

全国计算机等级考试四级教程:2016年版.操作系统原理/教育部考试中心编.--北京:高等教育出版社,2015.12(2016.1重印)

ISBN 978-7-04-044311-0

I. ①全… II. ①教… III. ①电子计算机-水平考试-教材②操作系统-水平考试-教材 IV. ①TP3

中国版本图书馆CIP数据核字(2015)第273406号

策划编辑 何新权 责任编辑 何新权 柳秀丽 封面设计 杨立新 版式设计 马云
责任印制 毛斯璐

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街4号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	三河市华骏印务包装有限公司	网上订购	http://www.landaco.com
开 本	787mm×1092mm 1/16		http://www.landaco.com.cn
印 张	14.5	版 次	2015年12月第1版
字 数	350千字	印 次	2016年1月第2次印刷
购书热线	010-58581118	定 价	32.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物料号 44311-00

积极发展全国计算机等级考试 为培养计算机应用专门人才、促进信息 产业发展作出贡献

(序)

中国科协副主席 中国系统仿真学会理事长
第五届全国计算机等级考试委员会主任委员
赵沁平

当今,人类正在步入一个以智力资源的占有和配置,知识生产、分配和使用为最重要因素的知识经济时代,也就是小平同志提出的“科学技术是第一生产力”的时代。世界各国的竞争已成为以经济为基础、以科技(特别是高科技)为先导的综合国力的竞争。在高科技中,信息科学技术是知识高度密集、学科高度综合、具有科学与技术融合特征的学科。它直接渗透到经济、文化和社会的各个领域,迅速改变着人们的工作、生活和社会的结构,是当代发展知识经济的支柱之一。

在信息科学技术中,计算机硬件及通信设施是载体,计算机软件是核心。软件是人类知识的固化,是知识经济的基本表征,软件已成为信息时代的新型“物理设施”。人类抽象的经验、知识正逐步由软件予以精确地体现。在信息时代,软件是信息化的核心,国民经济和国防建设、社会发展、人民生活都离不开软件,软件无处不在。软件产业是增长快速的朝阳产业,是具有高附加值、高投入高产出、无污染、低能耗的绿色产业。软件产业的发展将推动知识经济的进程,促进从注重量的增长向注重质的提高方向发展。软件产业是关系到国家经济安全和文化安全,体现国家综合实力,决定 21 世纪国际竞争地位的战略产业。

为了适应知识经济发展的需要,大力促进信息产业的发展,需要在全民中普及计算机的基本知识,培养一批又一批能熟练运用计算机和软件技术的各行各业的应用型人才。

1994 年,国家教委(现教育部)推出了全国计算机等级考试,这是一种专门评价应试人员对计算机软硬件实际掌握能力的考试。它不限制报考人员的学历和年龄,从而为培养各行业计算机应用人才开辟了一条广阔的道路。

1994 年是推出全国计算机等级考试的第一年,当年参加考试的有 1 万余人,2012 年报考人数已达 549 万人。截至 2013 年年底,全国计算机等级考试共开考 38 次,考生人数累计达 5 422 万人,有 2 067 万人获得了各级计算机等级证书。

事实说明,鼓励社会各阶层人士通过各种途径掌握计算机应用技术,并通过等级考试对他们的能力予以科学、公正、权威性的认证,是一种比较好的、有效的计算机应用人才培养途径,符合我国的具体国情。等级考试同时也为用人单位录用和考核人员提供了一种测评手段。从有关公司对等级考试所作的社会抽样调查结果看,不论是管理人员还是应试人员,对该项考试的内容和

前 言

本书是根据教育部考试中心最新制订的《全国计算机等级考试四级操作系统原理考试大纲(2013年版)》和计算机专业基础课程“操作系统原理”的教学要求进行编写的。

本书主要内容包括:操作系统的基本概念、基本结构,操作系统运行机制,进程线程模型,进程线程同步机制,内存管理方案,文件系统的设计、实现及性能提升,I/O设备管理,死锁的概念及解决方案。

操作系统课程是计算机科学与技术专业的专业基础课,该课程不仅理论体系经典严谨,与实践结合紧密,同时又具有相关设计原理与技术发展迅速、课程知识深度和广度强的特点。通过学习,要求考生能掌握操作系统的基本概念和主要特征,深入理解操作系统运行机理,掌握并应用进程线程模型、进程线程调度算法,掌握并应用进程线程同步机制,掌握虚拟存储技术及虚拟页式存储管理方案的设计与实现,掌握并应用文件系统设计与实现,理解I/O设备管理及典型实现技术,理解死锁的概念及解决方法,了解操作系统技术的发展演化。

本书由北京大学陈向群主编,参加编写的人员包括陈向群(第1、2、6、8章)、向勇(第3、4章)、王雷(第5章)、马洪兵(第7章)。清华大学史美林教授对书稿进行了审阅。

由于编写时间仓促,编者水平有限,疏漏之处在所难免,望读者提出宝贵意见,以便修订时改正。

编者

目 录

第 1 章 操作系统概论	1	2.1.4 程序状态字 PSW	29
1.1 操作系统的概念	1	2.2 存储体系	29
1.1.1 计算机系统	1	2.2.1 存储器的层次结构	30
1.1.2 操作系统的定义	2	2.2.2 存储保护	31
1.1.3 操作系统的特征	3	2.3 中断与异常机制	32
1.1.4 研究操作系统的观点	4	2.3.1 中断与异常的概念	32
1.1.5 操作系统的功能	6	2.3.2 中断系统	33
1.2 操作系统的发展	9	2.3.3 中断优先级与中断屏蔽	37
1.2.1 手工操作	9	2.4 系统调用	38
1.2.2 监控程序(早期批处理)	10	2.4.1 系统调用简介	38
1.2.3 多道批处理	10	2.4.2 系统调用的处理过程	39
1.2.4 分时系统	11	2.5 I/O 技术	41
1.2.5 UNIX 通用操作系统	11	2.6 时钟	43
1.2.6 个人计算机操作系统	11	第 3 章 进程线程模型	44
1.2.7 Android 操作系统	12	3.1 多道程序设计模型	44
1.3 操作系统分类	12	3.1.1 程序的顺序执行	44
1.3.1 批处理操作系统	13	3.1.2 多道程序系统中程序执行环境的变化	45
1.3.2 分时系统	16	3.1.3 程序的并发执行	47
1.3.3 实时操作系统	17	3.2 进程模型	47
1.3.4 嵌入式操作系统	18	3.2.1 进程的概念	47
1.3.5 个人计算机操作系统	18	3.2.2 进程的状态及其状态转换	48
1.3.6 网络操作系统	18	3.2.3 进程控制块	52
1.3.7 分布式操作系统	19	3.2.4 进程控制	53
1.3.8 智能卡操作系统	20	3.3 线程模型	57
1.4 操作系统结构	21	3.3.1 线程的引入	57
1.4.1 整体式结构	21	3.3.2 线程的基本概念	58
1.4.2 层次结构	22	3.3.3 线程实现机制	59
1.4.3 微内核(客户机/服务器)结构	23	3.3.4 Pthread 线程包	62
第 2 章 操作系统运行机制	25	3.4 进程(线程)调度	63
2.1 中央处理器(CPU)	25	3.4.1 概述	64
2.1.1 CPU 的构成与基本工作方式	25	3.4.2 调度算法设计原则	64
2.1.2 特权指令和非特权指令	27	3.4.3 进程(线程)调度算法	67
2.1.3 处理器的状态	28		

第4章 并发与同步	73	6.2.2 文件的物理结构	135
4.1 进程(线程)间相互作用	73	6.2.3 文件的存储介质	140
4.2 进程互斥	76	6.2.4 文件的存取方式	143
4.3 信号量(Semaphore)	80	6.3 文件目录	144
4.4 经典的进程同步问题	81	6.3.1 文件目录的组成	144
4.5 管程	88	6.3.2 文件目录结构	146
4.6 进程通信	93	6.3.3 树形目录	148
4.6.1 共享内存	94	6.3.4 路径名	150
4.6.2 消息机制	94	6.3.5 目录操作	151
第5章 内存管理	98	6.4 文件系统的实现	151
5.1 基本概念	98	6.4.1 存储空间的分配与回收	152
5.1.1 存储体系	98	6.4.2 实现文件系统的表目	155
5.1.2 存储管理的任务	99	6.4.3 记录的成组与分解	157
5.1.3 地址转换	101	6.4.4 文件的操作	159
5.1.4 内存管理方案简介	103	6.5 文件的保护和安全	161
5.2 分区存储管理方案	104	6.5.1 文件的共享	161
5.2.1 固定分区	104	6.5.2 文件的保护	162
5.2.2 可变分区	105	6.5.3 文件的存取权限	163
5.2.3 分区管理方案的优缺点	111	6.5.4 文件的保密	166
5.3 覆盖技术与交换技术	111	6.6 文件系统的性能	166
5.3.1 覆盖技术	112	6.7 Windows 的 FAT 文件系统和	
5.3.2 交换技术	113	UNIX 文件系统	170
5.4 页式存储管理方案	114	6.7.1 Windows 的 FAT 文件系统	171
5.4.1 基本思想	114	6.7.2 UNIX 文件系统	172
5.4.2 存储空间的分配与回收	115	第7章 I/O 设备管理	174
5.4.3 地址转换与快表	115	7.1 设备与设备分类	174
5.5 虚拟存储技术与虚拟页式存储		7.1.1 设备管理的重要性	174
管理方案的实现	119	7.1.2 设备管理的任务	175
5.5.1 虚拟存储技术	119	7.1.3 设备的分类	175
5.5.2 虚拟页式存储管理	120	7.2 I/O 硬件组成	176
5.5.3 段式与段页式存储管理方案	128	7.2.1 计算机 I/O 系统的结构	176
第6章 文件管理	130	7.2.2 I/O 设备数据传送控制方式	178
6.1 概述	130	7.3 I/O 软件的特点及结构	182
6.1.1 文件和文件系统	131	7.3.1 设备驱动程序	182
6.1.2 文件分类	132	7.3.2 与设备无关的系统软件	184
6.1.3 文件系统的功能	133	7.3.3 用户空间的 I/O 软件	185
6.2 文件的结构	134	7.4 典型的 I/O 技术	187
6.2.1 文件的逻辑结构	134	7.4.1 缓冲技术	187

7.4.2 设备分配技术	189	8.3.1 安全与不安全状态	203
7.5 I/O 性能问题及解决方案		8.3.2 银行家算法	204
小结	191	8.4 死锁检测与解除	207
第 8 章 死锁	193	8.4.1 死锁检测	207
8.1 死锁基本概念	193	8.4.2 死锁解除	208
8.1.1 死锁的概念	193	8.5 资源分配图	210
8.1.2 死锁产生的原因	197	8.5.1 死锁的表示——资源分配图	210
8.1.3 产生死锁的必要条件	198	8.5.2 死锁判定法则	211
8.1.4 解决死锁的方法	199	8.5.3 资源分配图化简法	212
8.2 死锁预防	199	附录 1 全国计算机等级考试四级操作	
8.2.1 破坏“互斥条件”	200	系统原理考试大纲(2013 年版)	214
8.2.2 破坏“不可剥夺”条件	200	附录 2 全国计算机等级考试四级操作	
8.2.3 破坏“请求和保持”条件	200	系统原理样题及参考答案	217
8.2.4 破坏“循环等待”条件	201	参考文献	220
8.3 死锁避免	203		

操作系统是一门理论与实践结合非常紧密的课程。操作系统也是一个可以从多个视角加以考察的一个软件系统。从计算机应用角度看,操作系统是几乎人人都要使用的系统界面和接口;而从软件设计和开发的角度看,操作系统起着系统软件开发基础和工具的作用;而在黑客和网络攻击者看来,操作系统是他们要攻破的第一道防线……

本章将对操作系统作一个概括性的介绍,使读者在学习和理解操作系统基本原理之前,能够对操作系统这门课程有一个整体的认识 and 了解。本章首先阐述操作系统的基本概念,分别介绍计算机系统、操作系统的定义、功能、特征和研究操作系统的几种观点等方面的内容,并按照时间顺序介绍操作系统的由来和发展历史,以便读者对操作系统的基本概念有较为具体的认识;同时对各类操作系统的工作原理和特点进行概要性介绍。本章的最后一节简要介绍典型的操作系统结构。

1.1 操作系统的概念

本节首先介绍计算机系统,然后分别讲述操作系统的定义、特征、功能以及研究操作系统的几种观点。

1.1.1 计算机系统

计算机系统是一种可以按用户的要求接收和存储信息,自动进行数据处理并输出结果信息的系统。计算机系统包括硬件(子)系统和软件(子)系统。硬件系统是计算机系统赖以工作的实体;软件系统则保障计算机系统按用户指定的要求协调工作。这两个部分构成了计算机系统的资源。因此,计算机系统的资源包括两大类:硬件资源和软件资源。

中央处理器(CPU)、内存储器、外存储器(磁盘、磁带等)以及各种类型的输入/输出设备(键盘、鼠标、显示器、打印机等)组成了计算机硬件系统;而各种程序和数据则组成了计算机的软件系统。一个完整的计算机系统如图 1-1 所示。

计算机系统	软件系统 (程序、数据)	应用软件:文字处理、图形图像处理、科学计算、MIS 等
		支撑软件:数据库、网络、多媒体等
		系统软件:操作系统、编译器等
	硬件系统	中央处理器(CPU)、内存储器、 外存储器(磁盘、磁带等) 输入/输出设备(键盘、鼠标、显示器、打印机等)

图 1-1 计算机系统

任何一个程序在计算机系统中执行前,必须先要得到计算机系统的内存空间后才能被装入内存,程序的执行要依靠中央处理器完成,并且程序在执行过程中需要调用公共服务子程序和共享文件,还要使用各种计算机外部设备,以完成信息的输入和输出。这里的中央处理器、内存空间、公共服务子程序和共享文件以及计算机外部设备等都是计算机系统资源。可见,任何程序的执行都要占用计算机系统的资源。

但是,各个程序对计算机资源的要求和使用是不同的。不同用户对计算机资源的请求和使用还可能会产生冲突。比如,正当一个用户在使用打印机时,另外一个用户也要求使用这台打印机。可以想象,如果没有一种方法或规则对用户使用打印机的要求进行管理,肯定会引起混乱。同样,对计算机系统资源中的其他资源也需要进行管理。因此,适当地控制和协调计算机资源的分配是必要的。

用户在使用计算机系统时,当然希望计算机能够按照用户的要求工作,控制程序按照用户各自的意愿执行。所以,计算机系统必须有识别用户要求的能力,还必须具有针对不同用户进行不同控制执行的能力。

在计算机系统中,集中了资源管理功能和控制程序执行功能的一种软件,称为操作系统。

1.1.2 操作系统的定义

下面给出操作系统的定义:

操作系统是计算机系统中的一个系统软件,它是这样一些程序模块的集合——它们能有效地组织和管理计算机系统中的硬件及软件资源,合理地组织计算机的工作流程,控制程序的执行,并向用户提供各种服务功能,使用户能够灵活、方便、有效地使用计算机,并使整个计算机系统能高效地运行。

从表现形式上看,操作系统是一个大型软件程序,其中有大量的为实现操作系统各项功能的软件程序,这些软件程序从设计和结构角度看,通常是模块化的。有的操作系统带有内核,其内核也由各种程序模块组成。有的操作系统本身是层次化的,每一个层次的程序通常也是模块化的。还有的操作系统是客户机/服务器结构的,而服务器或客户机内部程序一般也是模块化的。总结起来,操作系统这样一个大型软件程序,是具有各种功能的、大量程序模块的集合。

操作系统的任务之一是组织和管理计算机系统中的硬件及软件资源。在操作系统内部,为了掌握整个计算机系统的硬件和软件资源,设计了各种不同类型的表格或数据结构,将所有的硬件和软件资源一一加以登记。资源的名称、类型、数量、用途、完好状态以及目前使用状态等信

息,均在有关数据结构中保存,并动态、实时地不断更新着。记录数据的目的是让操作系统根据用户对各种资源的需求情况,资源的当前分配和使用情况以及有关的资源调度策略,对资源进行有效的组织和管理。

这里,“有效”是指根据用户的不同要求,在管理计算机资源时考虑到系统运行的效率和资源的利用率。要尽可能提高中央处理器的利用率,让它尽可能少地空转,应该在保证访问效能的前提下尽可能有效地利用其他资源,例如减少对内存、硬盘空间等的浪费。

在计算机系统中,不同的用户有不同的要求。系统中的作业有大有小,有的作业需要运算时间长,有的作业有大量的信息需要打印输出,有的事件需要立即处理。比如,笔记本电脑的电池还有30秒钟即将没电了,需要立即通知用户,要么给电脑提供新的电源,要么赶紧保存重要的文档。而有的事件推迟一些时间处理,不会引起什么大的乱子,比如,通知用户信箱中刚刚收到一封电子邮件,早10秒钟发出通知还是晚10秒钟发出通知,估计不会对用户造成多大影响。操作系统应该根据用户不同的要求,作业的各自特点以及紧急程度等,合理地组织计算机的工作流程,并控制程序的执行。

这里的“合理”是指操作系统要“公平”对待不同的用户程序,保证系统不发生“死锁”和“饥饿”的现象。

操作系统的另一项重要任务,是向用户提供各种服务功能。一方面,向程序开发和设计人员提供高效的程序设计接口;另一方面,向使用计算机系统的用户提供接口,使用户能够灵活、方便、有效地使用计算机。

这里的“方便”是指操作系统的人机界面要考虑用户使用界面和程序设计接口两个方面的易用性、易学性和易维护性。

1.1.3 操作系统的特征

作为一种系统软件,操作系统有着与其他软件相比所不同的特征。

1. 并发性

并发性是指在计算机系统中同时存在若干个运行着的程序。从宏观上看,这些程序在同时向前推进。

进一步分析,计算机程序的并发性体现在如下两个方面:用户程序与用户程序之间并发执行;用户程序与操作系统程序之间并发执行。

实际从微观上看,在单处理器的环境下,这些同时运行着的程序是交替在中央处理器上运行的。在多处理器系统的环境中,多个程序的并发特征,不仅在宏观上是并发的,而且在微观上,即在处理器一级上,程序也是并发执行的。而在分布式系统中,多个计算机的并存,使得程序的并发特征得到了更充分的体现,因为在每个计算机上都可以有程序执行,它们共同构造了程序并发执行的图景。

2. 共享性

共享性是指操作系统程序与多个用户程序共用系统中的各种资源。这种共享性是在操作系统控制下实现的。资源的共享性主要针对计算机系统中的如下几项重要资源:

(1) 中央处理器。中央处理器是所有程序都必须使用的最重要的资源,操作系统必须采用恰当的调度策略,对多个并发程序分配处理器资源。

(2) 内存储器。任何一个程序必须首先调入内存之后,才能执行。管理内存的公共使用是操作系统的重要功能之一。

(3) 外存储器。外存储器主要用来保存各种程序和数据。这些程序和数据一般以文件的形式存放在外存储器上。外存储器有各种类型,如硬盘、软盘、磁带、可读写光盘等,这些外存储器都是为所有的程序和数据共享的。操作系统同样必须对外存储器的共享进行管理,并且保证在外存储器中所有程序和数据的完整性和正确性。

(4) 外部设备。我们知道,计算机系统的外部设备是供所有用户使用的,这些用户包括操作系统、系统用户(如管理员)和普通用户。操作系统必须对这些用户对外部设备的使用要求作出管理,使所有的外部设备能够按照一定的策略被全部用户所共享。

在计算机系统中,对资源的共享一般有两种形式:互斥共享和同时共享。

- 互斥共享。系统中的有些资源比如打印机、磁带机、扫描仪等,虽然可以供多个用户程序同时使用,但是在一段特定的时间内只能由某一个用户程序使用。当这个资源正在被使用的时候,其他请求该资源的程序必须等待,并且在这个资源被使用完了以后才由操作系统根据一定的策略再选择一个用户程序占有该资源。通常把这样的资源称为临界资源。许多操作系统维护的重要系统数据都是临界资源,它们都要求被互斥共享。

- 同时共享。系统中还有一类资源,它们在同一段时间内可以被多个程序同时访问。需要说明的是,这种同时访问是指宏观上的同时,微观上这些程序访问这个资源有可能还是交替进行的,而且它们交替访问这个资源的顺序对访问结果没有什么影响。一个典型的可以同时共享的资源就是硬盘,当然,那些可以重入的操作系统代码也是可以同时共享的。

3. 随机性

操作系统的运行是在一种随机的环境下进行的。这种随机环境的含义是,操作系统不能对所运行的程序的行为以及硬件设备的情况作出任何事先的假定。显然,一个设备可能在任何时候向处理器发出中断请求,而用户提交的、运行着的程序要做什么事情也无法事先知道。因而一般来说,操作系统正处于什么样的状态之中是无法确切知道的,这就是随机性的含义。但是,随机性并不意味着操作系统不能很好地控制资源的使用和程序的运行,随机性突出强调了在进行操作系统的设计与实现时要充分考虑各种各样的可能性。操作系统本身应该稳定、可靠、安全、高效,实现程序并发和资源共享的目的。

➤ 1.1.4 研究操作系统的观点

操作系统是计算机系统中的一个极为重要的部分,担负着非常重大的职责,可以说操作系统身兼数职。要理解操作系统,有必要深入地了解 and 认识操作系统的本质,即操作系统究竟是什么。

对操作系统本质的不同看法,会影响操作系统的设计思想。本节介绍研究操作系统的几种观点。

1. 软件的观点

从软件的观点来看,操作系统是一种大型软件系统,它是多种功能程序的集合。作为一种大型软件系统,操作系统有软件的外在特性和内在特性。

所谓操作系统的软件外在特性是指,操作系统是一种软件,它的外部表现形式,即它的操作

命令定义集和它的界面,完全确定了操作系统这个软件的使用方式。比如,操作系统的各种命令、各种系统调用及其语法定义等。从操作系统的使用界面上,即从操作系统的各种命令、系统调用及其语法定义等方面学习和研究操作系统,我们才能从软件的外部特征上把握住操作系统的性能。

所谓操作系统的软件内在特性是指,操作系统既然是一种软件,它就具有一般软件的结构特点。然而这种软件又不是一般的应用软件,它具有一般应用软件所不具备的特殊结构。因此,学习和研究操作系统时,就需要研讨其软件结构上的特点,从而更好地把握操作系统的内部结构特点。比如,操作系统是直接同硬件打交道的,那么操作系统中同硬件交互的软件是如何组成的,每个组成部分的功能作用和各部分之间的关系等,都是需要研究的。

2. 资源管理的观点

在计算机系统中,硬件和软件资源可以分成以下几部分:中央处理器、存储器(内存和外存)、外部设备和信息(文件)。现代计算机系统支持多个用户、多道作业,那么,面对众多的用户程序、作业争夺处理器、存储器、设备和共享软件资源,如何协调这些用户程序和作业呢?如何有条不紊地进行资源分配呢?

操作系统就是负责记录谁在使用什么样的资源,系统中还有哪些资源空闲,当前响应了谁对资源的要求,以及收回哪些不再使用的资源等。操作系统要提供一些机制去协调程序间的竞争与同步,提供机制对资源进行合理使用,施加保护,以及采取虚拟技术来“扩充”资源等。图1-2示意了操作系统管理的基本资源。

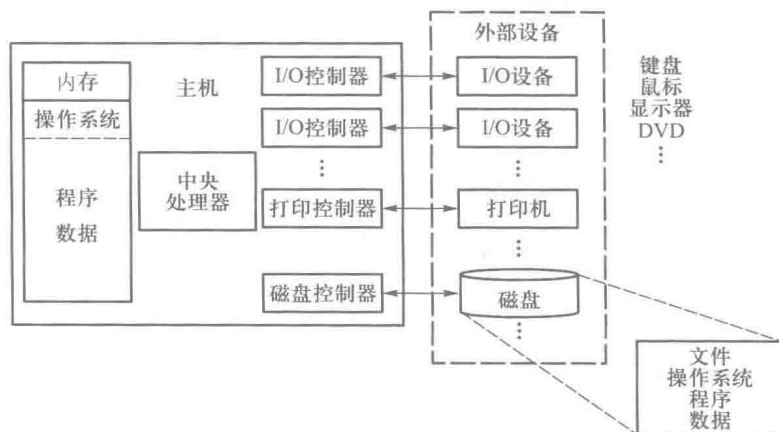


图 1-2 操作系统管理的基本资源

3. 进程的观点

简单地讲,进程可以看作运行中的程序。采用进程的观点,操作系统可看作是由多个可以同时独立运行的程序和一个对这些程序进行协调的核心所组成。每个进程都完成某一特定任务,任务可以来自用户提交的作业,也可来自操作系统(例如,控制用户作业的运行,处理某个设备的输入/输出……)。可见,进程可以分为用户进程和系统进程两大类,而操作系统则控制和协调这些进程的运行。

采用进程的观点时,侧重于分析计算机系统各部分的并行工作,研究处理各项管理任务的分

割以及这些管理任务相互之间的关系。进程在共享资源时所产生的竞争问题,通过进程之间的通信来解决。

4. 虚拟机的观点

在操作系统的支持下,用户不需要直接操作硬件机器(称为裸机),而是通过操作系统提供的各种手段来控制和使用计算机。操作系统把原来的计算机(裸机)扩充成为功能强、使用方便的计算机系统,我们把这种计算机系统称为虚拟计算机,而把操作系统的全部功能,包括系统调用、命令、作业控制语言等,统称为操作系统虚拟机。

虚拟机的观点是从系统功能分解的角度出发来考虑操作系统的结构。这种观点将操作系统的功能分成若干个层次,每一层次完成特定的功能,从而构成一个虚拟机,并为上一层次提供支持,构成它的运行环境。通过逐个层次的功能扩充,最终完成整个操作系统虚拟机的构造,向用户提供各种服务,完成用户的作业要求。

5. 服务提供者观点

操作系统提供了一系列的功能和便利的工作环境为用户服务,所以可以把操作系统看作是服务提供者。从用户的角度,站在操作系统之外观察操作系统,则可以认为该服务提供者为用户提供了比裸机功能更强,服务质量更高,更方便、灵活的虚拟机。为用户使用便利,该服务提供者提供了一组功能强大、方便、易用的广义指令(称为系统调用)。

➤➤ 1.1.5 操作系统的功能

操作系统在计算机系统中具有非常重要的地位,承担着计算机系统中全部的管理功能。按照资源管理的观点,操作系统的这些功能主要可以分为进程管理(处理器管理)、存储管理、文件管理、作业管理和设备管理。下面分别对它们进行简要的分析介绍。

1. 进程管理

进程管理的实质是对中央处理器进行管理,所以进程管理往往又称为处理器管理。CPU 是计算机系统中最宝贵的硬件资源,为了提高 CPU 的利用率,现代操作系统都采用了多道程序技术。如果一个程序因等待某一条件而不能运行下去时,就把处理器占用权转交给另一个可运行程序。或者,当出现了一个比当前运行的程序更重要的可运行的程序时,后者应能抢占 CPU。为了描述多道程序的并发执行,就要引入进程的概念。通过对进程的管理,协调多道程序之间的关系,解决有关处理器分配调度策略、分配的实施和回收等问题,以使 CPU 资源得到最充分的利用。

由于操作系统对处理器管理策略的不同,操作系统对其作业的处理方式也就不同,例如有批处理方式、分时处理方式和实时处理方式等,从而有各种不同性质的操作系统呈现在用户面前。

进程管理主要包括进程控制、进程同步、进程间通信和进程调度等几方面的内容。其中进程控制主要处理进程的创建、状态转换、进程撤销以及相关的进程资源的分配与回收等事务;进程同步主要处理进程之间的关系,包括进程的同步和互斥;进程间通信主要处理相互协作进程之间信息的交换问题;而进程调度则是按照一定的算法从就绪队列中挑选一个进程在处理器中真正执行它。下面分别叙述进程管理的主要内容。

1) 进程控制

在多道程序环境下,进程是操作系统进行资源分配的单位。在进程创建时,系统要为进程分

配各种资源,例如内存、外设等;在进程退出的时候,系统要从进程空间中回收被分配给它的资源。进程控制的主要任务就是创建进程、撤销结束的进程以及控制进程运行时候的各种状态转换。

2) 进程同步

多个进程的执行是并发的,它们以异步的方式运行,它的执行进度也是不可预知的。为了使多个进程可以有条不紊地运行,操作系统要提供进程同步机制,以协调进程的执行。一般有两种协调方式:互斥和同步。互斥是指多个进程对临界资源访问时采用互斥的形式;同步则是在相互协作共同完成任务的进程之间,用同步机制协调它们之间的执行顺序。

最简单的实现互斥的方法就是给资源加锁,并提供操纵锁变量的原语。包括开锁和关锁原语。所谓原语就是指具有某种功能,运行时有“原子性”的小段程序。原子性可保证这一段程序要么全部被执行,要么全部不起作用,即这一操作不可以被进一步分割或者打断,如同人们一度认为原子是物质的最小单位,不可再分割一样。有关进程同步与互斥等内容在后面的章节中将详细讲述。

3) 进程间通信

进程间通信主要发生在相互协作的进程之间。既然需要协作,那么协作的进程之间,就存在着信息或数据的交换。由操作系统提供的进程间通信机制是协作的进程之间相互交换数据和消息的手段。一个比较典型的例子是通过网络的在线流式媒体播放。流式媒体源获取线程(线程是现代操作系统中处理器时间分配的基本单位,它代表一个指令的执行流及执行的上下文环境,一个进程可以包含多个线程,这个概念将在后面的章节中详细描述),负责将媒体数据从远端的服务站点下载到本地的数据缓存中;流媒体的播放线程负责将缓存中的数据进行流媒体数据分离(视频流和音频流)和解码,还原成实际的图像帧和声音数据;然后渲染线程将把前面的数据送往显示设备和声音设备。整个播放过程中,为了保证播放的流畅,这三个线程的执行是有一定关系的:原始数据获取的速度要比解码的速度快,解码又要和渲染输出的速度匹配。这种执行速度的协调以及数据在不同线程之间的传递就需要进程间通信和进程同步机制来共同保证。

4) 调度

调度又称为处理器调度,通常包括进程调度、线程调度和作业调度等。进程(线程)调度的任务就是从进程(线程)的就绪队列中按照一定的算法挑选出一个,把处理器资源分配给它,并准备好特定的执行上下文让它执行起来。作业调度的基本任务则是从作业后备队列中按照一定的算法挑出若干个作业,并依照作业说明书为它们分配一定的资源,把它们装入内存并为每个作业建立相应的进程。

2. 存储管理

存储管理的任务是管理计算机内存的资源。在当代计算机系统中,计算机内存空间越来越大,2 GB 的内存空间已是很普通的容量。虽然内存空间可以扩充,但是内存空间再大也跟不上运行程序对内存空间的巨大需求。而且内存空间还受到 CPU 寻址能力的限制,内存的容量不能无限扩充。

概括地说,存储管理有三个方面的任务。首先,当多个程序共享有限的内存资源时,要考虑如何为多个程序分配有限的内存空间;其次,存放在内存中的多个程序和数据应该彼此隔离、互不侵扰;第三,解决内存扩充的问题,即将内存和外存结合起来管理,为用户提供一个容量比实际



内存大得多的虚拟存储器。下面分别就这三项存储管理功能作进一步的叙述。

1) 内存的分配与回收

操作系统要为每个进程分配内存空间,在分配的过程中,还要尽可能提高内存资源的使用效能。对于已经退出运行的进程所占据的内存空间,操作系统还要把它们加以回收,重新使用。人们在使用计算机时经常会发现,原本很大的内存在使用一段时间以后,似乎丢失了一部分空间,仔细检查会看到在空间使用图中有一大堆空间碎片,这种现象是由内存的分配与回收算法中的缺陷造成的。可见内存的分配与回收对计算机系统的有效运行的影响是很大的。

2) 存储保护

由于内存是供多个程序共享的,从理论上讲,每个程序都应该在分配给自己的内存区中运行。但是,从实际上看,必须考虑某个程序可能发生越界的情况。这是因为,首先,程序自身可能存在一些错误或隐患,造成程序的突然越界运行,不仅自身发生问题,而且影响或破坏了其他用户程序的运行。如果发生故障的程序越界到操作系统运行的内存空间,那么受到危害将更加严重,整个计算机系统有可能崩溃。其次,如果某个程序具有恶意,比如是病毒程序,或者是外界的人侵程序,那么就更有必要防止它们危害其他用户和整个计算机系统了。

在现代的计算机系统中,通常提供由硬件实现的存储保护机制,操作系统则利用这一机制实现进程的地址保护。

3) 内存扩充

我们已经知道,物理内存的大小总是有限的,各种程序对运行空间的需求经常超过实际内存的大小。内存扩充功能就是借助于虚拟存储技术在逻辑上增加进程运行空间的大小,这个大小比实际的物理内存要大得多。例如,在 x86 系统中目前常见的台式个人计算机系统中,物理内存的大小不超过 1 GB 字节,而一个进程的地址空间可有 4 GB 字节大小,各个进程空间总和则可以达到 64 TB 字节,远远大于 1 GB 字节的物理内存。

在虚拟存储技术中,操作系统通常将空间划分为一个个 4 KB 字节大小的页面,并且以页面为单位进行存储空间的调度。进程所需的总页面数所对应的空间可以超出实际物理空间。但是操作系统只把正在使用的页面保持在内存中,或把即将使用的页面调入到内存中,而将不用的页面调换到外存上,用户感受不到实际内存对使用空间的限制,仿佛有一虚拟的大容量内存存在。当然,系统要实现这一机制,必须提供请求页面调入的功能和页面置换的功能。

3. 文件管理

在计算机系统上的信息资源(如程序和数据)是以文件的形式存放在外存储器(如磁盘、磁带)上的,需要时再把它们装入内存。操作系统一般都提供很强的文件系统。文件管理的任务是有效地支持文件的存储、检索和修改等操作,解决文件的共享、保密和保护问题,以使用户方便、安全地访问文件。

1) 文件存储空间的管理

任何文件存放在外存储器上,都需要以某种形式占据外存储器的空间。文件系统的一个很重要的功能就是为每个文件分配一定的外存空间,并且尽可能提高外存空间的利用率和文件访问的效能。文件系统设置专门的数据结构记录文件存储空间的使用情况。为了提高空间利用率,存储空间的分配通常采用离散分配方式,以 512 字节或者几千字节的块为基本单位进行分配。

2) 目录管理

目录管理的主要任务就是给出组织文件的方法,它为每个文件建立目录项,并对众多的目录项加以有效的组织,以便为用户提供方便的按名存取。

3) 文件系统的安全性

安全性包括文件的读写权限管理以及存取控制,用以防止未经核准的用户存取文件,防止越权访问文件,防止使用不正确的方式访问文件。

4. 设备管理

在现代计算机系统中,通常存在着大量的外部设备,从键盘、鼠标、显示器到彩色打印机、数字音响设备和 DVD 等。这些外部设备的种类繁多,功能差异很大。如果要求每个用户亲自对连接到计算机系统各种不同的外部设备进行控制操作,实际是行不通的。

操作系统应该向用户提供设备管理。这里的设备管理是指计算机系统中除了 CPU 和内存以外的所有输入、输出设备的管理。由操作系统的设备管理功能负责外部设备的分配、启动和故障处理,用户不必详细了解设备及接口的技术细节,就可以方便地通过操作系统提供的设备管理手段,对设备进行操作。

在操作系统中,为了提高设备的使用效率和整个系统的运行速度,需要采用一系列的技术,包括中断技术、通道技术、虚拟设备技术和缓冲技术等,尽可能发挥设备和主机的并行工作能力。此外,设备管理应为用户提供一个良好的界面,使用户不必涉及具体设备的物理特性即可方便灵活地使用这些设备。

除此之外,操作系统还要具备中断处理、错误处理等功能。操作系统的各功能之间并非是完全独立的,它们之间存在着相互依赖的关系。

5. 用户接口

除了上述四项功能之外,操作系统还应该向用户提供使用系统本身的手段,这就是用户与计算机系统之间的接口。

从用户的观点来看,操作系统是用户与计算机系统之间的接口。因此,接口管理的任务是为用户提供一个使用系统的良好环境,使用户能有效地组织自己的工作流,并使整个系统能高效地运行。

1.2 操作系统的发展

操作系统如同任何其他事物一样,也有它的诞生、成长和发展的过程。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。了解操作系统发展的历史,有助于更深刻地认识操作系统基本概念的内在含义。下面来看一看操作系统的发展过程。

1.2.1 手工操作

由于第二次世界大战对武器装备设计的需要,美国、英国和德国等国家陆续开始使用真空管建造数字电子计算机。在这个阶段,通过在一些插板上的硬连线来控制计算机的基本功能,程序设计全部采用机器语言,没有程序设计语言(甚至没有汇编语言),更谈不上操作系统。这时实