

21世纪高等教育计算机规划教材

COMPUTER

计算机操作系统

Operating System

冯裕忠 方智 周舸 主编

全面了解计算机操作系统知识

以 UNIX 操作系统为应用案例

介绍云计算与操作系统相关内容



人民邮电出版社
POSTS & TELECOM PRESS



21世纪高等教育计算机规划教材

COMPUTER

计算机操作系统

Operating System

冯裕忠 方智 周舸 主编



人民邮电出版社

北京

图书在版编目 (C I P) 数据

计算机操作系统 / 冯裕忠, 方智, 周舸主编. -- 北京: 人民邮电出版社, 2013.9
21世纪高等教育计算机规划教材
ISBN 978-7-115-32712-3

I. ①计… II. ①冯… ②方… ③周… III. ①操作系统—高等学校—教材 IV. ①TP316

中国版本图书馆CIP数据核字(2013)第185532号

内 容 提 要

本书详细介绍了计算机系统的重要组成部分——操作系统。全书共分10章。第1章介绍操作系统的发展过程、基本特征、功能、结构和分类。第2章和第3章阐述进程和线程的同步、通信、调度和死锁等基本概念。第4章讲解存储管理方式和虚拟存储器的概念。第5章到第7章讲述设备和文件的管理及用户接口。第8章详细讲述UNIX操作系统的基本组成、特点和常用命令等。第9章介绍计算机安全方面的内容。第10章介绍云计算的有关知识。本书在附录中还为读者提供了详实的参考内容。

本书可作为高等学校计算机工程和应用专业的教材,也非常适合作为IT类相关专业的教材或参考书。

-
- ◆ 主 编 冯裕忠 方 智 周 舸
责任编辑 刘 博
责任印制 彭志环 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
大厂聚鑫印刷有限责任公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 16.75 2013年9月第1版
字数: 460千字 2013年9月河北第1次印刷
-

定价: 39.00元

读者服务热线: (010)67170985 印装质量热线: (010)67129223
反盗版热线: (010)67171154

前言

在计算机系统中,操作系统是最核心、最基础的计算机系统软件,也是计算机系统资源的管理者。计算机操作系统的设计原理与实现技术是计算机专业人员有必要掌握的基本知识。当前最为流行、应用最为广泛的计算机操作系统是 Windows 和 UNIX 及各兼容版本(如 Linux)。前者是单用户/多任务/分时操作系统,主要用于个人计算机;后者是多用户/多任务/分时操作系统,主要运行在大、中、小型计算机的硬件环境中,为银行、证券、票务和统计等业务的提供处理平台。

本书共分 10 章。第 1 章旨在让读者全面了解操作系统的发展过程、基本特征、功能、结构和分类等基本知识,为此结合具体的操作系统进行了介绍。第 2 章~第 6 章是本书的重中之重,讲述了应用程序是怎样在操作系统的控制和管理下完成运行的。第 7 章阐述了用户多种接口的应用方式。第 8 章介绍了 UNIX 操作系统的基本组成、特点和系统调用等内容。第 9 介绍了涉及计算机安全方面的内容,读者可以根据自己的情况选择第 7 章~第 9 章中的有关内容。第 10 章对云计算作了介绍,读者可以从中了解到云计算与操作系统相关的内容。

为了让读者能对操作系统的主要内容有一个从理性到感性的认识,本教材在附录中给出了进程和存储管理的实验内容,通过实验来加深对操作系统中重要内容的理解和掌握。

本书课堂授课建议为 64 学时,实验 8 学时(课外 16 学时)。读者可以根据授课对象来选择、安排所需的内容和课时。

冯裕忠提出了本书的编写大纲,编写了第 1 章~第 3 章、第 7 章~第 10 章和附录 1~附录 4 的内容;方智编写了第 4 章和附录 5(实验内容并上机验证);周舸编写了第 5 章和第 6 章。

在本教材的编写过程中,得到了本校计算机系的同事和同行的大力支持,在此我们表示真诚的感谢。

由于时间较紧,作者水平有限,书中难免出现错误和疏漏,敬请广大读者批评指正。

作 者
2013 年 6 月

目 录

第 1 章 操作系统概述.....1

1.1 操作系统的定义.....2
1.2 操作系统的功能及服务对象.....3
1.2.1 操作系统的主要功能.....3
1.2.2 操作系统的服务对象.....7
1.3 操作系统的结构.....8
1.3.1 操作系统的层次结构.....8
1.3.2 操作系统的模块结构.....9
1.3.3 操作系统的微内核结构.....10
1.4 操作系统的发展过程.....11
1.5 操作系统的分类.....13
1.5.1 单道批处理系统.....13
1.5.2 多道批处理系统.....14
1.5.3 分时操作系统.....15
1.5.4 实时操作系统.....16
1.5.5 网络操作系统.....17
1.5.6 分布式操作系统.....19
1.5.7 嵌入式操作系统.....21
1.5.8 操作系统的基本特性.....22
习 题.....24

第 2 章 进程和线程.....26

2.1 进程的基本概念.....26
2.1.1 程序的顺序执行及其特征.....26
2.1.2 前趋图.....27
2.1.3 程序并发执行和特征.....28
2.1.4 进程的定义、特征与状态.....28
2.2 进程的管理.....32
2.2.1 进程控制块 PCB 简介.....33
2.2.2 进程的调度.....37
2.2.3 进程的同步与互斥.....40
2.2.4 进程的通信.....50
2.3 线程的定义.....55

2.3.1 线程的引入.....55
2.3.2 线程的属性.....56
2.3.3 什么是“超线程”.....58
习 题.....61

第 3 章 处理机的调度与死锁.....62

3.1 处理机调度的基本概念及设计原则.....62
3.2 常用的几种调度方式.....63
3.2.1 高级调度.....63
3.2.2 中级调度.....63
3.2.3 低级调度.....64
3.2.4 进程调度的功能.....64
3.2.5 引起进程调度的原因.....65
3.3 几种常用的调度队列模型.....65
3.3.1 仅有低级调度的调度队列模型.....65
3.3.2 具有高级和低级 调度的调度队列模型.....66
3.3.3 同时具有三级调度的 调度队列模型.....67
3.4 调度算法的若干准则.....68
3.5 常用的调度算法.....69
3.5.1 先来先服务调度算法.....69
3.5.2 短作业（进程）优先调度算法.....70
3.5.3 高优先权优先调度算法.....70
3.5.4 基于时间片的轮转调度算法.....72
3.6 多处理机的调度.....73
3.6.1 多处理机的类型.....73
3.6.2 多处理机系统中的进程分配方式.....74
3.6.3 多处理机系统中的进程（线程） 调度方式.....74
3.7 产生死锁的原因与必要条件.....77
3.7.1 产生死锁的原因.....77
3.7.2 产生死锁的必要条件.....78
3.7.3 预防死锁的基本方法.....78
3.7.4 系统运行的安全状态.....79

3.7.5 利用银行家算法避免死锁	80	5.1.1 设备的分类	114
习 题	84	5.1.2 设备管理的任务和功能	115
第 4 章 存储管理	85	5.2 设备控制器	115
4.1 存储器的组成	85	5.2.1 设备控制器的功能	116
4.2 存储管理的功能	86	5.2.2 设备控制器的组成	116
4.2.1 内存分配与回收	86	5.3 输入、输出的控制方式	117
4.2.2 内存共享保护	87	5.3.1 程序直接控制方式	117
4.2.3 地址映射	88	5.3.2 中断控制方式	118
4.2.4 内存扩充	88	5.3.3 DMA 控制方式	118
4.3 什么是重定位	88	5.4 中断技术	120
4.3.1 名字空间	88	5.4.1 中断的基本概念	120
4.3.2 逻辑地址空间	88	5.4.2 中断的分类和优先级	120
4.3.3 内存存储空间	89	5.4.3 中断处理的过程	121
4.3.4 地址重定位	90	5.5 缓冲技术	121
4.4 存储空间的划分	92	5.5.1 缓冲的引入	121
4.4.1 单一连续分配	92	5.5.2 缓冲的种类	122
4.4.2 固定分区分配	92	5.6 设备分配技术	125
4.4.3 动态分区分配	94	5.6.1 设备分配中的数据表	125
4.4.4 可重定位分区分配	96	5.6.2 设备分配策略	126
4.4.5 分区保护	97	5.6.3 设备分配程序	127
4.5 覆盖技术	98	5.7 SPOOLing 技术	128
4.6 交换技术	98	5.7.1 什么是 SPOOLing	128
4.7 分页存储管理	99	5.7.2 SPOOLing 系统的组成	128
4.7.1 页表	99	5.7.3 SPOOLing 系统的特点	129
4.7.2 分页的地址变换	100	习 题	129
4.7.3 两级页表及多级页表	102	第 6 章 文件管理	131
4.8 分段存储管理	103	6.1 文件系统概述	131
4.8.1 段表	103	6.1.1 文件和文件系统	131
4.8.2 分段的地址变换	104	6.1.2 文件的类型和属性	132
4.9 段页式存储管理	105	6.1.3 文件系统的基本功能	133
4.10 虚拟存储管理	106	6.2 文件结构与存储设备	133
4.10.1 什么是虚拟存储	106	6.2.1 文件的逻辑结构	133
4.10.2 请求分页存储管理	107	6.2.2 文件的物理结构	134
4.10.3 页面置换算法	109	6.2.3 文件的存取方式	136
4.10.4 请求式分段存储管理	111	6.2.4 文件的存储设备	137
习 题	112	6.3 文件存储空间的管理	138
第 5 章 输入、输出设备管理	114	6.3.1 外存的主要技术参数	138
5.1 设备管理概述	114	6.3.2 空闲块的管理	138
		6.3.3 空闲块的分配策略	140

6.4 文件目录的管理.....140	8.3 进程的描述和控制.....179
6.4.1 文件目录.....140	8.3.1 进程的PCB.....179
6.4.2 单级目录结构.....141	8.3.2 进程的状态与进程映射.....181
6.4.3 二级目录结构.....141	8.3.3 进程控制.....182
6.4.4 多级目录结构.....142	8.3.4 进程调度与切换.....184
6.5 文件的共享与保护.....143	8.4 进程同步与通信.....186
6.5.1 文件的共享.....144	8.4.1 sleep与wakeup同步机制.....186
6.5.2 文件的保护.....145	8.4.2 信号机制.....187
习 题.....146	8.4.3 管道机制.....187
第7章 操作系统接口.....148	8.5 存储器管理.....188
7.1 命令接口.....148	8.5.1 请求调页管理的数据结构.....188
7.1.1 联机命令的类型.....148	8.5.2 换页进程.....190
7.1.2 键盘终端处理程序.....149	8.5.3 请求调页.....190
7.1.3 命令解释程序.....150	8.6 Linux操作系统简述.....190
7.2 程序接口.....151	8.6.1 Linux内核.....191
7.2.1 系统调用的类型.....152	8.6.2 Linux内核定义的常量.....194
7.2.2 系统调用的执行步骤.....153	8.6.3 Linux编程.....196
7.2.3 主程序被中断时的环境保护.....155	习 题.....206
7.3 图形用户接口.....156	第9章 计算机系统安全.....207
习 题.....158	9.1 计算机系统安全的基本概念.....207
第8章 UNIX操作系统.....159	9.2 计算机系统安全的内容和性质.....207
8.1 UNIX操作系统概述.....159	9.3 系统安全的评价准则.....208
8.1.1 UNIX操作系统的发展过程.....159	9.4 现代数据加密技术.....209
8.1.2 UNIX操作系统的主要版本.....161	9.5 信息的认证技术.....211
8.1.3 UNIX操作系统的特征.....161	9.6 信息的访问技术.....212
8.1.4 UNIX操作系统的结构.....162	9.7 防火墙.....212
8.1.5 UNIX操作系统的启动流程.....163	9.7.1 包过滤防火墙.....212
8.1.6 UNIX操作系统用户和职责的 划分.....167	9.7.2 代理服务器.....213
8.1.7 UNIX操作系统的运行示意图.....168	9.7.3 规则检查防火墙.....214
8.1.8 UNIX操作系统用户的登录与 退出.....168	9.8 Windows操作系统的安全隐患.....214
8.2 UNIX操作系统的文件系统和文件.....170	9.9 DES简介.....214
8.2.1 磁盘组织.....170	习 题.....217
8.2.2 文件系统.....170	第10章 云计算.....218
8.2.3 UNIX操作系统的文件类型.....172	10.1 云计算的概念.....218
8.2.4 文件名和路径名.....177	10.2 云计算的关键技术.....222
8.2.5 文件和目录的层次结构.....178	10.3 云计算安全管理平台的主要功能.....224
	10.4 云计算应用举例.....227
	习 题.....229

附录 1	DOS/UNIX 操作系统的	附录 3	UNIX/Linux 系统下网卡等
	常用命令230		外设的安装 234
附录 2	计算机系统和网络	附录 4	计算机术语的解释..... 241
	中的常用标准231	附录 5	操作系统实验指导书..... 249

第 1 章

操作系统概述

在当今信息社会时代，计算机的发展和应用给人们的生活、学习和工作带来了举足轻重的作用。人们平常所说的计算机，严格地讲应该是指计算机系统。它包含两部分：一部分指计算机硬件（摸得着、看得见的物理部件，例如主板、CPU、内存、显示器、打印机和各种接口等）；另一部分指计算机软件（就是在硬件中运行的各种程序），又分为系统软件和应用软件。系统软件是管理和控制计算机系统运行的核心，通常把这部分称为计算机操作系统，例如我们平时所用到的 Windows、UNIX/Linux 系统软件等。Windows 通常是运行在个人计算机上的单用户、多任务、分时操作系统；而 UNIX 是在大中型计算机（由于现在个人计算机的硬件环境非常完善且功能强大，也可以在个人计算机上安装 UNIX 操作系统）系统中使用的多用户、多任务、分时操作系统。通常，系统软件都是由专业的软件开发公司研发的。

应用软件就是人们平时用于管理和完成各种业务的程序（如办公软件 Office、银行业务管理程序和各种游戏程序等）。用户可以自己开发应用软件，也可以从软件市场上购买所需的软件。

用专业术语讲，通常把没有配置系统软件的计算机（即硬件）称为裸机（这个时间段基本上指计算机问世以来到 20 世纪 50 年代），它的功能非常弱，仅能完成简单的运算（只能进行 0 和 1 的二进制运算，计算机所运行的程序是用 0 和 1 编写的，这些程序非常繁杂，不容易看懂又容易出错，只有编写人员和运行此程序的计算机可以识别，故将这种编写程序的语言称为低级语言或机器语言）。把计算机操作系统加载到计算机的硬件上（这个时间段是指 20 世纪 70 年代到现在），就可以使计算机硬件的功能变得强大、服务质量高、使用方便，为使用计算机的人们提供一个非常安全可靠的应用环境来满足各类用户的应用需求。同时，操作系统可以有效而合理地组织和安排多个用户共享计算机系统的各类设施（通常把这些设施称为资源），最大限度地提高资源的利用率。

本章将分别介绍计算机的发展过程，操作系统在计算机系统中的地位与作用，让读者对计算机操作系统有一个总的认识 and 了解。图 1.1 给出了操作系统课程在计算机专业中的位置。

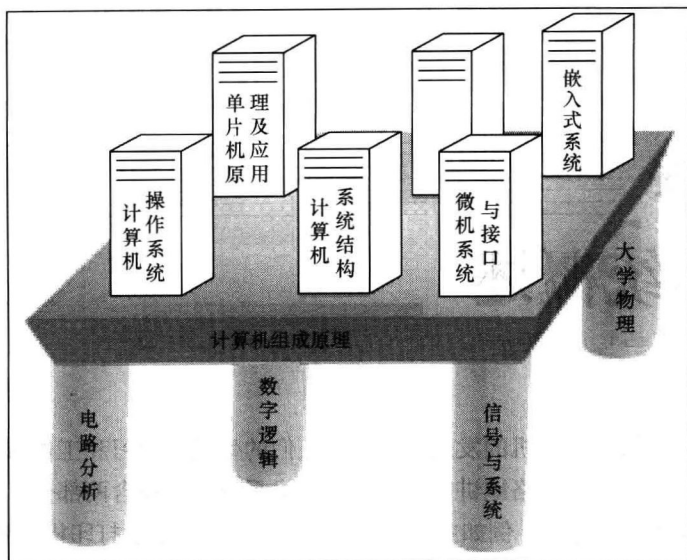


图 1.1 计算机操作系统课程在计算机专业中的位置

1.1 操作系统的定义

众所周知，根据冯·诺依曼的指导思想，计算机硬件系统是由运算器、控制器、存储器、输入/输出等安装在计算机主板上的部件，通过逻辑连接而构成的，如图 1.2 所示。随着计算机技术的应用和发展，在运算器部分又增加了若干用于存放临时信息的寄存器等部件，即 CPU。

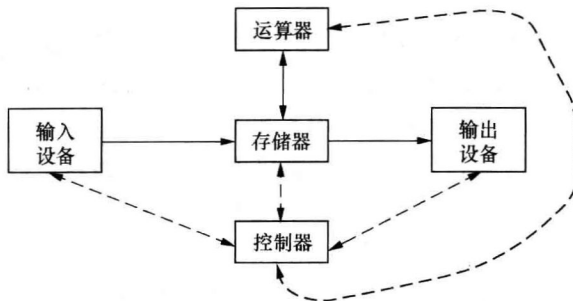


图 1.2 计算机硬件的基本组成



图 1.2 给出了数据的处理过程。当信息经过输入设备到达存储器中后，再送到运算器运算，并将运算结果返回存储器，再由存储器经输出设备输出。图中的“—”代表所传输的是数据；“---->”代表传输的是控制信号。控制器将控制信号送到被控设备，而被控设备又将其状态信息反馈给控制器。

要使这些部件能够充分发挥其效能，尽可能地按人们预期的目的和要求来运行各类程序，就需要有一套管理（控制、分配）硬件和组织程序有序运行的程序。这套程序就是人们通常所称的操作系统。这非常类似于人们修高速公路，为了提高运输和通行能力，就必须有一套管理高速公路的规章制度，在高速公路行驶的车辆必须严格执行规章制度，只有这样才能充分发挥高速公路的作用。当然计算机操作系统要比一套高速公路的管理制度复杂得多。操作系统是建立在计算机

硬件的基础上的。

有的书是这样描述：操作系统是一个用来和硬件打交道并为用户程序提供一个有限服务集的低级支撑软件。一个计算机系统是一个硬件和软件的共生体，它们互相依赖，不可分割。计算机的硬件，包括外围设备、处理器、内存、硬盘和其他电子设备，这些组成了计算机的发动机。但是没有软件来操作和控制它，硬件自身是不能工作的，完成这个控制工作的软件就称为操作系统，在 Linux 的术语中被称为“内核”，也可以称为“核心”。Linux 内核的主要模块（或组件）分为以下几个部分：存储管理、CPU 和进程管理、文件系统、设备管理和驱动、网络通信以及系统的初始化（引导）、系统调用等。

通常，操作系统是由许多可供用户调用的程序（也称为系统调用、命令）组成的。这些程序可分为三类。

1. 信息管理

信息管理（Information Management, IM），主要提供对信息的存储、检索、修改和删除等功能，通常涉及以下几个方面：

- （1）创建文件；
- （2）创建目录；
- （3）操作文件（打开/关闭、读/写文件）；
- （4）创建链表/管道。

2. 进程管理

进程管理（Process Management, PM），主要为程序的执行而创建进程、调度进程、挂起进程、终止进程和重启进程等。

进程通常是指在计算机系统中正在运行或欲运行的某一程序的一部分。

3. 内存管理

内存管理（Memory Management, MM），主要为进程分配所需的内存和回收进程运行结束后所释放的内存。

1.2 操作系统的功能及服务对象

在计算机系统上配置操作系统，就是让硬件的功能得到成百上千倍的提升。配置（安装）什么样的操作系统，与计算机硬件的规模、用户使用计算机的环境和用途密不可分。平时人们在办公室或家中单独使用的是个人计算机系统，安装小规模、单用户、多任务的分时操作系统即可，如 DOS、Windows、Mac 等操作系统；通常银行、证券、交通运输行业中的信息处理对系统要求较高，就需要安装大规模、多用户、多任务的分时操作系统，如 UNIX/Linux 操作系统。

1.2.1 操作系统的主要功能

操作系统的主要功能包括：处理机管理、存储器管理、输入/输出设备管理和文件管理。

1. 处理机管理

在传统的多程序系统中，处理机的分配和执行都以进程为基本单位，所以对处理机的管理可视为对进程的管理；在引入了线程的操作系统中，除包括了对线程的管理外，还实现对存储器、I/O 设备及文件的管理。

处理机管理的主要功能包括：创建和撤销进程（线程），对诸进程（线程）的运行进行控制，实现进程（线程）之间的信息交换，按照一定的算法条件把处理机分配给进程（线程）。当进程（线程）使用完 CPU 后，操作系统又要将 CPU 回收（也称为资源回收），再按照分配策略将 CPU 分配给需要使用 CPU 的用户（有的书中把进程（线程）称为用户）。这样，周而复始地进行下去，直至整个系统中的各类用户任务得以完成。

处理机管理主要涉及如下四个方面。

（1）进程控制。在传统的多道程序环境下，要使作业（程序）能够运行，就必须先为作业创建一个或数个进程，并为该进程分配必要的资源。多进程运行结束时，系统会立即撤销该进程，以便能及时回收该进程所占用的各类资源。进程控制的主要功能就是为作业创建进程、撤销已经结束的进程和控制进程在运行过程中的状态转换。在现代的操作系统中，进程控制还应该具有为一个进程创建若干个线程的功能和撤销（终止）已经完成任务的线程的功能。

（2）进程同步。进程是以异步方式运行的，并以人们不可想象的速度向前推进。为了使多个进程能有条不紊地运行，操作系统必须设置进程同步机制。进程同步的主要任务是为多个进程（含线程）的运行进行协调，通常有两种协调方式：①进程互斥方式，就是诸进程（线程）在对临界资源进行访问时所采用的互斥方式；②进程同步方式，就是在相互合作完成共同任务的诸进程（线程）之间，由同步机制对各进程（线程）的执行顺序进行协调的方式。

在进程同步机制中，通过对临界资源设置一个控制变量 W （锁），当 $W=1$ 时，即认为锁打开，进程就可以访问该临界资源；当 $W=0$ 时，即认为锁关上，则禁止进程（线程）对该临界资源的访问。通常把这种控制方式称为“信号量”机制，这种控制方式就是通过信号量机制实现的。

（3）进程通信。在传统的多道程序系统中，为了加速应用程序的执行，不仅应该在系统中建立多个进程，还应该为每个进程建立若干个线程，由这些进程（线程）相互合作去完成一个共同的任务。当然，在这些进程（线程）之间还需要交换信息，例如，当一个有输入进程、计算进程和打印进程的程序在系统中运行时，输入进程把输入的数据传给计算进程，计算进程运行后把结果传给打印进程，打印进程再把计算结果输出给用户。这种信息传递方式就是所谓的进程通信。

（4）进程调度。在后备队列上等待的每个作业都需要经过调度才能运行。在传统的操作系统中，调度分为作业调度和进程调度两步：①作业调度的主要任务是从后备队列中按照一定的算法，选择若干作业，为这些被选择的作业分配必要的资源（首先分配一定内存空间），再将这些被选作业调入内存、为创建进程，并按照一定的算法把这些进程放入就绪队列；②进程调度的主要任务是按照一定的算法从就绪队列中选择一个满足运行条件的进程，把处理机（CPU）分配给该进程让其执行。

在多线程的操作系统中，通常是把线程作为一个独立运行和分配处理机的基本单位，这样，就需要把就绪线程排成一个队列，每次调度时，就从就绪线程队列中选择一个线程，把处理机（CPU）分配给它，让它执行。

2. 存储器管理

存储器管理（包括内存和外存，这里主要是指内存），为需要运行的程序（任务）按照一定的分配算法为其分配所需的存储空间。当使用完后，操作系统又负责将这部分存储空间回收。

存储器管理通常有以下主要功能。

（1）内存分配。主要是为每个要运行的程序分配必要的内存空间，可采取静态分配、动态分配两种方式。静态分配就是在作业装入内存时就决定其所需内存空间的大小，在整个运行过程中不再允许该作业申请新的内存空间，也不允许其在内存中“移动”。动态分配就是在作业装入内存

后,允许该作业在运行时可以根据需要申请新的内存空间,并允许该作业在内存中“移动”。

要实现内存的分配,操作系统的内存管理机制中应该有这样的结构和功能:①内存分配数据结构(表格),该表格记录内存空间的使用情况(包括内存的剩余空间容量、地址等),作为内存分配的依据;②内存分配功能,操作系统按照一定的内存分配算法,为需要内存空间的用户分配必要容量的内存空间;③内存回收功能,即当用户不需要内存时,操作系统的内存管理把此部分内存容量回收,并修改内存分配数据结构中的相关信息。

(2) 内存保护。就是确保每个用户程序仅在属于自己的内存空间中运行,彼此互不干扰。为了实现这一功能,操作系统的内存管理模块中必须设置内存保护机制。通常,一种简单而实用的内存保护机制就是设置两个界限寄存器,分别用于存放正在执行的程序的上界地址和下界地址。当CPU从内存取指令时,首先对要取的指令的地址进行上、下界地址检查,看是否有地址越界,如有地址越界,系统则发出地址越界中断请求,以停止该指令(程序)的执行。

(3) 地址映射。在计算机系统中,一个应用程序(通常是源程序)经过编译后,会形成若干个目标地址(这时的程序就称为目标程序),再经过链接后就形成了可装入程序(也称为可执行程序)。而这些程序的地址都是从“0”开始的,程序中的其他地址都是相对于起始地址计算的;由这些地址所形成的空间称为“地址空间”,其中的地址称为“逻辑地址”或“相对地址”。将地址空间的逻辑地址转换为内存空间中与之对应的地址,则该地址称为物理地址(有的书把物理地址称为实地地址、有效地址)。

计算机系统中往往运行着多道程序,而每个程序不可能都从内存的“0”地址开始存储(装入),由于程序的逻辑地址与其物理地址不一致,为了克服这个问题,操作系统的内存管理模块被赋予了可将程序的逻辑地址转换为物理地址的功能。通常把地址转换称为地址映射。

(4) 内存扩充。这里讲的内存扩充不是通过物理方法扩大计算机的内存容量,而是通过虚拟存储技术,从逻辑上扩充内存容量,让用户所感觉到的内存容量比计算机系统的实际内存容量要大若干倍。这样可以改善系统的性能,满足用户的需要,减少内存容量的投资。

为了实现内存扩充这一目标,要求系统应该具有:①请求调入功能,就是当程序要运行时,只装入其中的一部分,便开始执行,在运行过程中,若发现要继续运行的程序尚未到达内存,便向操作系统的管理程序发出请求,再由操作系统把所需的部分程序从外存(通常是硬盘)中调入内存;②置换功能,如果已经装入内存的程序无需现在运行,别的程序又急需运行却没有内存空间,操作系统的管理程序就应该把内存中暂时不能运行的程序调出到外存中,而把急需运行的程序从外存调入到内存。

由于计算机系统的运行速度非常高,程序的换进/换出基本上是在用户没有感觉的情况下完成的。

3. 输入/输出(I/O)设备管理

按照一定的分配策略,如FCFS(先来先服务策略),将计算机系统有限的可用设备(资源)分配给需要I/O设备的用户,满足用户的要求。

操作系统需要对系统中各类信息(通常也称为文件)进行管理。计算机系统的信息是五花八门的,有文字、图形、音乐等,怎样才能按照用户的要求对信息分门别类加以管理和控制、按照用户的需要检索文件、提供操作文件的各种命令,都是操作系统中完成这部分功能的软件(程序)应该具备的。

通常,I/O管理涉及三方面。

(1) 缓冲区管理。CPU的高速运行和I/O设备的低速工作之间的矛盾是自计算机问世以来就已经存在了。随着CPU的运行速度成数十倍的提高,两者之间速度不匹配的矛盾就更突出了。为

了解决此矛盾，引入了缓冲技术。也就是在主机与外部设备之间配置一定容量的数据存储空间，当主机要与外部设备交换信息时，主机先将要输出的信息传到缓冲存储空间，使得主机可以去执行别的程序，而外部设备就从缓冲区中读取信息，如果输入设备要输入信息到 CPU 进行处理，首先输入设备将信息送到输入缓冲区，当该缓冲区满后给 CPU 一个信号，CPU 就暂停其他工作，把输入缓冲区的信息取走。这样基本上做到了主机与外部设备并行工作，提高了主机的利用率。

(2) 设备分配。根据用户进程的 I/O 请求、系统现有的资源情况等，按照一定的设备分配算法为用户分配所需设备。

为了实现设备的分配，系统中还需设置设备控制表、控制器控制表等数据表格，用于登记设备、控制器的标识符和状态等（例如某设备现在是否被占用）。

(3) 设备处理。设备处理就是指设备驱动程序，用于实现主机与外部设备之间的通信，即由主机向设备控制器（俗称为“卡”，例如网卡、声卡）发出 I/O 命令，要求其完成指定的 I/O 操作；反之由主机接收从设备控制器发来的中断请求，并给予迅速的响应和相应的处理。

处理过程是：设备驱动程序首先检查 I/O 请求的合法性，了解设备状态是否为“忙”、了解有关的传递参数（例如输入/输出的格式）和设备的工作方式（例如是串行还是并行），然后向设备控制器发出 I/O 命令，启动具体的 I/O 设备去完成指定的 I/O 操作。设备驱动程序能够及时响应由控制器发来的中断请求，并根据请求的类型，调用相应的中断处理程序进行处理。

4. 文件管理

在现代计算机系统中，大量的程序和数据都是以文件的形式存储在磁盘上（绝大多数是指硬盘）供用户需要时调用的。为此，操作系统中就必须配置文件管理机制。文件管理的主要任务：对用户文件和系统文件进行管理，方便使用并保证文件的安全。

文件管理涉及对文件存储空间的管理、文件目录的管理、文件存取的管理和文件共享、保护等。

(1) 文件存储空间的管理。为了方便用户，当前需要使用的系统文件和用户文件通常都存储在可随机存取的磁盘上。

在多用户（多任务）的环境中，需要由操作系统的文件系统来完成对诸多文件及文件存储空间的统一管理。管理的主要任务就是按照一定的分配算法为每个文件分配必须的外存空间，满足快速检索和读取的需要。

为了实现存储空间的管理，系统应该具有相应的数据结构，用于记录文件存储空间的使用情况。

(2) 文件目录管理。为了便于用户方便快速地查找自己所需的文件，系统为每个文件建立一个目录项。目录项包括文件名、文件属性、文件在磁盘上的物理地址等。由若干个文件目录项组成一个目录文件。

文件目录管理的主要任务就是为每个文件建立其目录项，同时对众多的目录进行组织，以实现快速按名存取。

文件目录管理的另一个任务就是实现主目录可以嵌套子目录。

(3) 文件的存取管理和保护。根据用户的请求，系统可以从外存储器中读取信息或将信息存储在外存储器中。

① 在进行文件的读（写）时，系统先根据用户给出的文件名去检索文件目录，从中获取文件在外存储器上的地址，然后利用文件的读（写）指针对文件进行读（写）。读（写）完成便修改文件读（写）指针，为下次读（写）做好准备。

② 文件保护。文件系统应该提供文件保护功能，以防止未经核准的用户存取文件，防止冒名顶替存取文件或以不正确的方式使用文件。

(4) 文件共享。计算机中都存储器大量的文件。其中有许多文件诸如编辑程序、数据库等是可供多用户共享的。这里就涉及为共享文件提供运行的环境、共享用户的管理、共享文件的访问等。

上述功能的完成,例如如何分配 CPU、怎样在用户的指示下一步一步地完成程序的运行,是靠控制器与操作系统的配合而实现的。

这些内容将在后面的章节中分别加以阐述。

综上所述,计算机硬件安装了操作系统,就可使计算机系统具有如下特征。

1. 方便性

计算机硬件配置了操作系统后,人们使用计算机系统就更方便。如果计算机硬件不配置(安装)系统软件,计算机就会非常难以使用(例如现在的个人计算机如果不安装操作系统,恐怕连启动机器都是不可能的),这是由于计算机硬件只能识别 0 和 1 二进制代码(也称为机器代码),人们要在计算机上运行所编写的程序,该程序就必须是用机器语言编写的,这也给用户编写程序带来极大的麻烦和困难。如果计算机硬件配置了操作系统,人们就可以用操作系统所提供的各类命令来操作计算机,也可以用计算机高级语言来编写所需的各种程序,然后将所编写的源程序交给操作系统提供的编译程序编译成计算机硬件能识别的代码。这样,用户使用计算机系统就非常方便了。

2. 有效性

在没有配置操作系统或者说操作系统的功能很弱的情况下,计算机硬件中的 CPU、内存和 I/O 设备经常会处于空闲状态而得不到充分利用,造成计算机硬件大量浪费。在配置了操作系统后,CPU、I/O 设备等在操作系统的控制下得到了充分利用,有序地完成单(多)个用户多个任务的运行。在大型机上,有效性就比方便性更显突出。

3. 可扩充性

随着超大规模集成电路(VLSI)和计算机技术的应用和发展,计算机硬件和计算机的体系结构(例如原来在一台计算机中采用单 CPU,而现在的计算机中可以采用多 CPU 结构)发生了非常大的变化。配置了操作系统的计算机,可以根据用户的需要来增加硬件(如多媒体、网络等)或扩充系统的某一部分软件(游戏、数据库等)。

4. 兼容性和开放性

随着计算机系统应用环境的变化(如从单机环境到多机的网络环境),各种不同结构体系的计算机相互传递信息,这就需要计算机系统(包括硬件和操作系统)具有兼容性和开放性。

综合上述,操作系统的主要功能表现在以下几方面。

(1) 管理 CPU,就是要运行的诸多程序创建和撤销进程(线程),对诸进程(线程)的运行进行协调,实现进程(线程)间的切换,并按一定的算法把 CPU 分配给具备条件的进程(线程)。也就是实现:进程控制、进程同步、进程通信和进程调度。有关进程的概念将在第 2 章介绍。

(2) 管理内存。为用户程序分配内存,实现内存保护、地址映像和内存扩充。

(3) 管理 I/O 设备,实现对缓冲的管理、设备分配和设备处理等的管理。

(4) 管理文件,对用户文件和系统文件进行管理,就是对文件的存储空间、目录、文件读/写和文件的共享/保护的管理。

(5) 操作系统为用户使用计算机系统提供软接口,即相关命令、系统调用等,人们通过操作系统提供的命令和系统调用来完成所需要的各种控制和计算。

1.2.2 操作系统的服务对象

操作系统用于完成计算机系统硬件的管理和控制,对各类信息的编辑、运行、输入/输出等实

行控制。由于计算机系统中，有多个程序（作业）并发执行，竞相使用资源，操作系统需根据资源的状态和程序运行的优先级等条件按一定的算法将资源分配给具备的程序（进程或线程）。例如在单 CPU 系统中，一个 CPU 要满足众多用户，而任何时刻仅有一个用户可以占用 CPU 时间。究竟是哪一个用户占用 CPU 时间，是由操作系统根据用户现场情况和采用的调度算法等多方面条件而定。所以，有的书中把操作系统称为计算机系统资源的管理者。

在计算机系统中，人们是通过操作系统或者说是通过操作系统提供的各种相关命令来使用计算机的，所以说，操作系统是用户与计算机硬件的接口。

1.3 操作系统的结构

操作系统是一个十分复杂而庞大的系统软件。为了控制该软件的复杂性，可以用软件工程的概念、原理、规范，来开发、运行和维护软件，以杜绝开发软件的随意性、编程冗余和维护困难等问题。为此，人们经过长期的探索，把做工程的思路和方法等应用到了软件（尤其是系统软件）的开发过程中。下面展示较为普遍的系统层次结构和模块结构。

1.3.1 操作系统的层次结构

在层次结构中，整个操作系统的构成通常以分层的结构来实现，各个部分关系非常清晰，一目了然。通常用图 1.3 和图 1.4 来划分计算机系统的结构，按照层次结构可以非常清楚地知道操作系统在整个计算机系统的位置。

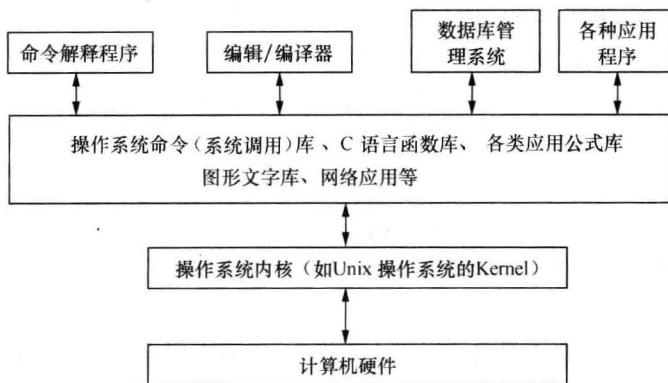


图 1.3 计算机系统中的层次结构

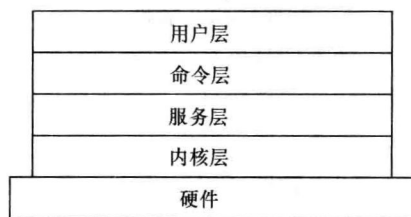


图 1.4 计算机系统的分层示意图

内核层：它是操作系统的最里层，是唯一直接与计算机硬件打交道的层。它使得操作系统和计算机硬件相互独立。也就是说，只要改变操作系统的内核层就可以使同一操作系统运行于不同的计算机硬件环境下。内核层提供了操作系统中最基本的功能，包括了装入、执行程序以及为程序分配各种硬件资源的子系统。软件与硬件所传递的各类信息在内核层进行处理，这样，对普通用户来讲，复杂的计算机系统便变得简单易操作了。

图 1.4 中间的命令层、服务层和内核层实际上就是操作系统部分。

服务层：服务层接受来自应用程序或命令层的服务请求，并将这些请求译码为传送给内核执行的指令，然后再将处理结果回送到请求服务的程序。通常，服务层是由众多程序组成，可以提供如下的服务。

- (1) 访问 I/O 设备：将数据进行输入/输出。
- (2) 访问存储设备（内存或外存）：从磁盘读或将处理后的数据写入磁盘。
- (3) 文件操作：通常指打开（关闭）文件、读写文件。
- (4) 特殊服务：窗口管理、网络通信和数据库访问等。

命令层：提供用户接口，是操作系统中唯一直接与用户（应用程序）打交道的部分（如 UNIX 操作系统的 shell）。

用户层：这里通常是指应用程序。

1.3.2 操作系统的模块结构

模块结构是指在开发软件（尤其是像计算机操作系统这样的大型软件）时，由于其功能复杂、参加开发工作的人员众多，要使每个人都能各负其责、各尽所能，有序地完成开发任务，通常会根据软件的大小、功能的强弱和参与人员等具体情况，把开发工作按功能（任务）划分若干模块，分散开发，集中组合、调式，使所开发的软件功能完善、结构优化。图 1.5 给出了操作系统的模块结构示意图。

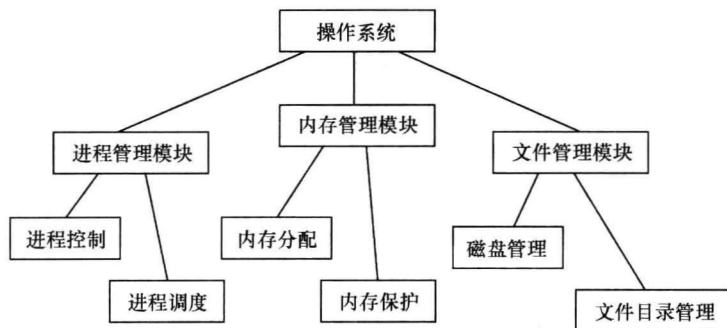


图 1.5 操作系统的模块化结构

不管是层次机构还是模块化结构的操作系统，都可以用图 1.5 来说明其构造体系。

从图 1.5 中可以看到，操作系统一般由“进程管理、内存管理、文件管理”三个子模块组成。这些子模块分别又由若干子模块组成。这样的结构类似一个倒树型，较为清晰，非常利于操作系统的修改、扩充和维护。例如，现在要增加一个 I/O 子模块，只要把 I/O 子模块连接在操作系统的主模块上就可以了，不需要修改系统的其他模块。

从图 1.6 可以看出，最底层实际上是操作系统控制和管理的计算机硬件、各类信息（文件）和需要运行的程序等部分。

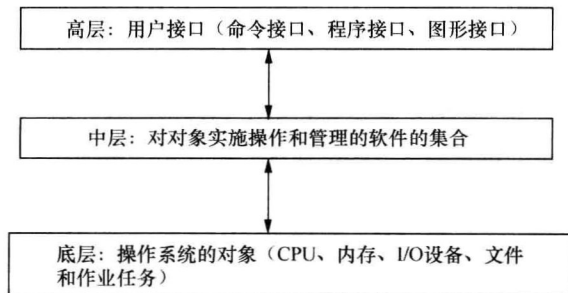


图 1.6 整个计算机系统的结构体系

中间部分（中层），是完成底层任务所需的各类程序，是一个庞大的软件体。