

高等院校“十三五”应用技能培养规划教材
• 移动应用开发系列

操作系统原理与应用 (第2版)

王育勤 主编
刘智珺 苏莹 鲁力 副主编



清华大学出版社



非
外
借

高等院校“十三五”应用技能培养规划教材·移动应用开发系列

操作系统原理与应用 (第2版)

王育勤 主 编
刘智珺 苏 莹 鲁 力 副主编

清华大学出版社
北 京

内 容 简 介

本书是作者在多年教学实践积累的基础上,吸收国内外操作系统新的理论和技术,依据操作系统教学大纲的要求进行编写的。

本书重点讲述了操作系统的一般原理和实施技术与方法。在讲授方法上,注重理论与实际相结合,以 UNIX 操作系统为实例,介绍操作系统中的主要管理方法和服务功能的实施技术和技巧。在内容安排上,由总体到具体,先介绍操作系统的概念和服务功能,然后以 UNIX 系统为例讲述这些功能的具体实现算法,最后以 Windows 和 Linux 操作系统为基础,通过 Windows 10 系统中的进程管理、内存管理、程序管理和网络管理几个方面的实例与具体操作,将操作系统原理中涉及的相关部分理论具体化,加深对操作系统理论的理解;同时介绍了 Linux 操作系统中常用命令的使用,编辑工具 vi 以及 Shell 脚本编程的基本方法;最后选取 Windows 操作系统中 5 个典型的项目实验来加强实践环节,使学生进一步加深对操作系统关键功能实现方法的理解与应用。

本书可作为高等院校计算机领域各专业、电子信息类专业本科生(若操作系统课程为 40~60 授课学时,主讲教师可根据实际情况作选学处理)和非计算机专业研究生操作系统课程的教材,也可供有关专业人员参考使用。

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:010-62782989 13701121933

图书在版编目(CIP)数据

操作系统原理与应用/王育勤主编. —2 版. —北京:清华大学出版社, 2019

(高等院校“十三五”应用技能培养规划教材·移动应用开发系列)

ISBN 978-7-302-53529-4

I. ①操… II. ①王… III. ①Windows 操作系统—高等学校—教材 IV. ①TP316.7

中国版本图书馆 CIP 数据核字(2019)第 166164 号

责任编辑:汤涌涛

装帧设计:杨玉兰

责任校对:王明明

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62791865

印装者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:17.75 字 数:430 千字

版 次:2013 年 8 月第 1 版 2019 年 8 月第 2 版 印 次:2019 年 8 月第 1 次印刷

定 价:49.00 元

产品编号:084270-01

前 言

操作系统是计算机系统的基本组成部分，它在整个计算机系统软件中占据核心地位。对操作系统的概念、理论和方法的研究，以及对操作系统的使用、分析、开发和设计，历来是计算机领域中最主要的课题和任务之一，因而，操作系统是计算机科学教育的基本课程之一。它涉及对各种资源(包括硬件和软件资源)的有效管理，又为用户及高层软件的运行提供良好的工作环境，起着承上启下、纵横贯通的作用。

本书以典型的 UNIX 操作系统为例，重点讲述操作系统的一般原理和实施技术与方法，最后在实践中，以 Windows 操作系统为基础，介绍了操作系统的安装、常用命令及其网络设置与通信等内容。

本书共分 10 章。

第 1 章介绍操作系统的发展历史、分类、功能、体系结构及 UNIX 系统的特点。

第 2 章介绍进程的基本概念、有关进程的操作、进程间的相互作用和通信及中断处理。

第 3 章介绍处理机管理，包括常用调度算法、UNIX 系统中的进程调度等。

第 4 章~第 6 章分别介绍存储管理、设备管理，以及文件系统的概念、功能及其主要实现技术。

第 7 章介绍死锁的概念和解决死锁问题的基本方法。

第 8 章~第 10 章是 16 学时的上机实践内容。其中，第 8 章介绍 Windows 系统管理，该章主要是通过对 Windows 10 操作系统的常用设置，进一步加深对操作系统中进程、内存、程序和网络相关知识的理解和认识。第 9 章介绍 Shell 和 Linux 的常用命令。第 10 章主要是进行 Windows 操作系统中的项目实训，选取了操作系统中 5 个典型的项目实验，在操作系统理论的基础上，将操作系统中进程与作业调度、动态分区存储管理、模拟页式虚拟存储管理中硬件地址转换、虚拟存储器中页面淘汰算法，以及死锁问题中银行家算法 5 个知识点，通过项目实验程序完成对实现方法的仿真，使学生进一步加深对操作系统关键功能实现方法的理解与应用。

本书每章的开头，先交代所要讨论的问题、环境和意义，然后逐层展开论述，在讲授理论的基础上，辅以 UNIX 系统的实例，从而加深对概念的理解和形象化思维。第 1 章~第 7 章后面都附有大量习题，这些有代表性的习题对学习巩固正文中的知识是有益的。本书在介绍 UNIX 系统的各功能模块的实现方法时，突出了重点和难点，并结合以往教学的实践体会，对难以理解的部分作出了较为详细的说明及生活中的实例，便于学生自学复习有关内容。本书还为学生在 Linux 和 Windows 系统环境下的上机实习、应用开发提供了指南。

本书由王育勤教授任主编，刘智珺、苏莹、鲁力任副主编。在本书编写过程中，第 1~3 章由苏莹老师编写，第 4~7 章由鲁力老师编写，第 8~10 章由刘智珺老师编写，在此表示衷心的感谢！全书由王育勤教授统稿。

由于编者水平有限，书中的疏漏之处在所难免，恳请广大读者给予指正。

编 者

目 录

第 1 章 操作系统概述1	
1.1 操作系统的发展过程.....2	
1.1.1 手工操作阶段.....2	
1.1.2 早期批处理阶段.....3	
1.1.3 执行系统.....4	
1.1.4 多道程序系统阶段.....4	
1.2 什么是操作系统(Operating System).....5	
1.2.1 概念(定义).....5	
1.2.2 设置 OS 的目的.....5	
1.2.3 操作系统的目标和作用.....6	
1.2.4 操作系统的主要功能.....7	
1.2.5 操作系统的服务功能.....8	
1.3 操作系统的结构.....10	
1.3.1 外部结构(环境).....10	
1.3.2 内部结构(体系结构).....10	
1.4 操作系统的分类.....11	
1.4.1 多道批处理系统.....11	
1.4.2 分时系统.....12	
1.4.3 实时系统.....13	
1.4.4 现代操作系统.....14	
1.5 操作系统的特征.....15	
1.6 操作系统的性能——如何评价一个操作系统.....16	
1.7 当前比较流行的几种微机操作系统.....17	
1.7.1 当前微机上的主流操作系统.....17	
1.7.2 用户如何选用操作系统.....20	
1.8 UNIX 系统的特点和结构.....22	
本章小结.....24	
习题.....24	
第 2 章 进程管理25	
2.1 进程的概念.....26	
2.1.1 程序的顺序执行.....26	
2.1.2 程序的并发执行和资源共享.....26	
2.1.3 程序并发执行的特性.....28	
2.1.4 进程(process).....29	
2.1.5 用进程概念说明操作系统的并发性和不确定性.....31	
2.1.6 进程的状态及其变迁.....32	
2.1.7 进程的组成.....33	
2.1.8 UNIX 系统的进程映像.....35	
2.2 有关进程的操作.....38	
2.2.1 进程的创建.....38	
2.2.2 进程终止和父子进程的同步...41	
2.3 进程间的相互作用和通信.....43	
2.3.1 同步.....43	
2.3.2 互斥.....44	
2.3.3 进程的临界区和临界资源.....45	
2.3.4 实施临界区互斥的锁操作法...45	
2.3.5 信号量与 P、V 操作.....48	
2.3.6 高级通信机构.....53	
2.4 线程.....55	
2.4.1 线程的概念.....55	
2.4.2 线程的特点.....56	
2.4.3 线程的状态.....57	
2.4.4 线程与进程的区别.....57	
2.4.5 多线程编程.....58	
2.5 中断处理.....59	
2.5.1 中断及其一般处理过程.....59	
2.5.2 中断优先级和多重中断.....61	
2.5.3 中断屏蔽.....61	
2.5.4 中断在操作系统中的地位.....62	
2.5.5 UNIX 系统对中断和陷入的处理.....62	
本章小结.....65	
习题.....65	
第 3 章 处理机管理69	
3.1 概述.....70	
3.1.1 CPU 调度的三级实现.....70	

3.1.2 进程的执行方式.....	72	本章小结.....	122
3.1.3 CPU调度的基本方式.....	73	习题.....	122
3.2 常用调度算法.....	74	第5章 设备管理	125
3.2.1 先来先服务FCFS.....	74	5.1 概述.....	126
3.2.2 最短周期优先SFB.....	75	5.1.1 设备分类.....	126
3.2.3 优先级.....	77	5.1.2 设备管理的目标和功能.....	126
3.2.4 轮转法.....	77	5.1.3 设备分配技术.....	127
3.2.5 可变时间片轮转法.....	79	5.1.4 通道技术.....	128
3.2.6 多队列轮转法和多级反馈 队列法.....	79	5.1.5 缓冲技术.....	130
3.3 UNIX系统中的进程调度.....	80	5.2 UNIX系统的设备管理.....	132
本章小结.....	85	5.2.1 UNIX设备管理的特点.....	132
习题.....	85	5.2.2 与设备驱动有关的接口.....	133
第4章 存储管理	87	5.2.3 块设备管理中的缓冲技术.....	133
4.1 引言.....	88	5.2.4 块设备的读、写.....	136
4.1.1 二级存储器及信息传送.....	88	5.2.5 字符设备管理.....	140
4.1.2 存储器分配.....	88	本章小结.....	141
4.1.3 存储管理的基本任务.....	89	习题.....	141
4.1.4 存储空间地址问题.....	90	第6章 文件系统	143
4.1.5 地址转换.....	90	6.1 概述.....	144
4.1.6 存储管理的功能.....	92	6.1.1 文件及其分类.....	144
4.1.7 内存的扩充技术.....	93	6.1.2 文件系统的功能.....	145
4.2 分区式管理技术.....	95	6.1.3 文件系统的用户界面.....	145
4.2.1 固定分区法.....	95	6.1.4 文件系统的层次结构.....	147
4.2.2 可变分区法.....	96	6.2 文件的组织和存取方法.....	148
4.3 可重定位分区分配.....	98	6.2.1 文件的逻辑组织和物理组织.....	148
4.4 多道程序对换技术.....	99	6.2.2 文件的存取方式.....	151
4.5 分页存储管理.....	100	6.3 目录结构.....	151
4.5.1 分页管理.....	100	6.3.1 一级目录结构.....	152
4.5.2 请求分页管理.....	103	6.3.2 二级目录结构.....	152
4.6 段式存储管理.....	111	6.3.3 多级目录结构.....	153
4.6.1 分段和分段的地址空间.....	111	6.4 文件存储空间的管理.....	155
4.6.2 分段管理的实现.....	112	6.4.1 记住空间分配现状的 数据结构.....	155
4.6.3 分段共享.....	113	6.4.2 存储空间分配程序.....	157
4.6.4 段的动态链接.....	113	6.5 文件保护.....	158
4.7 段页式存储管理.....	116	6.5.1 文件系统的完整性.....	158
4.7.1 基本思想.....	116	6.5.2 文件的共享与保护保密.....	159
4.7.2 实现过程.....	116	6.6 对文件的主要操作.....	160
4.8 UNIX系统的存储管理.....	117		

6.6.1 创建文件	160	8.1.1 实验目的	200
6.6.2 文件的连接与解除连接	161	8.1.2 实验内容	200
6.6.3 文件的打开和关闭	161	8.2 系统管理内存	202
6.6.4 文件的读、写	162	8.2.1 实验目的	202
6.7 UNIX 文件系统的内部实现	163	8.2.2 实验内容	203
6.7.1 数据结构	163	8.3 程序管理	205
6.7.2 活动 i 节点的分配与释放	166	8.3.1 实验目的	205
6.7.3 目录项和检索目录文件	167	8.3.2 实验内容	205
6.7.4 文件的索引结构	169	8.4 系统管理网络	208
6.7.5 文件卷和卷专用块	170	8.4.1 实验目的	208
6.7.6 空闲 i 节点的管理	171	8.4.2 实验内容	209
6.7.7 空闲存储块的管理	173	本章小结	219
6.7.8 子文件系统装卸和 装配块表	174	第 9 章 Linux 操作系统中的 常用命令	221
6.7.9 各主要数据结构之间的 联系	176	9.1 使用 Linux 基本命令	222
6.7.10 管道文件(pipe)	176	9.1.1 常用简单命令	222
6.8 系统调用的实施举例	180	9.1.2 目录管理命令	224
本章小结	182	9.1.3 文件管理命令	227
习题	182	9.2 使用命令补齐和别名功能	228
第 7 章 死锁	185	9.2.1 命令行自动补齐	228
7.1 死锁的基本概念	186	9.2.2 命令别名	228
7.1.1 什么是死锁	186	9.3 使用重定向和管道	230
7.1.2 死锁的表示	187	9.3.1 重定向	230
7.1.3 死锁判定法则	188	9.3.2 管道	231
7.2 死锁的预防	189	9.4 熟悉 vi 三种模式下的操作命令	231
7.3 死锁的避免	191	9.4.1 vi 的三种工作模式	232
7.3.1 资源分配状态 RAS	191	9.4.2 vi 在三种模式下的 基本操作	232
7.3.2 系统安全状态	191	9.5 使用 vi 建立简单的 Shell 脚本 并运行	234
7.3.3 死锁避免算法	192	9.5.1 创建 Shell 脚本	235
7.3.4 对单体资源类的简化算法	193	9.5.2 运行 Shell 脚本	235
7.4 死锁的检测和解除	194	9.5.3 Shell 编程基础	236
7.4.1 死锁的检测	194	9.5.4 流程控制语句	239
7.4.2 死锁的解除	195	本章小结	242
本章小结	196	第 10 章 操作系统项目实验	243
习题	196	10.1 进程调度及作业调度	244
第 8 章 操作系统基础实验	199	10.1.1 项目实验目的和要求	244
8.1 进程管理	200		



10.1.2	实验内容.....	244	10.4	虚拟页式存储器页面淘汰算法	
10.1.3	实验知识点说明.....	244		模拟.....	259
10.1.4	实验分析.....	244	10.4.1	项目实验目的和要求.....	259
10.2	动态分区存储管理.....	251	10.4.2	实验内容.....	259
10.2.1	项目实验目的和要求.....	251	10.4.3	实验知识点说明.....	259
10.2.2	实验内容.....	251	10.4.4	实验分析.....	260
10.2.3	实验知识点说明.....	251	10.5	银行家算法.....	263
10.2.4	实验分析.....	251	10.5.1	项目实验目的和要求.....	263
10.3	模拟页式虚拟存储管理中硬件的		10.5.2	实验内容.....	263
	地址转换与缺页中断.....	255	10.5.3	实验知识点说明.....	263
10.3.1	项目实验目的和要求.....	255	10.5.4	实验分析.....	264
10.3.2	实验内容.....	255		本章小结.....	273
10.3.3	实验知识点说明.....	256	参考文献.....		274
10.3.4	实验分析.....	256			

第 1 章

本章要点

1. 操作系统的发展史。
2. 操作系统的功能和特征。

学习目标

1. 理解操作系统的发展过程。
2. 掌握操作系统学科的主要任务和功能。
3. 了解操作系统课程要学习的大致内容和知识。

1.1 操作系统的发展过程

操作系统这一学科的产生和出现与其他任何新观点、新概念一样,并不是突然产生的,也不是一有计算机就有了操作系统这一学科,它也有一个发展、演变的过程。为了加深对这门课程的了解,下面先来回顾一下操作系统的各个发展阶段,即操作系统这门学科是怎么产生的。

1.1.1 手工操作阶段

早期的计算机是十分庞杂的由控制台“指挥”的机器,它使用的是一种初级的人机交互方式。即先在输入设备(纸带机、卡片机等)上,由人工把程序装入内存,然后启动、执行程序。通过控制台上的显示灯来监视程序的执行情况(如有错,则报错灯亮),并且直接由控制台对程序进行一些调试。其相互之间的控制关系如图 1-1 所示。

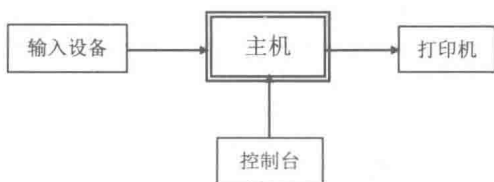


图 1-1 第一代计算机的控制关系

其中,主机包括中央处理器(CPU)和主存储器两部分。这种使用方法最明显的特点是:①资源独占(即计算机的全部硬件资源都由一个程序独自占用)。②串行工作(人的操作与计算机的运行,以及计算机各个部件之间都是按时间先后顺序工作的)。③人工干预(计算机是在人的直接联机干预下进行工作的)。

因为以上这些特点,使其存在非常严重的两个缺点:①资源浪费,②使用不便。这是为什么呢?因为资源独占和串行工作,导致一个程序在某一时刻不可能使用计算机的全部资源,它必须在程序全部装入内存之后,才能在主机上运行。所以当输入设备工作时,主机和输出设备都是空闲的,反之亦然。另一个原因,若是人工干预多的话,由于人工干预的操作速度远远低于主机的运算速度,而在人工干预的操作过程中,主机必须是停止运算状态,这不仅大大降低了计算机的使用效率,也使用户感到使用不便。随着计算机运算速度的大幅度提高,各种部件和设备日益增多,再加上计算机应用范围的普及,上述缺点就越来越突出了。

【例 1.1】 某个程序在两台具有不同运算速度的计算机甲、乙上的运行情况如表 1-1 所示。

表 1-1 不同运算速度的运行情况

计算机	运算速度	运行时间	人工操作时间	比例
甲	1000 次/S	1 小时	5 分钟	12 : 1
乙	60000 次/S	1 分钟	5 分钟	1 : 5

在计算机甲上运行这个程序时,机时的浪费还可容忍,在计算机乙上运行这个程序时,机时的浪费就不可容忍了。

为此就迫使计算机工程研究人员尽量克服这两个缺点。他们首先想到能否缩短建立作业和人工操作的时间,因而提出了从一个作业到下一个作业的自动转换方式,从而出现了早期批处理方式。

1.1.2 早期批处理阶段

早期批处理分为联机批处理和脱机批处理两种类型,一般将完成作业间自动转换工作的程序称为监督程序。

1. 早期联机批处理

在这种方式下,操作员(或用户)只需在输入设备上装入作业信息(或程序+数据+作业说明书),过一段时间后在打印机上取执行结果。其他操作都是由机器自动进行。如机器自动输入、编译和执行程序。当一个作业完成之后,由机器(监督程序)自动调入该批的第二个作业进行操作。因每次交给系统的作业是成批的,故称为批处理(就像流水线生产一样,进去一批作业,一个接一个完成后,又依次将结果打印出来)。

这种方式,比早期的人工操作要先进些,因为它多了监督程序,能实现一个作业到另一个作业的自动转换,从而缩短建立作业和人工操作时间。但这种方式还是存在着很严重的缺点:即将程序调入内存→计算结果→打印输出,都是由中央处理器(CPU)直接控制完成的。我们知道,CPU的速度较之输入(I)、输出(O)设备要快得多,那么由CPU去直接控制I/O设备,势必出现许多空等时间,就相当于将CPU的速度降之为I/O设备的速度。这种慢速外设与高速主机间串行工作的矛盾随着计算机运算速度的提高越来越突出。为了克服这一缺点,人们在批处理中引入了脱机技术(因为这一缺点从表面上看是由联机造成的,自然人们就想到了脱机),从而形成了早期脱机批处理。

2. 早期脱机批处理

1) 使用的方法

早期脱机批处理方式就是在主机之外另设一台小型机,称之为卫星机。由卫星机与外部设备打交道,而使主机腾出较多的时间来完成一些快速的业务,如图1-2所示。

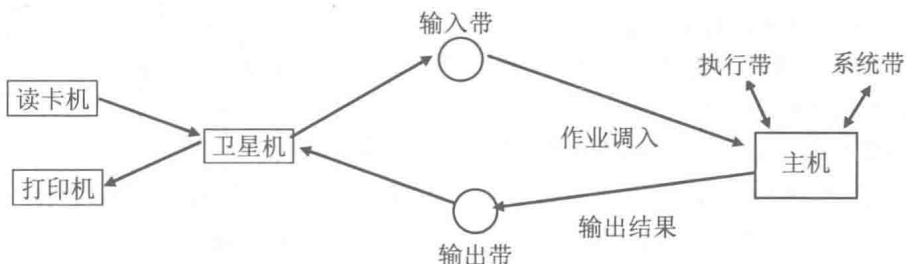


图 1-2 早期脱机批处理系统模型

2) 工作过程

读卡机上的作业通过卫星机逐个地传送到输入磁带上,而主机只负责把作业从磁带上

1.2 什么是操作系统(Operating System)

1.2.1 概念(定义)

操作系统(Operating System), 它不是讲解操作方法的, 因由名字可能会引起一些误解, 平时我们常常将操作系统简称为 OS。一个计算机系统是由硬件和软件两大部分组成。硬件通常指诸如 CPU、存储器、外设等这类用以完成计算机功能的各种部件。硬件部分是计算机系统必须具备的部分, 它是计算机赖以工作的基本部件, 不存在无硬件部分的计算机。通常将只有硬件的计算机称为“裸机”。用户直接使用裸机是非常困难而且很不方便的。有人说:“没有软件的计算机是一堆废金属”。这话虽然不准确, 但也反映了一些事实, 即裸机没有什么适用范围。必须给裸机穿上“衣服”——即编制不同的软件, 才能让它接待别人——即以更好的姿态面向用户。

计算机软件指为计算机编制的程序, 以及执行程序时所需要的数据和说明使用该程序的文档资料。因为程序是软件的核心部分, 所以人们往往在介绍软件时只讲程序, 即程序和软件是同意语。

计算机软件包括应用软件和系统软件两大部分。所谓应用软件是指针对某些特定应用领域所配置的软件。这些软件的应用范围往往要受到特定应用领域的限制(如用于计算机辅助设计的 CAD、用于企业管理等软件)。而系统软件则不然, 它是指计算机系统所必须配备的软件, 通常是在各种应用领域都可通用的软件(如编译程序、连接程序、操作系统等), 而操作系统又是这些软件中最基本的部分。

操作系统(OS)——是管理计算机系统资源(硬件和软件)的系统软件, 它为用户提供方便、有效和安全可靠的工作环境。

对操作系统, 至今尚未有严格定义, 但上面的解释, 既说明了操作系统的本质, 也说明了操作系统是干什么的, 以及它的功能。但这个解释并不全面, 下面再做几点说明。

(1) 从此定义上讲, 操作系统是软件而不是硬件。操作系统这一软件的重要任务之一是管理计算机本身的机器硬件, 因此, 在操作系统运行和实现其功能的过程中, 需要硬件强有力的支持, 而且操作系统的一部分功能就是由硬件直接完成的(如中断系统中, 有一部分功能就是由中断机构直接完成的)。从这个意义上讲, 操作系统又不完全是软件, 而是一个软、硬件结合的有机体——在软、硬件的相互配合下, 共同完成操作系统所应完成的任务。由于由硬件直接完成的功能只占很少一部分, 因此一般还是说操作系统是软件。

(2) 操作系统是系统软件而不是应用软件, 但它与其他系统软件不同。操作系统不仅与应用软件不同, 也与其他系统软件不同。一个完善的计算机系统, 通常都配有众多的系统软件, 如编辑程序、编译程序等, 所有这些程序, 虽然与操作系统一样都属于系统软件, 但它们都受 OS 的管理和控制, 并得到 OS 的支持和服务, OS 可以说是这些系统软件的领导者(控制者)。

1.2.2 设置 OS 的目的

具有一定规模的现代计算机系统一般都配备有一个或几个 OS, 而且 OS 的性能在很大

程度上决定了计算机系统工作的优劣。那么在计算机系统中,设置 OS 的目的是什么呢?主要有两个目的。

1. 方便用户(即为用户创造良好的工作环境)

因为用户直接使用裸机非常困难且不方便,设想一下:如果你所用的机器上没有装入 OS,无法使用命令,不能利用应用程序,连设备、内存等都需要自己去亲自管理,这显然会给你带来一些非常烦琐的程序设计工作,而这对于多数用户来说往往是无法胜任的,因为他必须掌握非常全面的计算机知识,而且具备很强的编程能力才行。但若在裸机的基础上设置了 OS,用户就可以用相当简便的方式,在 OS 的帮助下进行 I/O 操作,摆脱了烦琐的程序设计工作,所以从用户角度看,OS 是用户和裸机之间的一个界面。用户通过这一界面能方便地使用本来很难使用的计算机,也就是说 OS 向用户提供了一个方便而且强有力的使用环境。

2. 充分发挥计算机中各种资源的效率

这是从另一个观点(即资源管理观点)来看待 OS。因为 OS 是管理计算机系统中各种资源的软件,如果把一个计算机系统比作一个“国家”,则 OS 可以说是这个国家的“政府机构”。因现代计算机系统通常都是多道程序系统,所以 OS 就必须在多道程序之间合理地分配和回收各种资源,使资源得到合理有效的使用,使程序得以有条不紊地运行。

1.2.3 操作系统的目标和作用

1. 操作系统的设计目标

操作系统既要管理资源,又要为用户服务,所以,系统资源管理和提供用户界面是操作系统的功能要点。在资源管理中,操作系统的任务是使各种系统资源(硬件和软件资源)得到充分、合理的使用,解决用户作业因争夺资源而产生的矛盾。操作系统资源管理程序的设计目标有以下几方面。

1) 监视资源

操作系统作为用户作业的宏观调控者,必须时刻保持系统资源分配的全局信息,了解系统资源的总数、已分配和未分配的资源情况、资源的增减和变动情况、每类资源所具有的特点和适应性。这些资源信息是通过操作系统中各类数据结构和表格记录下来的,并且在系统运行过程中会不断更新。

2) 分配资源

操作系统必须对来自用户和应用程序的资源使用请求作出快速的响应,适当地处理这些请求,并且调解请求中的冲突,确定资源分配策略。当多个进程或多个用户竞争某个资源时,操作系统必须进行裁决,根据资源分配的条件、原则和环境,确定是否立即分配还是暂缓分配。对可以分配的资源,记录相应的分配情况,更新相应的分配记录。

3) 回收资源

当用户使用资源结束,提出释放请求时,操作系统按照与分配过程相反的操作回收用过的资源,同时更新相应的分配记录。

2. 操作系统的目标

(1) 有效性。包含两个方面的含义：①提高系统资源利用率，使 CPU 与 I/O 设备保持忙碌状态而得到有效利用，使内存与外存中的数据因有序而节省空间。②提高系统吞吐量。合理组织计算机工作流程，改善资源利用率，加速程序运行，缩短程序运行周期。

(2) 方便性：使计算机更容易使用。硬件只识别 0 与 1，那么如果没有 OS，用户要使用计算机就需要使用单纯的 0 与 1 字符串来操作机器。有了 OS，用户可以直接使用 OS 提供的各种命令来操作机器。

(3) 可扩充性：要能适应计算机硬件、网络、体系结构与应用发展的要求，保持对上接口可扩充；应采用层次化结构，能方便对 OS 进行扩充。比如现在采用的微内核结构与客户服务器模式。

(4) 开放性：系统能遵循世界标准规范，遵循开放系统互连(OSI)国际标准。

随着计算机技术的发展，在最开始的时候有效性最重要，但后来方便性更加重要。现在，可扩充性与开放性也是必须要考虑的目标。

3. 操作系统的作用

操作系统位于底层硬件与用户之间，是两者沟通的桥梁。用户可以通过操作系统的用户界面，输入命令。操作系统则对命令进行解释，驱动硬件设备，实现用户要求。所以它是用户和计算机之间的界面。一方面操作系统管理着计算机上的所有系统资源，另一方面操作系统为用户提供了一个抽象概念上的计算机。在操作系统的帮助下，用户使用计算机时，避免了对计算机系统硬件的直接操作。对计算机系统而言，操作系统是对所有系统资源进行管理的程序集合；对用户而言，操作系统提供了对系统资源进行有效利用的简单抽象的方法。安装了操作系统的计算机称为虚拟机(Virtual machine)，是对裸机的扩展。实际上，用户是不用接触操作系统的，操作系统管理着计算机硬件资源，同时按照应用程序的资源请求，分配资源，如划分 CPU 时间，内存空间的开辟，调用打印机等。

1.2.4 操作系统的主要功能

现代计算机系统中的重要资源包括硬件资源、软件资源与用户资源。在这些资源中，最重要的是与程序运行、数据处理、用户操作密切相关的资源，通常包括中央处理器(CPU)、主存储器、输入输出设备、数据与信息、交互环境及互连通信 5 部分。所以，常规操作系统的主要任务针对这 5 个部分，对应地有如下 5 类功能模块。此外随着计算机软硬件技术的发展和应用领域的需求，现代操作系统还对网络与通信资源、安全机构与设施资源、多媒体资源等的管理进行了新的功能扩充。

1) 处理器(处理机、CPU)管理

由于计算机系统的“心脏”是处理器，所有软硬件操作都由处理器分解执行。在单处理器的计算机系统中，存在着用户作业争用处理器的情况。如何对使用处理器的请求作出适当的分配，就是操作系统处理器管理功能模块要解决的问题。在实际工作中，操作系统将以进程和作业的方式进行管理，完成作业和进程的派遣和调度，分配处理机时间，控制作业和进程的执行。

2) 存储器管理

在计算机系统中,存储器(一般称为主存或内存)是程序运行、中间数据和系统数据存放的地方,由于硬件的限制,它们的存储容量是有限的。此外,如果有多个用户共享存储器,彼此之间不能相互冲突和干扰。操作系统的存储器管理模块就是对用户作业进行分配并回收存储空间,进行存储空间的优化管理。

3) 设备管理

设备管理是指计算机系统中除了CPU和主存以外的所有输入、输出设备的管理。除了进行实际I/O操作的设备外,还包括诸如设备控制器、DMA控制器、通道等支持设备。外围设备的种类繁多、功能差异很大。这样,设备管理的首要任务是为这些设备提供驱动程序或控制程序,使用户不必详细了解设备及接口的技术细节,就可方便地对这些设备进行操作。另一个任务就是利用中断技术、DMA技术和通道技术,使外围设备尽可能地与CPU并行工作,以提高设备的使用效率,提高整个系统的运行速度。

4) 文件管理

程序和数据是以文件形式存放在外存储器(如磁盘、磁带、光盘)上的,需要时再把它们装入内存。文件包括的范围很广,例如用户作业、源程序、目标程序、初始数据、结果数据等,而且各种系统软件甚至操作系统本身也是文件。因此,文件是计算机系统中除CPU、内存、外设以外的另一类资源,即软件资源。有效地组织、存储、保护文件,使用户方便、安全地访问它们,是操作系统文件管理的任务。

对上述4种资源的管理,其彼此之间并非是完全独立的,它们之间存在着相互依赖的关系。操作系统常借助于一些表、队列等数据结构来实施管理功能。

5) 工作管理(系统交互与界面的有效利用)

操作系统必须为用户提供一个良好的人机交互界面,用户通过命令操作和程序操作与计算机交互,而交互的环境界面将对用户产生极大的影响,包括心理和思维方面的影响,工作管理模块则极力解决用户操作问题,使计算机系统的使用更方便。

1.2.5 操作系统的服务功能

1. 服务功能

使用操作系统的目的就是为了方便用户,给用户提供一些服务。当然,各个操作系统给用户提供的服务并不是一模一样的,但既然都是操作系统,大部分的服务功能还是相同的。下面给出一般操作系统所具备的服务功能。

(1) 程序执行:启动执行用户程序,并有能力终止程序的执行。

(2) I/O操作:包括文件读写和I/O驱动。专用设备需要专门的程序(如倒带驱动、CRT的清屏等)。

(3) 文件系统管理:用户的程序和数据需要建立文件才能保存在系统中,以后可以按照名字删除它等。

(4) 出错检测:操作系统需要经常了解可能出现的错误。错误来源是多方面的。操作系统对每类错误都要检测到,并采取相应措施,保证计算的一致性。

(5) 资源分配:多个用户或者多道作业同时运行时,每一个用户或作业都必须分得相

应的资源。系统中各类资源都由操作系统统一管理，如 CPU 调度、内存分配、文件存储等都有专门的分配程序，而其他的资源(如 I/O 设备)有更为通用的申请和释放程序。

(6) 统计：通常是了解各个用户对系统资源的使用情况，如用什么类型的资源，用了多少等，以便简单地进行使用情况统计，作为进一步提高服务性能，对系统进行组合的有价值的依据。

(7) 保护：在多用户计算机系统中，用户主要对所创建的文件进行控制使用，并规定其他用户的存取权限。此外，当多个不相关作业同时执行时，一个作业不干扰另一个作业。在多道程序运行环境中，对各种资源的需求经常发生冲突，为此操作系统必须进行调节和合理的调度。

2. 服务方式

操作系统的服务可以通过不同的方式提供，其中两种基本的服务方式是系统调用和系统程序。下面我们来介绍这两种服务方式。

1) 系统调用

所谓系统调用，就是用户在程序中调用 OS 所提供的一些子功能。

系统调用有时也称为广义指令或管理程序调用，它是在用户态下运行的程序和 OS 的界面。用户态程序使用系统调用，可以获得 OS 提供的各种服务。若在运行程序中碰到系统调用命令，则中断现行程序而转去执行相应的系统子程序，以完成特定的系统功能。完成后，控制又返回到发出系统调用命令之后的一条指令，被中断程序将继续执行下去。

对于每一个具体的 OS，它们所提供的系统调用条数、具体格式及所执行的功能，都可能不同。即使是同一 OS 的不同版本，所提供的系统调用也可能有所增减，UNIX 系统在 C 语言和汇编语言级上都提供了系统调用，而大部分 OS 只在汇编级上提供，如 UNIX 系统第 6 版中提供了 42 种系统调用，UNIX S-5 中提供了 64 种系统调用。由此可知，OS 所提供的系统调用的大致范畴为以下 3 种。

(1) 与进程和作业控制有关的系统调用。例如，进程的创建、终止，进程间的同步，进程的睡眠等待，以及设置并获得系统或进程时间等。

(2) 与文件系统管理和设备管理有关的系统调用。例如，文件的创建、删除，文件的打开、关闭、读、写、重新置位等，以及对设备的申请和释放等。

事实上，I/O 设备和文件在很大程度是相似的，所以很多 OS 都把二者并入一类结构，如 UNIX 系统中，I/O 设备就作为特别文件来对待，对用户来说，除了 I/O 设备有专用的名称外，其他操作与普通文件相同。

(3) 与信息维护有关的系统调用。例如，返回当前时间和日期的系统调用，返回 OS 的版本号、存储器空闲区域情况的系统调用等。

2) 系统程序

用户还可以利用键盘命令以求得系统的服务。现代计算机系统都有系统程序包，其中含有系统提供的大量程序，它们解决带共性的问题，并为程序的开发和执行提供了更方便的环境。如很多操作系统都提供了绘画软件包、命令解释程序(UNIX 中是 Shell)等应用程序。命令解释程序是最重要的系统程序，键盘上的控制命令都是由它来进行识别的。