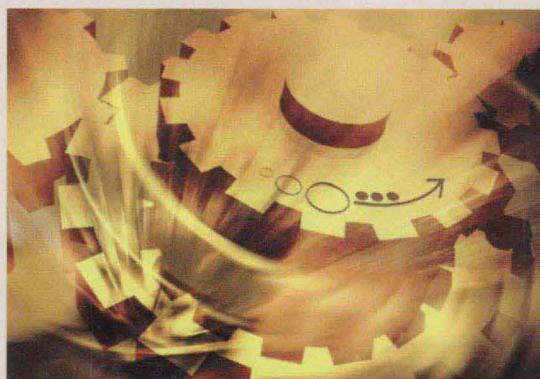


CAO ZUO XI TONG YUAN LI
YU YING YONG



操作系统原理与应用

王育勤 程海英 彭 焱 徐 鹏 © 编著



清华大学出版社

操作系统原理与应用

王育勤 程海英 彭 焱 徐 鹏 编著



清华大学出版社

北 京

内 容 简 介

《操作系统原理与应用》一书重点讲述了操作系统的一般性原理和应用技术。在理论上,力求系统、完整,尽量体现当代的先进研究成果;在讲授方法上,注意理论与实践的结合,特别是以当代最流行的UNIX操作系统为例,介绍了操作系统中主要服务功能的应用技术和技巧;在内容安排上,注意由浅入深,由一般到具体,先介绍操作系统的概念和服务功能,然后一一讲述这些功能的实现算法,并在最后以Linux操作系统为基础,讨论了Linux系统的安装、常用命令及其网络与通信,以其加强实践环节。

本书可作为高等学校计算机科学和工程类的专业教材,也可作为非计算机专业的研究生教材,还可作为从事计算机专业的广大科技工作者学习操作系统的参考用书。

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/王育勤等编著. —北京:清华大学出版社,2013

ISBN 978-7-302-32889-6

I. ①操… II. ①王… III. ①操作系统 IV. ①TP316

中国版本图书馆CIP数据核字(2013)第136594号

责任编辑:赵洛育

封面设计:刘洪利

版式设计:文森时代

责任校对:赵丽杰

责任印制:何 芊

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

印 刷 者:三河市君旺印装厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:19.75 字 数:462千字

版 次:2013年8月第1版 印 次:2013年8月第1次印刷

印 数:1~3000

定 价:39.00元

产品编号:054254-01

前 言

操作系统是计算机系统的基本组成部分，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，起到承上启下的作用。

本书是作者在多年操作系统教学实践积累的基础上，吸收国内外操作系统的新理论和技术，依据操作系统教学大纲的要求编写而成的。

本书重点讲述了操作系统的一般原理和应用技术与方法。在讲授方法上，注重理论与实践的结合，特别是以当代最流行的 UNIX 操作系统为例，介绍了操作系统中主要管理方法和服务功能的应用技术和技巧；在内容安排上，由一般到具体，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 Linux 操作系统为基础，讨论了 Linux 系统的安装、应用及其网络与通信，以其加强实践环节。

全书共分 13 章，具体内容如下：

第 1 章概述了操作系统的发展历史、分类、功能、体系结构及 UNIX 系统的特点等。

第 2 章介绍了进程的基本概念、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 章介绍了处理机管理。

第 4~6 章分别介绍了存储管理、设备管理和文件系统的概念、功能及其主要实现技术。

第 7 章介绍了死锁的概念和解决死锁问题的基本方法。

第 8~13 章是 16 学时的上机实践内容，分别介绍了 Linux 系统的安装和初步使用、使用 Shell 和 Linux 的常用命令、Linux 系统管理、文件服务器与打印服务器、Internet 接入与代理服务器的配置、Linux 服务器配置。

在具体讲述过程中，每章均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。其中第 1~7 章后面附有习题，这些有代表性的习题对巩固正文中的知识是有益的。该书在介绍 UNIX 系统各功能模块的实现方法时，突出重点、难点，结合以往教学的实践体会，对其难以理解的部分给出了较为详细的说明，并给出了生活中的实例，便于学生自学和复习有关内容；还为学生在 Linux 系统环境下的上机实习、应用开发提供了指南。

需要说明的是：使用本书的教师、学生和其他计算机爱好者，可以通过登录清华大学出版社网站 <http://www.tup.com.cn> 下载相关的教学资源来辅助教学和自学，这些资源包括教学大纲、电子教案、PPT 课件和习题答案等。

本书在编写过程中得到了华中科技大学武昌分校信息科学与工程学院欧阳明星院长的

大力支持，在此表示衷心的感谢！

由于编者水平有限，加之时间仓促，书中难免会有疏漏甚至错误之处，恳请广大读者给予指正。

编 者

于华中科技大学武昌分校

目 录

第 1 章 操作系统概述	1	2.1.4 进程	27
1.1 操作系统的发展过程	1	2.1.5 用进程概念说明操作系统的 并发性和不确定性	28
1.1.1 手工操作阶段	1	2.1.6 进程的状态及其变迁	29
1.1.2 早期批处理阶段	2	2.1.7 进程的组成	30
1.1.3 执行系统	3	2.1.8 UNIX 系统的进程映像	33
1.1.4 多道程序系统阶段	3	2.2 有关进程的操作	36
1.2 什么是操作系统	4	2.2.1 进程的创建	36
1.2.1 概念(定义)	4	2.2.2 进程终止和父/子进程的同步	38
1.2.2 设置操作系统的目的	5	2.3 进程间的相互作用和通信	41
1.2.3 操作系统的主要功能	5	2.3.1 同步	41
1.2.4 操作系统的服务功能和方式	7	2.3.2 互斥	42
1.3 操作系统的结构	8	2.3.3 进程的临界区和临界资源	43
1.3.1 环境(外结构)	8	2.3.4 实施临界区互斥的锁操作法	43
1.3.2 体系结构(内结构)	9	2.3.5 信号量与 P、V 操作	46
1.4 操作系统的分类	10	2.3.6 高级通信机构	51
1.4.1 多道批处理系统	10	2.4 中断处理	54
1.4.2 分时系统	11	2.4.1 中断及其一般处理过程	54
1.4.3 实时系统	11	2.4.2 中断优先级和多重中断	57
1.4.4 网络操作系统	13	2.4.3 中断屏蔽	57
1.4.5 区别	13	2.4.4 中断在操作系统中的地位	58
1.5 操作系统的特征	13	2.4.5 UNIX 系统对中断和陷入的处理	58
1.6 操作系统的性能	14	习题	61
1.7 当前比较流行的几种微机操作系统	15	第 3 章 处理机管理	64
1.7.1 当前微机上的主流操作系统	15	3.1 概述	64
1.7.2 如何选用操作系统	18	3.1.1 CPU 调度的三级实现	64
1.8 UNIX 系统的特点和结构	19	3.1.2 进程的执行方式	67
习题	22	3.1.3 CPU 调度的基本方式	67
第 2 章 进程管理	23	3.2 作业概念	68
2.1 进程管理的概念	23	3.2.1 作业管理的概念	68
2.1.1 程序的顺序执行	23	3.2.2 作业管理的功能	71
2.1.2 程序的并发执行和资源共享	23	3.3 常用的调度算法	74
2.1.3 程序并发执行的特性	25		

3.3.1 先来先服务.....	74
3.3.2 最短周期优先.....	75
3.3.3 优先级.....	77
3.3.4 轮转法.....	77
3.3.5 可变时间片轮转法.....	79
3.3.6 多队列轮转法和多级反馈 队列法.....	79
3.4 UNIX 系统中的进程调度.....	81
习题.....	86

第4章 存储管理.....87

4.1 引言.....	87
4.1.1 二级存储器及信息传送.....	87
4.1.2 存储器分配.....	87
4.1.3 存储管理的基本任务.....	88
4.1.4 存储空间地址问题.....	89
4.1.5 用户程序的装入.....	90
4.1.6 存储管理的功能.....	92
4.1.7 内存的扩充技术.....	92
4.2 分区式管理技术.....	95
4.2.1 固定分区法.....	95
4.2.2 可变分区法.....	96
4.2.3 硬件支持.....	98
4.3 可重定位分区分配.....	98
4.4 多道程序对换技术.....	99
4.5 请求分页存储管理.....	100
4.5.1 分页管理.....	100
4.5.2 请求分页管理.....	104
4.6 段式存储管理.....	112
4.6.1 分段和分段的地址空间.....	112
4.6.2 分段管理的实现.....	113
4.6.3 分段共享.....	114
4.6.4 段的动态链接.....	115
4.7 段页式存储管理.....	117
4.7.1 基本思想.....	117
4.7.2 实现过程.....	118
4.8 UNIX 系统的存储管理.....	119
4.8.1 对换.....	120

4.8.2 请求分页.....	122
习题.....	126

第5章 设备管理.....128

5.1 概述.....	128
5.1.1 设备分类.....	128
5.1.2 设备管理的目标和功能.....	129
5.1.3 通道技术.....	129
5.1.4 缓冲技术.....	132
5.2 设备分配技术与 Spooling 系统.....	133
5.2.1 设备分配技术.....	133
5.2.2 Spooling 系统.....	134
5.3 RK 磁盘设备.....	135
5.4 UNIX 系统的设备管理.....	136
5.4.1 UNIX 设备管理的特点.....	136
5.4.2 与设备驱动有关的接口.....	137
5.4.3 块设备管理中的缓冲技术.....	137
5.4.4 块设备的读、写.....	141
5.4.5 字符设备管理.....	145
习题.....	146

第6章 文件系统.....147

6.1 概述.....	147
6.1.1 文件及其分类.....	147
6.1.2 文件系统的功能.....	148
6.1.3 文件系统的用户界面.....	149
6.1.4 文件系统的层次结构.....	150
6.2 文件的组织和存取方法.....	151
6.2.1 文件的逻辑组织和物理组织.....	151
6.2.2 文件的存取方式.....	154
6.3 目录结构.....	155
6.3.1 一级目录结构.....	155
6.3.2 二级目录结构.....	156
6.3.3 多级目录结构.....	157
6.4 文件存储空间的管理.....	159
6.4.1 记住空间分配现状的数据结构.....	159
6.4.2 存储空间分配程序.....	161
6.5 文件保护.....	162
6.5.1 文件系统的完整性.....	162

6.5.2 文件的共享与保护保密	163	第 8 章 Linux 系统的安装和初步使用	202
6.6 对文件的主要操作	164	8.1 Linux 系统的安装	202
6.6.1 创建文件	164	8.1.1 安装前的准备工作	202
6.6.2 文件的连接与解除连接	165	8.1.2 Linux 分区及文件系统	203
6.6.3 文件的打开和关闭	165	8.1.3 Linux 安装步骤	203
6.6.4 文件的读、写	166	8.2 Linux 系统运行时相关的基本概念	207
6.7 文件系统的执行过程	166	8.2.1 Linux 引导程序	208
6.8 UNIX 文件系统的内部实现	168	8.2.2 加载 Linux 内核	209
6.8.1 i 节点	168	8.2.3 Linux Shell	210
6.8.2 活动 i 节点的分配与释放	169	8.2.4 init 进程	210
6.8.3 用户打开文件表和系统打开文件表	170	8.2.5 守护进程	211
6.8.4 目录项和检索目录文件	172	8.2.6 建立终端	212
6.8.5 文件的索引结构	175	8.3 Linux 系统的初步使用	213
6.8.6 文件卷和卷专用块	176	8.3.1 登录	213
6.8.7 空闲 i 节点的管理	177	8.3.2 虚拟控制台切换	213
6.8.8 空闲存储块的管理	179	8.3.3 注销	214
6.8.9 子文件系统装卸和装配块表	179	8.3.4 重启计算机	214
6.8.10 各主要数据结构之间的联系	181	8.3.5 关闭计算机	214
6.8.11 管道文件 (pipe)	181	第 9 章 使用 Shell 和 Linux 的常用命令	216
6.9 系统调用的实例	186	9.1 使用 Linux 基本命令	216
习题	188	9.1.1 常用简单命令	216
第 7 章 死锁	190	9.1.2 目录管理命令	218
7.1 死锁的基本概念	190	9.1.3 文件管理命令	220
7.1.1 什么是死锁	190	9.2 使用命令补齐和别名功能	222
7.1.2 死锁的表示	191	9.2.1 命令行自动补齐	222
7.1.3 死锁的判定法则	192	9.2.2 命令别名	222
7.2 死锁的预防	193	9.3 使用重定向和管道	223
7.3 死锁的避免	195	9.3.1 重定向	224
7.3.1 资源分配状态 RAS	195	9.3.2 管道	224
7.3.2 系统安全状态	196	9.4 熟悉 vi 三种模式下的操作命令	225
7.3.3 死锁避免算法	196	9.4.1 vi 的三种工作模式	225
7.3.4 对单体资源类的简化算法	198	9.4.2 vi 在三种模式下的基本操作	225
7.4 死锁的检测和清除	198	9.5 使用 vi 建立简单的 Shell 脚本并运行	228
7.4.1 死锁的检测	199	9.5.1 创建 Shell 脚本	228
7.4.2 死锁的解除	200		
习题	201		

9.5.2 运行 Shell 脚本	228	11.2.3 配置 vsftpd 服务器	272
9.5.3 Shell 编程基础	229	11.2.4 FTP 匿名登录实例	273
9.5.4 流程控制语句	232	11.3 Samba 服务器	277
第 10 章 Linux 系统管理	236	11.3.1 Samba	277
10.1 磁盘和文件系统管理	236	11.3.2 安装、启动与停止 Samba 服务	278
10.1.1 用户磁盘空间管理	236	11.3.3 配置 Samba 服务	278
10.1.2 文件系统管理	239	11.3.4 配置 Samba 服务的密码文件	281
10.2 账户管理和操作权限	244	11.3.5 Samba 的用户映射文件	282
10.2.1 账户管理	244	11.3.6 Samba 服务的日志文件	283
10.2.2 操作权限管理	250	11.3.7 Samba 应用实例	283
10.3 进程管理	253	11.3.8 Linux 访问 Windows 的共享 资源	289
10.3.1 进程基本概念	253	11.4 Linux 打印服务器	290
10.3.2 进程管理操作	253	第 12 章 Internet 接入与代理 服务器的配置	293
10.3.3 工作任务管理工具	255	12.1 调制解调器接入的配置	293
10.3.4 守护进程	257	12.2 ISDN 接入的配置	295
10.4 RPM 包管理	257	12.3 ADSL 接入的配置	297
10.5 TCP/IP 网络配置	259	12.4 Squid 代理服务器的配置	297
10.5.1 TCP/IP 网络配置文件	259	第 13 章 Linux DNS 服务器配置	300
10.5.2 TCP/IP 网络配置命令	262	13.1 BIND 的安装检查	300
10.5.3 常用网络测试工具	266	13.2 高速缓存 DNS 服务器的配置	300
第 11 章 文件服务器与打印服务器 ...	268	13.3 主 DNS 服务器的配置	301
11.1 RPM 的使用	268	13.4 辅助 DNS 服务器的配置	303
11.1.1 初始化 RPM 数据库	268	13.5 DNS 客户机的配置	304
11.1.2 RPM 软件包的查询功能	268	13.6 DNS 的测试	304
11.1.3 RPM 软件包的配置文件	270	参考文献	305
11.2 FTP 匿名下载	270		
11.2.1 FTP 命令	270		
11.2.2 安装、启动与停止 vsftpd 服务 ..	271		

第 1 章

操作系统概述

1.1 操作系统的发展过程

操作系统这一学科的出现与其他任何新观点、新概念一样，并不是突然产生的，但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（操作系统这门学科是怎样产生的）。

1.1.1 手工操作阶段

早期计算机是十分庞杂的由控制台“指挥”的机器，它使用的是一种很初级的人机交互方式，即先在输入设备（如纸带机、卡片机等）上由人工把程序装入内存，然后启动执行。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。

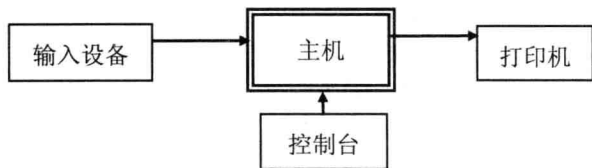


图 1-1 第一代计算机的控制关系

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。

因为具有这些特点，导致它有两个非常严重的缺点：资源浪费和使用不便。因为资源独占和串行工作，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；这就导致资源浪费。另一方面，若是人工干预多，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，以及计算机应用的普及，上述缺点就越来越突出。

【例 1.1】 设有一个程序在两个具有不同运算速度的计算机甲和乙上的运行情况如表 1-1 所示。

表 1-1 程序运行情况

计 算 机	运 算 速 度	运 行 时 间	人工操作时间	比 例
甲	1000 次/秒	1 小时	5 分钟	12:1
乙	60000 次/秒	1 分钟	5 分钟	1:5

在计算机甲上运行该程序时，机时的浪费还可容忍，但在计算机乙上运行该程序时，机时的浪费就不可容忍了。为此就迫使计算机的工程研究人员去花费精力以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），过一段时间后再在打印机上取执行结果，其他操作都是由机器自动进行的，如机器自动输入、编译和执行程序。当一个作业完成后，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，又顺序将结果打印出来）。

这种方式比早期的手工操作先进，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的时间。但这种方式存在很严重的缺点：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

这种方式所用的方法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。

其工作过程为：读卡机上的作业通过卫星机逐个传送到输入带上，而主机只负责把作业从输入带调入内存并运行作业，作业完成后，主机把计算结果和其他信息记录在输出带上，由卫星机负责把输出带上的信息读出来，交由打印机打印。这样，就使主机与慢速设

备间由串行工作变成了并行工作，而且主机是与较快速的磁带机进行信息交换，从而提高了主机效率。由于卫星机仅处理简单的输入/输出工作，因而只需采用较小型的机器即可。

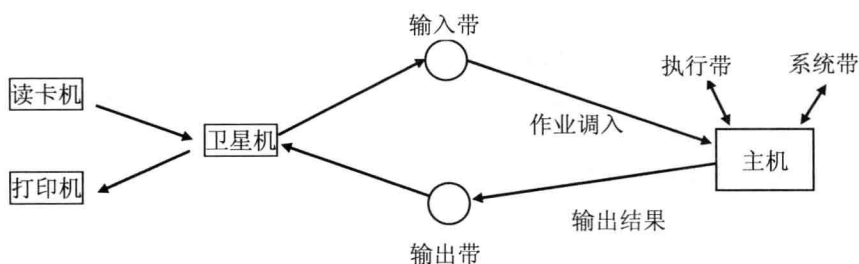


图 1-2 早期脱机批处理系统模型

虽然批处理系统有很明显的优点，如减少人工干预、提高主机效率等，但它又带来一些新的问题，如保护问题。因为在这种系统中，用户程序与监督程序等一些程序之间可以相互调用，无主次之分，这就无法防止用户程序可能破坏监督程序和系统处理程序，所以必须想办法保护监督程序和系统处理程序，这就过渡到了后来的执行系统。

1.1.3 执行系统

其实，执行系统是伴随着通道和中断技术的出现而出现的，有了通道和中断之后，就迫使人们不得不编写中断处理程序和 I/O 控制程序来处理中断和控制通道工作。因为这些程序对其他很多程序都起控制和指挥作用，需常驻内存。一般将具有这种执行程序（控制程序）的系统称为执行系统。

执行程序与监督程序的显著区别是：执行程序对其他程序拥有指挥控制权，它和用户程序之间的关系不是平等的调用关系，而是控制与被控制的关系，其他程序是在其指挥和控制下工作的。这样，系统就可以对不合法的要求进行检查，因而提高了系统的安全性。另一方面，由于执行系统发挥了通道和主机的并发性，亦提高了系统的效率（通道和中断在《计算机原理》课上已接触过，且在以后各章中我们还要介绍，故在此不再介绍）。

不管是批处理系统还是执行系统，由于是单道作业运行，无论如何想办法以提高主机的效率，主机时间的浪费仍然很严重。所以，要想提高主机的利用率，就必须在单道上下工夫，这就促使系统发展到多道程序系统阶段。

1.1.4 多道程序系统阶段

所谓多道程序系统，就是能够控制多道程序并行的系统。多道程序设计的基本思想是在内存里同时存放若干道程序，它们可以并行运行，也可以交替运行。多道程序设计的特点是多道程序并行。从宏观上看，多道程序都处在运行状态，它们之间是并行的；但从微观上看，每道程序又是交替地在 CPU 上运行，它们分时占有 CPU。

【例 1.2】 当 CPU 对第一道程序进行处理后，需要输出时，CPU 在处理完它的 I/O 请求后就转去执行第二道程序的处理工作，这就使第一道程序的 I/O 操作与第二道程序的处理工作并行，第二道程序需要输出时，CPU 处理完它的 I/O 请求后又转向第三道程序，使第三道程序的处理工作与第一、二道程序的 I/O 操作并行，以此类推。

在多道程序系统中，并发程序要共享系统内资源，使系统管理变得很复杂，从而对软硬件管理都提出了许多新课题，也促进了它们的进一步发展，如要解决系统保护、存储分配和简单的动态地址翻译等问题。那么如何有条不紊地来管理计算机系统就作为一门新的学科脱颖而出，这就是我们正在学习的操作系统。下面就来解释什么叫操作系统。

1.2 什么是操作系统

1.2.1 概念（定义）

操作系统是由英文 Operating System 翻译而来的，常简称为 OS，它不是讲解操作方法的，不要因其名字而误解。一个计算机系统是由硬件和软件两大部分组成的。硬件通常指 CPU、存储器、外设等这样一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”，这话固然不很正确，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”——编制不同的软件，才能让它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，即程序和软件是同义语。

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、用于企业管理的软件等）。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。

对操作系统至今尚未有严格定义，现给出以下解释：

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，它为用户使用计算机提供方便、有效和安全可靠的工作环境。

这样一个解释，既说明了操作系统是什么，也说明了操作系统的作用和功能。但它并不十分全面，下面再作简单说明。

（1）从此定义上讲，操作系统是软件而不是硬件，但实际上它是一个软、硬件结合的有机体。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,操作系统不完全是软件,它是一个软、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要完成的任务。由于由硬件直接完成的功能只占很少一部分,故一般说操作系统是软件。

(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。我们知道,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。“万事皆有头”,OS 可以说是这些系统软件的领导者(控制者)。

1.2.2 设置操作系统的目的

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?其主要有如下两个目的。

(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,那么你就无法使用命令,不能使用应用程序,连设备、内存等都需要自己亲自去管理,这显然给用户带来一些非常繁琐的程序设计工作,而这对于多数用户而言往往是无法胜任的,因为必须掌握非常全面的计算机知识并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,摆脱了繁琐的程序设计工作。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。

1.2.3 操作系统的主要功能

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如下。

(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。这些资源信息通过操作系统中各类数据结构和表格记录下来，并且在系统运行过程中不断更新。

(2) 分配资源

操作系统必须对来自用户和应用程序的对资源的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. 操作系统的功能简介

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、交互环境以及互连通信等。所以，常规操作系统的主要任务也针对这 5 个部分，对应地有如下 5 类功能模块。

(1) 处理器(处理机、CPU)管理

计算机系统的“心脏”是处理器，所有软硬件操作都必须由处理器分解执行。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，这就是操作系统处理机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，它们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，以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，使外围设备尽可能与 CPU 并行工作，以提高设备的使用效率并加快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)的，需要时再把它装入内存。文件包括的范围很广，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,以使用户方便、安全地访问它们,是操作系统文件管理的任务。

上述 4 种资源的管理,其彼此之间并不是完全独立的,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,包括心理上 and 思维上的影响,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、适用。

1.2.4 操作系统的服务功能和方式

1. 服务功能

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,各个具体的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,大部分功能还是相同的。下面给出一般操作系统所应有的服务功能。

(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,以后还可以按照名字删除等。

(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,操作系统要检测到每类错误,并采取相应的措施,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:通常希望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,以使用户付款或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工具。

(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,并规定其他用户对它的存取权限。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,操作系统必须作调节和合理的调度。

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,下面分别进行介绍。

(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，则中断现执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、具体格式以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。但不管怎么样，所提供的系统调用的大致范畴有以下 3 种。

① 与进程和作业控制有关的系统调用

如进程的创建、终止，进程间的同步，进程的睡眠等待，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

如文件的创建、删除，文件的打开、关闭、读、写、重新置位等，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上是相似的，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

如返回当前时间和日期、OS 的版本号、空闲内存或盘空间数量的系统调用等。

(2) 系统程序

用户还可以利用键盘命令来求得系统的服务。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。

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。

1.3 操作系统的结构

操作系统的结构包括其环境和体系结构，下面进行介绍。

1.3.1 环境（外结构）

OS 的外部环境主要是指硬件、其他软件 and 用户（人）。OS 与外部环境的关系如图 1-3 所示。

由图 1-3 可知，OS 是核心的系统软件，它与硬件的关系最为密切，且有些功能就是由硬件和软件共同配合完成的，如中断系统。OS 是整个计算机系统的控制管理中心，其中包