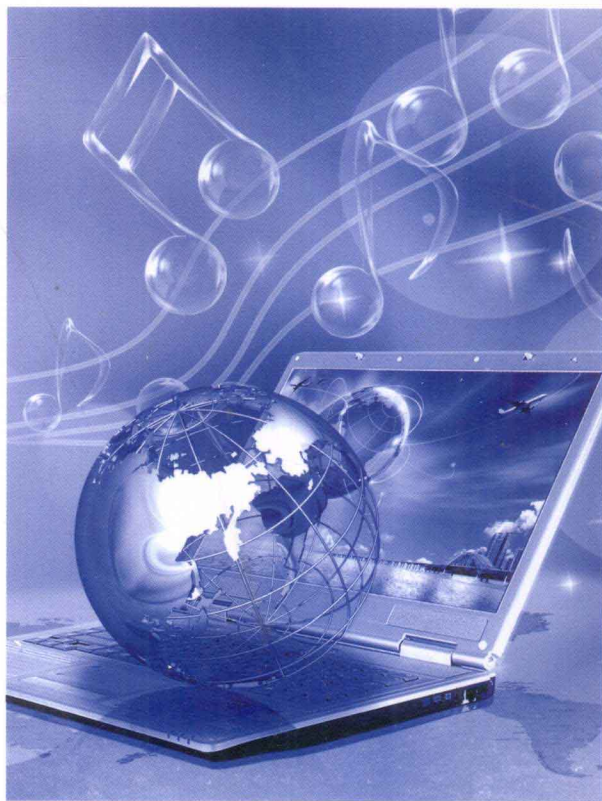


操作系统原理与 实践教程 (第二版)

- ◆ 操作系统概论
- ◆ 操作系统的界面
- ◆ 处理器管理
- ◆ 进程同步与死锁
- ◆ 存储管理
- ◆ 文件管理
- ◆ 设备管理
- ◆ 操作系统的安全和保护



史韦航 编著

高等学校计算机应用规划教材

操作系统原理与实践教程

(第二版)

陶永才 史韦航 编著

清华大学出版社

北 京

内 容 简 介

本书全面系统地介绍了现代计算机操作系统的基本理论和最新技术。全书共分为 8 章,第 1 章介绍了操作系统的概念、特征和功能、发展历史、操作系统结构和设计的相关问题,并介绍了常用的操作系统产品;第 2 章介绍了操作系统的生成与引导、操作系统向用户提供的接口和服务;第 3、4 章详细介绍了进程和线程的概念、处理器调度、同步和通信机制、调度与死锁等问题;第 5、6、7 章分别介绍了操作系统的存储管理、文件管理和设备管理功能;第 8 章介绍了操作系统的安全和保护问题。

本书可作为高等院校计算机科学与技术、软件工程、信息管理等专业本、专科生的教材和考研、考证参考书,也可供从事计算机科学、工程和应用等方面工作的科技人员参考。

本书每章对应的电子教案和习题答案可以到 <http://www.tupwk.com.cn/downpage/index.asp> 网站下载。

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

操作系统原理与实践教程(第二版)/陶永才,史韦航 编著. —北京:清华大学出版社,2010.10

(高等学校计算机应用规划教材)

ISBN 978-7-302-23697-9

I. ①操… II. ①陶… ②史 III. ①操作系统—高等学校—教材 IV. TP316

中国版本图书馆 CIP 数据核字(2010)第 165182 号

责任编辑:胡辰浩(huchenhao@263.net) 袁建华

装帧设计:孔祥丰

责任校对:成凤进

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

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 装 者:三河市春园印刷有限公司

经 销:全国新华书店

开 本:185×260 印 张:17.75 字 数:410 千字

版 次:2010 年 10 月第 2 版 印 次:2010 年 10 月第 1 次印刷

印 数:1~5000

定 价:29.00 元

产品编号:034871-01

前 言

操作系统是计算机系统中最重要系统软件，它管理整个计算机系统的软件和硬件资源，是其他软件和程序的运行基础，是沟通用户与计算机硬件的桥梁。操作系统因其在计算机系统中所处的地位决定了它的重要性，它是计算机科学与技术专业的一门专业基础课，是计算机相关专业学生的必修课程。操作系统是计算机领域中最活跃的学科之一，其发展极为迅速。

操作系统具有如下特点：

(1) 内容庞杂，涉及面广。操作系统是一个庞大的系统软件，它管理系统中所有的软件、硬件资源，控制计算机的工作流程，提供用户与计算机之间的接口。因此，本课程的内容非常庞大且复杂。

(2) 内容较抽象。操作系统在计算机系统中处于裸机与应用层之间，对下与硬件接口，对上提供简单便捷的用户界面。但是对于什么是操作系统、操作系统应具有哪些功能、这些功能如何实现等问题并不是很清楚，因而对操作系统的内容倍感抽象、费解。

(3) 发展变化快。操作系统是计算机领域中最活跃的学科之一，其发展极为迅速，随着计算机的发展而不断发展，是计算机软件中变异、更新最快的软件，因而更加重了学习难度。

正是由于操作系统的上述特点，使得这门课程的学习具有相当的难度。为了解决这些问题，提高操作系统课程的教学质量，在广泛汲取国内外优秀教材和研究成果的基础上，笔者编写了本教材。在编写过程中，力求覆盖面广、内容新颖、重点突出。本书共分为 8 章，参考学时数为 60~80 学时。各章内容简述如下：

第 1 章介绍了操作系统的概念、功能、特征、发展历史和结构，并从操作系统的发展入手分析操作系统的发展方向，引入不同结构的操作系统的性能比较和分析，对最为流行的操作系统实例——Windows 系列和 Unix/Linux 系统进行简要的介绍。

第 2 章介绍了操作系统的生成和启动，以及操作系统提供的服务和接口。

第 3 章首先介绍进程与线程的基本概念，重点介绍进程的定义、状态等知识，并说明进程控制过程和方法。然后对处理器调度的实现和调度算法进行阐述。最后介绍在现代网络环境和实时系统中使用的多处理器调度和实时调度的一些基本原理。

第 4 章首先介绍进程同步和互斥的基本概念。然后阐述如何通过信号量机制和管程来实现进程的同步与互斥。最后介绍进程死锁的基本概念，并阐述如何预防和避免进程死锁的发生。

第 5 章介绍存储管理的基本概念和常见的存储管理方法，并分别介绍了各种内存管理技术的实现思想、算法和硬件支持，探讨各种算法的优缺点。

第 6 章介绍了文件系统中有关文件管理的基本概念、文件的逻辑结构和物理结构、文

件存储空间的管理、文件的共享和保护等内容。

第 7 章介绍了设备管理的基本概念、I/O 控制方式、中断技术、缓冲技术、设备分配和 I/O 软件原理、磁盘调度和管理等内容。

第 8 章首先介绍信息系统安全的概念。然后阐述操作系统的安全机制,包括数据加密、数字签名、身份认证、授权机制和审计。最后介绍计算机病毒的基本概念、常见的计算机病毒类型以及如何预防和检测计算机病毒。

本书可作为普通高校计算机科学与技术及其相关专业本、专科学生的教材,也可作为自学参考书和考研参考书。本书为任课教师免费提供电子教案(此教案用 PowerPoint 制作,可以任意修改),需要者请到清华大学出版社网站 <http://www.tupwk.com.cn/downpage> 下载。

本书由陶永才、史韦航编著整理,在编写和出版过程中,感谢郑州大学石磊教授提出的许多宝贵、中肯的意见,使我们获益匪浅。马秀华、卫琳等给予了大力支持和帮助,同时王战红、杨仕飞、史海振、张聪、陈文远、屠卫、李颖秋、姚培娟、薛正元、郭东东、王晖等付出了努力,在此一并向他们表示诚挚的感谢。另外,对清华大学出版社的领导及有关同志深表谢意,谢谢他们对此书的出版和发行做出的大量工作。

本书的编写参阅了大量书籍和资料,主要的参考文献列于书后,在这里对这些书籍和资料的编著者表示诚挚的谢意。

由于编者水平有限,难免会有错误和不当之处,恳请读者批评指正。我们的联系方式 huchenhao@263.net; 电话 010—62796045。

作 者

2010 年 6 月

目 录

第 1 章 操作系统概论	1	2.1.3 实例分析: Linux 系统启动	32
1.1 操作系统的概念	1	2.2 操作系统的用户接口	34
1.1.1 操作系统的定义	1	2.2.1 命令接口	34
1.1.2 操作系统的特征	3	2.2.2 图形用户接口	35
1.1.3 操作系统的功能	4	2.2.3 程序接口	36
1.1.4 操作系统的设计目标	8	2.3 操作系统提供的服务	37
1.1.5 操作系统的性能指标	9	2.3.1 操作系统提供的基本服务	37
1.2 操作系统的形成和发展	9	2.3.2 操作系统提供的公共函数	38
1.2.1 人工操作阶段	9	2.4 小结	39
1.2.2 单道批处理	10	2.5 思考练习	39
1.2.3 多道批处理	10	第 3 章 处理器管理	41
1.2.4 分时操作系统	12	3.1 进程的引入	41
1.2.5 个人计算机操作系统	13	3.1.1 程序的顺序执行	41
1.2.6 实时操作系统	14	3.1.2 程序的并发执行	42
1.2.7 网络操作系统	15	3.2 进程的概念	43
1.2.8 分布式操作系统	16	3.2.1 进程的定义	43
1.2.9 操作系统的进一步发展	17	3.2.2 进程的状态	44
1.3 操作系统的结构	19	3.2.3 进程控制块	46
1.3.1 整体结构	19	3.3 进程控制	47
1.3.2 分层结构	20	3.3.1 进程控制的概念	47
1.3.3 虚拟机	21	3.3.2 进程控制机构	47
1.3.4 外核	22	3.3.3 进程创建	48
1.3.5 客户机/服务器结构	23	3.3.4 进程撤销	48
1.4 常用操作系统简介	24	3.3.5 进程阻塞	49
1.4.1 Windows 系列操作系统	24	3.3.6 进程唤醒	49
1.4.2 UNIX 和 Linux 系统	25	3.4 线程	49
1.5 小结	27	3.4.1 线程的概念	49
1.6 思考练习	28	3.4.2 线程的特点与状态	50
第 2 章 操作系统的界面	29	3.4.3 线程的实现	50
2.1 用户工作环境	29	3.4.4 多线程模型	51
2.1.1 用户环境	29	3.4.5 线程池	52
2.1.2 系统的生成与引导	30	3.5 处理器调度	53

3.5.1	处理器调度的层次	53	4.4.2	进程通信的方式	92
3.5.2	选择调度算法的准则	54	4.4.3	消息传递系统	93
3.5.3	作业调度	55	4.4.4	消息缓冲队列通信机制	95
3.5.4	进程调度	56	4.4.5	管道通信方式	97
3.6	调度算法	57	4.5	死锁	98
3.6.1	先来先服务	57	4.5.1	死锁的概念	98
3.6.2	短作业(进程)优先	58	4.5.2	死锁产生的原因和必要条件	99
3.6.3	优先级调度	58	4.5.3	死锁的描述—— 资源分配图	101
3.6.4	时间片轮转法	59	4.5.4	处理死锁的方法	103
3.6.5	多级队列算法	60	4.6	死锁的预防和避免	103
3.6.6	多级反馈队列调度	61	4.6.1	死锁的预防	103
3.6.7	高响应比优先调度	62	4.6.2	死锁的避免	105
3.7	多处理器调度和实时调度	63	4.7	死锁的检测和解除	111
3.7.1	多处理器调度	63	4.7.1	死锁的检测	111
3.7.2	实时调度	64	4.7.2	死锁的解除	113
3.8	小结	64	4.8	小结	114
3.9	思考练习	65	4.9	思考练习	115
第4章	进程同步与死锁	67	第5章	存储管理	119
4.1	进程的同步和互斥	67	5.1	存储管理的概念	119
4.1.1	进程的同步	67	5.1.1	多级存储结构	120
4.1.2	进程互斥	68	5.1.2	程序的运行过程	122
4.1.3	信号量机制	71	5.1.3	存储管理的任务和功能	124
4.1.4	信号量的使用方法	76	5.1.4	存储管理方式	128
4.1.5	信号量的应用实例	77	5.2	连续内存分配	130
4.2	经典同步问题	78	5.2.1	单一连续分配	130
4.2.1	生产者-消费者问题	79	5.2.2	固定分区分配	130
4.2.2	读者-写作者问题	80	5.2.3	动态分区分配	132
4.2.3	哲学家进餐问题	83	5.2.4	可重定位分区分配	136
4.2.4	理发师问题	86	5.3	内存不足时的管理	138
4.3	管程	87	5.3.1	覆盖	138
4.3.1	管程的基本概念	88	5.3.2	交换	139
4.3.2	条件变量	89	5.4	分页存储管理	140
4.3.3	使用管程解决生产者- 消费者问题	90	5.4.1	分页存储管理的基本原理	140
4.4	进程通信	91	5.4.2	地址映射	143
4.4.1	进程通信的概念	91	5.4.3	页表的结构	145
			5.4.4	页面的共享	147

5.5	分段存储管理	148	6.5.3	随机文件	190
5.5.1	分段存储管理方式的引入	148	6.5.4	文件物理结构比较	192
5.5.2	分段存储管理的基本原理	149	6.6	文件存储空间的分配	192
5.5.3	段的共享和保护	151	6.6.1	连续分配	192
5.5.4	段页式存储管理	153	6.6.2	链接分配	194
5.6	虚拟存储器	155	6.6.3	索引分配	198
5.6.1	虚拟存储器的引入	155	6.7	文件存储空间的管理	200
5.6.2	虚拟存储器的实现方法	157	6.7.1	空闲文件目录	201
5.6.3	虚拟存储器的特征	157	6.7.2	空闲链表法	202
5.7	请求分页存储管理技术	158	6.7.3	位示图	203
5.7.1	请求分页存储管理的硬件支持	158	6.8	文件系统	204
5.7.2	页面置换算法	160	6.8.1	文件系统的结构	204
5.7.3	请求分页系统内存块的分配	165	6.8.2	文件系统的实现	205
5.7.4	工作集理论和抖动问题	167	6.9	文件的共享和保护	206
5.8	小结	168	6.9.1	文件的共享	206
5.9	思考练习	170	6.9.2	文件保护	208
			6.9.3	文件系统的可靠性	209
第 6 章	文件管理	173	6.10	小结	210
6.1	文件的概念	173	6.11	思考练习	210
6.1.1	文件及其分类	173	第 7 章	设备管理	213
6.1.2	文件的属性	176	7.1	设备管理的概念	213
6.2	文件目录	176	7.1.1	设备的分类	213
6.2.1	文件控制块和文件目录	177	7.1.2	设备管理的任务和功能	215
6.2.2	文件目录结构	178	7.1.3	I/O 系统结构	216
6.2.3	目录的实现	182	7.1.4	设备控制器	217
6.3	文件和目录操作	183	7.2	I/O 控制方式	219
6.3.1	文件操作	183	7.2.1	程序直接控制方式	219
6.3.2	目录操作	184	7.2.2	中断方式	220
6.4	文件的逻辑结构	185	7.2.3	DMA 方式	222
6.4.1	文件逻辑结构的类型	186	7.2.4	通道控制方式	223
6.4.2	顺序文件	187	7.3	中断技术	225
6.4.3	索引文件	187	7.3.1	中断的基本概念	226
6.4.4	索引顺序文件	188	7.3.2	中断的作用	226
6.5	文件的物理结构	189	7.3.3	中断的分类与优先级	227
6.5.1	连续文件	189	7.3.4	软中断	227
6.5.2	链接文件	190	7.3.5	中断处理过程	228

7.4 缓冲技术.....	229	7.9 思考练习.....	256
7.4.1 缓冲的引入.....	229	第 8 章 操作系统的安全和保护.....	257
7.4.2 缓冲的种类.....	230	8.1 计算机系统和操作系统安全.....	257
7.4.3 缓冲池的管理.....	230	8.1.1 计算机系统安全概述.....	257
7.5 设备分配.....	233	8.1.2 操作系统安全及信息 安全评价准则.....	258
7.5.1 设备分配中的数据结构.....	233	8.1.3 计算机安全威胁分类.....	259
7.5.2 设备分配的原则.....	234	8.2 操作系统安全机制.....	260
7.5.3 SPOOLing 系统.....	236	8.2.1 数据加密的基本概念.....	260
7.5.4 虚拟设备—共享打印机.....	237	8.2.2 数字签名和身份认证.....	262
7.5.5 SPOOLing 系统的特点.....	237	8.2.3 授权机制.....	264
7.6 I/O 软件.....	238	8.2.4 审计.....	264
7.6.1 I/O 软件的设计目标 和原则.....	238	8.3 计算机病毒.....	265
7.6.2 I/O 中断处理程序.....	239	8.3.1 计算机病毒的基本概念.....	265
7.6.3 设备驱动程序.....	240	8.3.2 计算机病毒的类型.....	267
7.6.4 设备无关软件.....	242	8.3.3 病毒的预防和检测.....	268
7.6.5 用户层 I/O 软件.....	244	8.4 访问控制机制.....	270
7.7 磁盘调度和管理.....	245	8.4.1 保护域.....	270
7.7.1 磁盘结构.....	245	8.4.2 访问矩阵.....	271
7.7.2 磁盘访问时间.....	246	8.4.3 访问矩阵的实现.....	272
7.7.3 磁盘调度.....	247	8.5 小结.....	273
7.7.4 磁盘高速缓存.....	250	8.6 思考练习.....	274
7.7.5 磁盘管理.....	251	参考文献.....	275
7.7.6 廉价磁盘冗余阵列.....	253		
7.8 小结.....	255		

第1章 操作系统概论

本章学习目标

- 掌握操作系统的定义及其在计算机系统中的作用。
- 掌握操作系统的特征及其功能。
- 了解操作系统的形成过程及其发展趋势，掌握批处理系统、分时系统和实时系统的特点。
- 了解操作系统的设计结构。
- 了解常见操作系统的特点。

本章概述

在现代计算机系统中，一个或多个处理器、主存、外存、网络接口以及各种不同的输入/输出设备共同协作，完成用户的各项需求。用户需求的响应过程是十分复杂和关键的，对编写和监督管理上述各种部件的程序员能力要求极高。为了将部分关键的操作封装起来，同时也达到简化程序员工作的目的，计算机体系中出现了操作系统这个软件层次。它能在管理并正确使用上述部件的同时，为程序员提供一个通用的、相对简单的、能够驱动硬件工作的软件接口。

本章首先从操作系统的定义、特征、功能、设计目标、性能指标等方面阐述操作系统的概念。然后从操作系统的发展分析操作系统的发展方向，由此引入不同结构的操作系统的性能比较与分析。最后对最为流行的操作系统实例——Windows 系列和 Unix/Linux 系统进行简要的介绍。

1.1 操作系统的概念

1.1.1 操作系统的定义

在现代计算机体系结构中，操作系统起着至关重要的作用。如图 1-1 所示为操作系统在计算机体系结构中的位置。操作系统是硬件之上的第一层软件，在操作系统之上的是各种应用程序。其中每个层次又可以细分为更多的子层，如硬件层从底向上可分为物理设备、由各种寄存器和数据通道组成的微体系层以及主要由指令集组成的机器语言层，提供的是基本的计算资源。应用程序层则通常是基于特定操作系统的、满足特定功能的、直接面向用户的软件，这些软件能够根据用户的具体需求申请特定的资源，并按照应用程序规定的方法来使用这些资源。操作系统处于这两个层次之间，用来协调与控制应用程序对硬件资

源的使用。

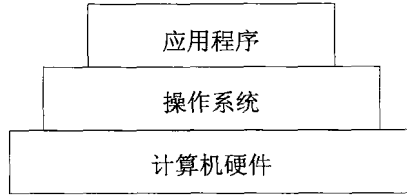


图 1-1 计算机体系结构

在当今社会，几乎每个人都与操作系统打过交道，但是要精确地给出操作系统的定义却并非易事。由于每个人看待操作系统的角度不同，使用操作系统的目的不同，看到的操作系统也就表现出不同的特征。下面我们将从不同角度来探讨这个问题，并总结操作系统的定义。

1. 资源管理角度

从资源管理角度来看，操作系统可以被视为资源管理与分配器。操作系统是硬件之上的第一层软件，可以与硬件直接交互，对硬件资源具有最直接有效的控制和管理权限。同时，作为应用程序层的基础，操作系统又要为应用程序提供各种使用硬件的方法，即应用程序接口。因此，这个层次的软件应该能够直接操控各种计算机资源。

计算机资源分为硬件资源与软件资源。硬件资源是指作为计算机运算基础的所有物理设备，以及为方便用户所使用的鼠标、键盘、打印机等各种不同类型的外部设备。这类资源使用特定的电子信号来指挥，由电子工程师设计并提供相应的驱动程序。而在操作系统中则使用这些驱动程序以及特定的指令集来告知硬件如何工作，同时接收硬件发送来的反馈数据与状态信息。根据硬件资源的功能不同，又将其分为处理器、存储器、I/O 设备。相应地，操作系统也针对不同类型的硬件专门规划了处理机管理模块、存储器管理模块以及 I/O 设备管理模块。计算机的软件资源通常是指各种程序与数据资源，它们以程序的形式或各种不同类型的文件形式存放于外存上，操作系统要将其进行合理化存储，以保证空间利用率和读写效率之间的均衡与有效。

2. 用户观点

从用户观点来看，操作系统是用户与计算机硬件系统之间的接口，该接口在使用便捷性、资源利用率方面表现突出。

由于操作系统是一般用户可以接触到的最底层的软件，只有通过它所提供的接口，用户才能使用各种硬件资源。换句话说，操作系统将复杂的底层机器语言和操作屏蔽起来，并将常用操作和指令序列组合后以命令、系统函数调用、图形用户界面等方式呈现给用户，帮助用户以更安全、高效、便捷的方式使用系统资源。因此我们称操作系统是一种人机交互接口。

在大型机和工作站端，这个接口除了能够帮助本机用户更方便地使用资源外，通常还肩负着为该用户与其他联机用户分配资源的重任。而分配资源的最重要原则就是确保 CPU 时间、内存和 I/O 设备得到最充分的利用，以达到资源利用率最大化的目的。

3. 机器扩充角度

由计算机所完成的工作，无论繁简，总是可以分解为各个不同硬件的序列性动作。这些动作通过控制器命令来完成。控制器命令有不同的种类，可以完成数据读写、磁头臂移动、磁道格式化、状态检测等不同工作。每条控制器命令均需要读写特定位置的数据，并从中分析所要执行的动作和被操作的数据等信息，然后按照分析结果完成命令动作，最后反馈新的状态信息和返回值到指定位置。显然，要求一般程序员使用控制器命令完成任务是不现实的，程序员需要的是高度抽象的、简单的操作方法。

基于上述原因，一个专门用来隐藏硬件的实际工作细节，并提供了一个可以读写的、简洁的命名文件视图的软件层次被引入了计算机体系结构中，这就是操作系统。

综上所述，操作系统就是一组管理与控制计算机软硬件资源并对各项任务进行合理化调度，并且附加了各种便于用户操作的工具的软件层次。

1.1.2 操作系统的特征

操作系统种类繁多，但却有一些共同特征，这些特征也是操作系统这一软件层次与应用软件的区别所在。现代操作系统都具有并发、共享、虚拟和异步特性，其中并发性是操作系统最重要的特征，其他三个特性均基于并发性而存在。

1. 并发性

在说明什么是并发性之前，首先要区分两个概念：并发和并行。如果在一个时间段内发生了一个以上的事件，则称这几个事件具有并发性；而并行性指的是这多个事件在同一时刻点发生。

在不同类型的操作系统中，并发性的含义有一定的区别。在单处理机系统中，每个特定时刻只能有一个程序占用 CPU。但一个较长的时间段可以被分为多个小的时间碎片，这些碎片可以按照一定的原则分发给多个不同的程序，使得在这个时间段内有多个程序得到一定程度的执行。这些程序是具有并发性，不具有并行性的。而在多处理机系统中，每个特定时刻有多个 CPU 可以使用，在这样的时刻点，多个可以独立运行的程序就能够并行执行。

2. 共享性

操作系统中的共享指的是多个并发执行的程序能够按照一定的规则共同使用操作系统所管理的软硬件资源。由于这些资源具有不同的使用要求，因此其共享方式也有所不同。操作系统所管理的软硬件资源按照使用方式可以分为同时访问方式和互斥访问方式。

同时访问方式是指在一段时间内允许多个程序同时访问。这里的“同时”指的是宏观上的一个“时间段”内的同时访问。从微观上讲，这些程序可能是顺序或者轮替地使用该资源。常见的使用同时访问方式进行共享的资源有磁盘、某些程序的公共缓冲区等。

互斥访问方式是指在一段时间内只允许一个程序访问的资源，而这类资源被称为临界资源。临界资源通常需要较长时间来处理一个不可被中断的任务，例如打印机、某些程序

的公用数据等。当临界资源空闲时，会对到达的第一个资源请求给出回应，在处理该请求的过程中，若又有新的资源请求来到，则不会予以理会。这种方法可以保证一个连续任务的无误处理，避免交叉打印或计算错误的发生。

3. 虚拟性

这里所谓的虚拟性并不是虚拟机，而是将计算机体系结构中的各种物理设备映射为多个逻辑设备。这种映射通常是利用分时共享的方式实现的，被映射的物理设备有多种，如内存、外设、CPU 等。每个不同的设备，由于其工作模式不同，所使用的映射方法也不尽相同。

使用虚拟存储器技术，可以将较小的物理内存“扩充”为较大的虚拟内存。这种方法的核心思想就是仅将程序当前运行所需的数据和代码载入内存，当这个程序的一个相对独立的功能模块运行完成或暂时无法继续进行，这个模块所对应的数据和代码将被暂存到外存的指定区域，其释放的内存空间将被重新分配给本程序的其他功能模块或其他应用程序的功能模块使用。使用这种方法可以确保对空间有较大要求的应用程序也能正常运行于小内存的机器上。

类似地，虚拟处理器技术也利用了分时方法，且使用多道程序设计技术保证多道程序并发执行。在一段时间内，CPU 将分时间片处理多个程序请求，但提出这些请求的各个用户并不会感觉到有其他人和自己共用 CPU 时间，而是感觉自己独占了资源。在硬件能力快速发展的今天，这种方法能够最大限度地发挥联网机器的效用，提高处理效率。

对设备的虚拟则是将一个物理设备映射为多个虚拟的逻辑设备，尽管进行不同程序处理的仍然是一台机器，但在用户感知上则是有多台机器同时进行数据处理。这种映射的实现仍然依靠分时共享方法。

通过上面的介绍可以看出，虚拟特性的实现主要依靠分时共享方法和多道程序设计技术，通过虚拟操作系统可以将一个设备映射为多个，将一个设备的能力均分到不同的逻辑设备上，以便多用户共享资源。虚拟方法不会造成多个用户长期等待其他用户的情况，同时也极大地提高了资源利用率。

4. 异步性

异步性是指操作系统中的各个程序的推进次序无法预知。异步性的产生是由现代操作系统的并发性引起的。在并发执行的多个进程之间，何时能够获得所需资源，何时等待哪些进程释放资源，以及当前占有资源的进程何时释放资源等因素都是不确定的，因此用户是无法预知各个进程的执行时间。异步性是现代操作系统的一个重要特征。

1.1.3 操作系统的功能

现代操作系统的主要任务就是维护一个优良的运行环境，以便多道程序能够有序地、高效地获得执行，而在运行的同时，还应该尽可能地提高资源利用率和系统响应速度，并

保证用户操作的方便性。因此，操作系统的基本功能包括处理器管理、存储器管理、设备管理和文件管理。此外，为了给用户提供一个统一、方便、有效地使用系统能力的手段，现代操作系统还需要提供一个友好的人机接口。在互联网不断发展的今天，操作系统通常还具备基本的网络服务功能和信息安全防护等方面的支持。

1. 处理器管理功能

处理器是计算机软硬件体系结构的“心脏”，是制约整个计算机体系性能的最重要器件，因此，处理器性能是否被充分发挥关系着整个计算机体系的性能。操作系统的主要任务之一就是合理有效地管理处理器，使其在现有环境下尽可能地发挥最大功效，提供更高的处理效率。

处理器管理功能主要体现在创建、撤销进程，并按照一定的算法为其分配所需资源，同时还要管理和控制各用户的多个进程的协调运行，确保各个进程可以正确的通信。在多道程序操作系统中，这些管理功能最终通过对进程的控制和管理来实现，而在具有线程机制的操作系统中，这些功能的实现还依赖于对线程的管理和控制。

2. 存储器管理功能

存储器是用来存放程序和数据的容器，它是为计算机系统提供运作数据和具体指令序列的器件。操作系统所管理的存储器包括内存、外存等，存储器管理的主要任务就是将各种存储器件统一管理，保证多道程序的良好运行环境，同时，还要兼顾内存利用率、逻辑上扩充内存的需求以及用户的感受，提供优良的控制、存取功能，为用户提供控制存储器的手段。

为了实现上述要求，存储器管理应该具有内存分配、内存回收、内存保护、地址映射和虚拟内存等功能。

(1) 内存分配

内存分配是指为每道程序分配合适的内存空间，使其能在运行期间将运行所需数据放置在内存指定区域，以保证 CPU 能够顺利地获取指令并存取指定数据。分配内存空间时应尽量提高内存的利用率，减少不可用内存空间。此外，还应能响应正在运行的程序发出的动态内存申请，以便满足新增指令和数据对新空间的需求。

内存分配通常采用动态和静态两种方式。静态分配方式指的是程序在装入内存时需要估计所需空间，一旦进入内存开始运行，就不能再申请新的空间，也不能将该程序所占用的空间“搬运”到其他位置；动态分配方式指的是尽管程序装入内存时申请了一定的空间，但在程序运行期间还可以为运行过程中所需的新的程序和数据再申请额外的空间，以满足程序空间动态增长的需要。

(2) 内存回收

内存回收是指当程序运行完毕后，将各个程序在装入内存时所分配的空间重新置为空闲分区，并交由操作系统统一管理，以备其他程序申请使用。

在内存的分配和回收过程中，为了记录当前内存的使用和分配情况，操作系统通常还

要配置内存分配数据结构, 以便为后期分配和回收提供依据。

(3) 内存保护

在多道程序环境中, 为了保证每个用户的各个程序独立运行, 不会相互影响, 需要提供内存保护机制。该机制的主要任务就是确保每道程序都在自己的内存空间中运行, 决不允许任何程序访问或存取其他程序的非共享程序和数据。

内存保护机制的实现有多种方式, 常见的一种处理方法是利用上下界限寄存器。这两个寄存器中存放的数值是当前正在运行的程序的内存空间的起始地址和终止地址, 每当 CPU 要求访问某个地址的程序或数据时, 操作系统会先利用上下界限寄存器与之相比较, 若在这两个界限内, 则可以正确访问, 否则就拒绝此次的内存访问。通过这种方式, 可以确保在进程运行期间不会误访问无权限空间。

(4) 地址映射

在多道程序环境中, 每个程序都使用自己的独立空间。这些空间分布于内存的不同位置, 但每个程序员在编码时通常并不知道自己的程序进入内存后会被放置在什么位置, 因此也不可能在程序中直接使用内存单元地址来操作所需的指令或数据。为了解决这个矛盾, 当前操作系统都提供了地址映射机制。该机制的基本思想是将用户与内存分隔, 即在程序员编码时, 直接以“0”作为程序中出现的其他任何地址的初始位置, 该位置被称为逻辑基址。当该程序被编译和连接之后, 形成可装入的可执行文件。根据内存的使用情况, 操作系统会在可执行文件真正装入内存时为其分配合适大小的空闲空间, 此空间的初始位置称为物理基址。当程序运行时, CPU 需要查询某位置的数据或指令时, 只需给出相对于逻辑基址的偏移量, 操作系统就会根据逻辑空间内容的分布情况自动将该逻辑地址转换为内存中对应的物理地址。地址映射功能需要硬件机构的协助, 以保证数据的快速定位与存取。

(5) 虚拟内存

虚拟内存技术在当今多数操作系统中都有涉及, 它指的是利用特殊技术将磁盘的一部分空间实现较快的存取, 从逻辑上扩充内存容量, 使得用户感觉到内存的容量比实际物理内存所提供的空间要大。这种方式可以提高多道程序的速度, 提升系统吞吐量, 获得更好的系统性能。而为了实现虚拟内存, 只需配置简单的内存扩充机制即可。该机制的核心内容是请求调入功能和置换功能。

请求调入功能允许程序仅向内存装入保证启动的必需数据和指令, 当程序在运行过程中需要新的数据和程序时, 先中断自身运行, 并向操作系统提出调入请求, 由其从磁盘将所需数据和指令调入内存, 然后继续从被中断的地方执行。

置换功能是指在操作系统将所需新数据或指令调入内存时, 若发现内存空间不足, 需要从正处于内存中的数据或程序中选择部分暂时不用的调出到磁盘上, 腾出的空间则用来调入当前的急需数据。

3. 设备管理功能

在计算机系统中, 外围设备的地位举足轻重, 它是用户直接接触的对象, 可用来进行人机交互。

设备管理的主要作用是使用统一的方式控制、管理和访问种类繁多的外围设备。设备管理功能主要体现在：接收、分析和处理用户提出的 I/O 请求，为用户分配所需 I/O 设备，同时，还要做到尽量提高 CPU 和 I/O 设备利用率、I/O 处理效率，为用户提供操控 I/O 设备的便捷界面和手段。根据设备管理模块的功能要求，可以将其功能分为设备分配、缓冲管理、设备处理、虚拟设备等。

设备分配的主要功能是根据用户的 I/O 请求和系统的设备分配策略，从系统当前空闲资源中选择所需类型设备，并将其使用权限交付给用户。如果 I/O 子系统中还包括通道和设备控制器，则设备分配还要负责选择空闲通道和控制器并交付用户使用。

缓冲管理的主要功能是合理组织 I/O 设备与 CPU 之间的缓冲区，并提供获取和释放缓冲区的有效手段。在计算机系统中，凡是数据到达速度和离开速度存在差异的地方均可设置缓冲区以缓解速度矛盾，高速 CPU 和低速 I/O 设备之间就符合这个条件，并且由于二者的利用率关系到整个系统的处理效率和响应速度，因此在操作系统中均为其设置缓冲层次。设置缓冲区的作用不仅体现在缓解速度矛盾上，还可以显著改善系统的性能。常见的缓冲区机制有单缓冲机制、双缓冲机制、缓冲池机制等。

设备处理程序即通常所说的设备驱动程序，它是 CPU 与 I/O 设备之间的通信程序。其工作过程如下：当设备驱动程序接收到上层软件发送来的 I/O 请求时，要先检查其合法性，然后查看设备是否空闲、设备的工作方式等信息，接着，按照要求的参数格式向设备控制器发送具体的 I/O 命令，指挥控制器启动设备按照顺序完成指定动作。为了保证通信，设备驱动程序还应能接收从控制器发来的中断请求，分析该中断请求的类型，接着启动处理该中断类型的相应中断处理程序，由其完成最终的处理过程。如果是具有通道的操作系统，设备处理程序中还要配备根据用户请求构造通道程序的功能。

虚拟设备功能是通过特殊的虚拟技术实现的，该技术可以将一台物理设备虚拟为多台逻辑设备，每个用户使用一台逻辑设备，即将独占的物理 I/O 设备交由多个用户共享使用。这种方法能够极大地提高 I/O 速度，改善设备利用率，对每个用户而言也感觉自身具有一台独享的物理设备，改善了用户请求的响应感受。

4. 文件管理功能

在现代操作系统中，程序运行所需的代码和数据量十分庞大，而内存空间有限且无法长期保存信息，因此这些资源通常以文件形式存储在磁盘、光盘等外部介质上，只有在程序运行需要时才调入内存。为了保证和核准用户可以正确使用这些资源，所有的操作系统都提供了文件管理机制。其主要功能就是管理外存上的静态文件，提供存取、共享和保护文件的手段，以方便用户使用，同时禁止无权限用户对他人资源的误访问或者有权限用户对资源的误操作。文件管理机制还要能有效地管理外存空闲区域，根据文件的大小为其分配和回收空闲区。为了满足用户对响应时间的要求，文件管理机制还应实现目录管理，以便快速地定位文件。

文件管理机制能有效保护文件安全，提高资源利用率，为用户提供快速检索和使用文件的手段，是操作系统不可或缺的组成部分。

5. 人机接口

为了更大程度地减少操作人员的次要工作、方便用户使用系统功能,现代操作系统无一例外地配置了用户界面,即所谓“用户与操作系统的接口”。该接口分为图形用户接口、命令接口和程序接口三类。

图形用户接口使用文字、图形和图像来形成文件,同时使用各种图标将操作系统的多种类型的文件直观形象地表示出来。使用时只需单双击鼠标左右键即可完成全部操作。这种接口使得用户可以简单快捷地完成工作,即使是刚接触计算机的人员也能使用。该接口的实现使得计算机在社会生活的多种领域得到广泛普及,计算机变得非常简单易用,甚至非计算机专业的人士也可以利用计算机的高速处理和运算能力加速本专业的工作流程。因此,在上世纪最后的十年间,图形界面已经成为所有主流操作系统的必备模块。

命令接口是操作系统提供给用户的另外一种直接或间接控制自身工作的途径。用户使用命令接口向自身工作发送命令,控制工作运行。常见的命令接口有联机命令接口和脱机命令接口两种。脱机接口为批处理用户使用,由一组作业控制语言组成。作业控制语言可用来定义作业说明书。由于当批处理作业运行时,用户不能直接与自身作业通信,只能由系统完成作业控制和管理,此时系统的控制和干预方法均来自于作业说明书。联机命令接口为联机用户使用,由一组键盘指令和命令解释程序组成。用户在使用联机命令接口时,需要利用键盘顺序输入多条指令。而命令解释程序每接收到一条指令后就对其进行解释并执行。命令接口的好处在于直接驱动和控制相关设备,能得到更高的执行效率。

程序接口是出现于应用程序中的接口,它用来保证用户程序能够获得操作系统服务,由一组系统调用组成。这些系统调用均是完成某些具体系统功能的子程序,对用户而言,这些系统调用表现为对应的库函数,当用户需要使用系统功能时只需像使用一般函数那样调用这些库函数即可。

1.1.4 操作系统的设计目标

现代操作系统的设计目标是有效性、方便性、开放性、可扩展性等特性。

其中有效性指的是操作系统应能有效地提高系统资源利用率和系统吞吐量。方便性指的是配置了操作系统后的计算机应该更容易使用。这两个性质是操作系统最重要的设计目标。早期由于硬件的昂贵,设计人员更关注的是有效性,即使得系统中的资源利用率尽可能高。但随着硬件价格的不断降低以及计算机在各领域的广泛使用,在当今主流的微型计算机系统中,尤其是个人用户计算机中,设计人员更关注的是方便性,以便非计算机专业人士也能正确地使用计算机。

开放性指的是操作系统应遵循世界标准规范,如开放系统互联 OSI 国际标准。这是因为,随着 Internet 的快速发展,计算机操作系统早已从传统封闭的单机环境变为开放的多机环境。为了使不同厂家生产的计算机和设备能够通过网络集成与共享,保证应用程序的可移植性和互操作性,操作系统必须提供统一的开放环境,遵循相同或相似的国际标准,