

- 由浅入深介绍了操作系统的基本原理
- 剖析典型操作系统Windows和Linux
- 内容安排合理，做到循序渐进
- 基本功能均配有实际案例，培养方案设计理念
- 习题形式多样且内容丰富，并附有答案
- 为任课教师免费提供PPT课件

操作系统教程

王伟◎主编
徐克奇 林捷◎副主编

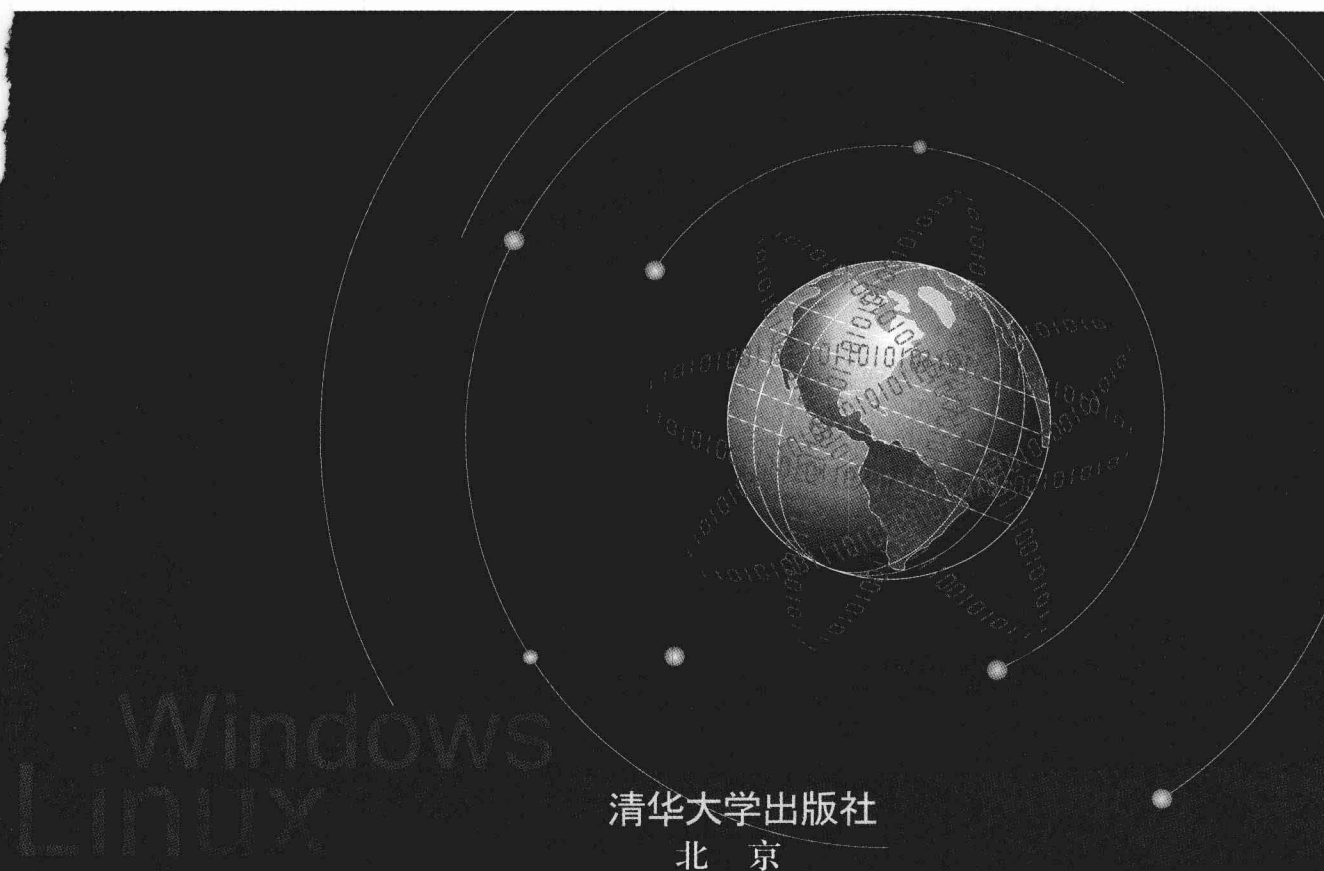
Windows
Linux



清华大学出版社

操作系统教程

武 伟◎主 编
徐克奇 林 捷◎副主编
李兴鹏 张成姝◎编 著



清华大学出版社
北 京

内 容 简 介

本书详细介绍了操作系统的基本原理,剖析了两个典型的操作系统 Linux 和 Windows。主要内容包括处理器管理、存储器管理、I/O 设备管理、文件系统、Linux 和 UNIX 剖析及 Windows 2000 剖析等。各章均配有习题,有助于读者领会和掌握已学过的相关知识。本书在编写过程中力求做到结合实际、突出应用、便于教学,同时注重内容的先进性。

本书可作为高等学校计算机科学与技术、软件工程、网络工程等专业的本科教材,也可作为非计算机专业的教学参考书。

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

操作系统教程/武伟主编. —北京:清华大学出版社,2010.8

ISBN 978-7-302-23192-9

I. ①操… II. ①武… 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(2010)第 124278 号

责任编辑:朱英彪

封面设计:张 岩

版式设计:魏 远

责任校对:王 云

责任印制:王秀菊

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:北京市清华园胶印厂

经 销:全国新华书店

开 本:185×260 印 张:18.25 字 数:419 千字

版 次:2010 年 8 月第 1 版 印 次:2010 年 8 月第 1 次印刷

印 数:1~4000

定 价:32.00 元

产品编号:035313-01

前 言

操作系统是计算机系统中最重要系统软件,它是围绕着如何提高计算机资源利用率和改善用户界面的友好性而形成、发展和不断成熟的。进入 21 世纪以来,我国的高等院校和 IT 界对操作系统的关心和重视达到了前所未有的程度,因为人们普遍认识到,计算机操作系统是整个 IT 领域中的重要基础,要构建现代化的、稳固的、可靠的信息技术大厦,必须掌握计算机操作系统原理。

本书针对培养技术型人才的特点,在注重操作系统原理的基础上引入了操作系统发展过程中的最新技术,以 Linux 和 Windows 操作系统为实例材料,剖析了现代操作系统所采用的最新技术。

本书是作者在多年教学实践和科学研究的基础上,参阅了大量国内外操作系统教材,编写的一本适用于计算机科学与技术、软件工程、网络工程专业的本科教科书。其编写思路及特点如下:

(1) 以最新的主流操作系统 Linux 和 Windows 为实例,从操作系统原理的角度对其做了详尽的介绍,并在讲授原理时注重理论联系实际。

(2) 根据作者的教学经验,对于难以理解的部分,均以实例引出,语言深入浅出,使学生能够从简单的实例入手,比较容易地掌握操作系统的内部工作原理。

(3) 本书配有大量经过精选的习题,以帮助读者检验和加深对内容的理解。

本书参考教学时数为 70~80 学时。要求先修课程为《数据结构》、《汇编语言》、《C 语言》和《计算机组成原理》。

本书的内容是按照理工科院校计算机科学与技术专业的教学大纲编写的。对于非计算机专业的本科专业及非本科的计算机专业,可适当删减内容。

全书共分 11 章。第 1、6 章由徐克奇编写,第 3 章由林捷编写,第 9、11 章由李兴鹏编写,第 10 章由张成姝编写,第 2、4、5、7、8 章由王伟编写。全书由王伟统一编排定稿。

上海师范大学胡金初教授对本书的初稿进行了认真的审阅,并提出了修改意见。在此,谨向胡金初老师表示诚挚的谢意。

由于时间仓促,加之作者水平所限,书中难免会有错误和不足之处,敬请各位老师和读者批评指正。同时也恳请各位老师和读者一旦发现错误,及时与编者联系,以便尽快更正,编者的电子邮件地址是 weiwu@sit.edu.cn。

编 者
2010 年 3 月于上海

目 录

第1章 引论	1	2.1.3 联机用户接口	27
1.1 计算机系统组成	1	2.1.4 脱机用户接口	29
1.1.1 计算机硬件系统结构	1	2.2 程序级接口	30
1.1.2 计算机软件系统	3	2.2.1 用户态和核心态	30
1.1.3 计算机系统结构	3	2.2.2 特权指令和访管指令	30
1.2 操作系统的作用和定义	3	2.2.3 系统调用	31
1.2.1 操作系统的作用	3	2.2.4 系统调用的使用和执行过程	32
1.2.2 操作系统的定义	5	本章小结	32
1.3 操作系统的发展过程	5	习题	33
1.4 操作系统的分类	7	第3章 进程的描述与控制	35
1.4.1 批处理系统	8	3.1 程序执行方式	35
1.4.2 分时操作系统	10	3.1.1 程序顺序执行	35
1.4.3 实时操作系统	11	3.1.2 程序并发执行	36
1.4.4 多模式操作系统	12	3.2 进程描述	39
1.4.5 微机操作系统	12	3.2.1 进程定义	39
1.4.6 网络操作系统	13	3.2.2 进程特性	39
1.4.7 分布式操作系统	15	3.2.3 进程与程序的区别	40
1.4.8 嵌入式操作系统	16	3.2.4 进程控制块	40
1.4.9 智能卡操作系统	17	3.3 进程状态	42
1.5 操作系统的功能和特性	17	3.3.1 进程执行	42
1.5.1 操作系统的功能	17	3.3.2 进程的基本状态	42
1.5.2 操作系统的特性	19	3.3.3 进程的挂起	44
1.6 操作系统的结构模型	20	3.4 进程控制	45
1.6.1 整体式模型	20	3.4.1 内核	45
1.6.2 层次式模型	21	3.4.2 微内核	46
1.6.3 客户/服务器系统模型	22	3.4.3 进程控制	47
1.6.4 面向对象的程序设计	23	3.5 线程	48
本章小结	23	3.5.1 线程引入	49
习题	24	3.5.2 线程定义	49
第2章 用户与操作系统的接口	26	3.5.3 线程的状态	50
2.1 作业控制级接口	26	3.5.4 线程和进程的比较	50
2.1.1 作业	26	3.5.5 线程分类	51
2.1.2 作业的类型	27	3.5.6 线程的模型	52

3.6 Linux 进程管理	53	4.7.2 管道通信系统	90
3.6.1 Linux 进程状态	53	4.7.3 消息传递系统	91
3.6.2 task_struct 结构	54	本章小结	94
3.6.3 Linux 进程的创建	55	习题	95
3.6.4 进程使用的虚拟内存	57		
3.6.5 Linux 的进程调度	58	第 5 章 处理器调度	98
3.7 Windows 的进程管理	58	5.1 三级调度的概念	98
3.7.1 Windows 的进程和线程模型总述	59	5.1.1 作业的状态及其转换	98
3.7.2 Windows 中进程的实现	59	5.1.2 调度的层次	100
3.7.3 Windows 中线程的实现	59	5.1.3 调度模型	100
3.7.4 Windows 的线程调度	60	5.1.4 作业和进程的关系	102
本章小结	64	5.2 作业调度	102
习题	65	5.2.1 作业调度的功能	103
第 4 章 进程通信	67	5.2.2 作业调度的目标与性能衡量	103
4.1 进程的同步与互斥	67	5.3 进程调度	105
4.1.1 进程合作	67	5.3.1 进程调度的功能	105
4.1.2 共享资源	68	5.3.2 进程调度的方式	106
4.1.3 与时间有关的错误	68	5.3.3 进程调度的时机	107
4.1.4 临界资源与临界区	70	5.4 常用的调度算法	107
4.1.5 同步机构设计准则	71	5.4.1 先来先服务调度算法	107
4.2 互斥的软件方法	71	5.4.2 短作业（进程）优先调度算法	109
4.3 硬件指令机制	75	5.4.3 时间片轮转调度算法	110
4.3.1 测试与设置技术	75	5.4.4 高优先权优先调度算法	112
4.3.2 TS 指令	75	5.4.5 最高响应比优先调度算法	113
4.3.3 利用 TS 实现进程互斥	75	5.4.6 多级队列调度算法	114
4.4 信号量机制	76	5.4.7 多级反馈队列调度算法	114
4.4.1 整型信号量	77	5.5 实例分析：UNIX 进程调度	115
4.4.2 结构型信号量	77	5.5.1 调度时机	115
4.4.3 AND 型信号量集	78	5.5.2 调度标记设置	116
4.5 用信号量机制实现互斥与同步	79	5.5.3 优先数计算	116
4.5.1 用信号量实现互斥	79	5.5.4 调度的实现	116
4.5.2 用信号量实现同步	79	本章小结	117
4.6 经典进程同步问题	81	习题	117
4.6.1 生产者—消费者问题	82		
4.6.2 哲学家就餐问题	83	第 6 章 死锁	120
4.6.3 读者—写者问题	85	6.1 死锁的基本概念	120
4.6.4 睡眠的理发师问题	88	6.1.1 死锁的定义	120
4.7 进程通信方式	89	6.1.2 死锁产生的原因	120
4.7.1 共享存储区系统	89	6.1.3 死锁的必要条件	122
		6.1.4 处理死锁的策略	123

6.2 死锁预防.....	123	8.4.1 Linux 的内存空间.....	178
6.3 死锁避免.....	125	8.4.2 Linux 的页表机制.....	178
6.3.1 安全状态和不安全状态.....	125	8.4.3 Linux 内存页的分配.....	178
6.3.2 利用银行家算法避免死锁.....	126	8.4.4 Linux 内存页的置换算法.....	179
6.4 死锁检测.....	129	8.5 Windows 存储管理.....	179
6.5 死锁解除.....	131	8.5.1 Windows 的虚拟地址空间.....	179
6.6 死锁综合处理.....	132	8.5.2 页面状态.....	180
本章小结.....	132	8.5.3 地址转换机制.....	180
习题.....	133	8.5.4 页面目录、页表和 PTE 的结构.....	182
第 7 章 实存储管理技术.....	135	8.5.5 页面调度策略与工作集管理.....	183
7.1 存储管理的基本概念.....	135	8.5.6 Windows 的存储保护.....	184
7.1.1 存储管理要解决的问题.....	135	本章小结.....	184
7.1.2 存储管理的分类.....	135	习题.....	184
7.1.3 地址重定位.....	137	第 9 章 设备管理.....	187
7.2 连续分配存储管理方式.....	139	9.1 I/O 设备管理的基本概念.....	187
7.2.1 单一连续分配方式.....	139	9.2 I/O 系统的组成.....	188
7.2.2 固定分区存储管理方式.....	139	9.2.1 I/O 设备概述.....	188
7.2.3 可变分区存储管理方式.....	141	9.2.2 设备控制器.....	189
7.3 离散分配存储管理方式.....	147	9.2.3 I/O 通道.....	190
7.3.1 分页存储管理方式.....	148	9.3 I/O 控制方式.....	191
7.3.2 分段存储管理方式.....	154	9.3.1 程序 I/O 方式.....	191
7.3.3 段页式存储管理方式.....	158	9.3.2 中断驱动 I/O 方式.....	191
本章小结.....	160	9.3.3 直接存储器存取方式.....	192
习题.....	161	9.3.4 I/O 通道方式.....	193
第 8 章 虚拟存储管理技术.....	164	9.4 缓冲管理.....	193
8.1 虚拟存储器的基本概念.....	164	9.4.1 缓冲的引入.....	193
8.1.1 局部性原理.....	164	9.4.2 缓冲区及其管理.....	194
8.1.2 虚拟存储器.....	165	9.5 设备分配.....	197
8.2 请求分页式存储管理方式.....	165	9.5.1 设备分配中的数据结构.....	197
8.2.1 请求分页式存储管理的基本概念.....	166	9.5.2 设备分配策略.....	198
8.2.2 页面分配策略.....	168	9.5.3 设备独立性.....	199
8.2.3 页面调入时机.....	169	9.5.4 独占设备分配方法.....	200
8.2.4 页面置换算法.....	169	9.5.5 SPOOLing 技术.....	202
8.2.5 请求分页系统的性能分析.....	173	9.6 I/O 软件的组成.....	203
8.3 请求分段式存储管理方式.....	175	9.6.1 中断处理程序.....	204
8.3.1 请求分段式存储管理的基本概念.....	175	9.6.2 设备驱动程序.....	206
8.3.2 分段共享与保护.....	177	9.6.3 与设备无关的系统软件.....	208
8.4 Linux 存储管理.....	178	9.6.4 用户空间的 I/O 软件.....	209

9.7 磁盘 I/O	210	10.4.6 目录查询技术.....	244
9.7.1 磁盘性能概述.....	210	10.5 文件存储空间的管理	245
9.7.2 数据的组织.....	210	10.5.1 空闲表法.....	245
9.7.3 磁盘访问时间.....	211	10.5.2 空闲链表法.....	246
9.7.4 磁盘调度算法.....	212	10.5.3 位示图.....	246
9.8 Linux 的设备管理	214	10.5.4 成组链接法.....	248
9.8.1 设备和设备特殊文件.....	214	10.6 文件的共享	249
9.8.2 设备驱动程序的动态安装.....	215	10.6.1 目录结构中的共享.....	249
9.8.3 设备驱动程序.....	215	10.6.2 打开文件结构中的共享.....	251
9.8.4 设备文件的存取权限和 系统安全	217	10.6.3 管道文件.....	252
9.9 Windows 的设备管理	218	10.7 文件系统的安全性	254
9.9.1 Windows 的 I/O 设计目标	218	10.7.1 文件的存取控制.....	254
9.9.2 Windows 的 I/O 结构	218	10.7.2 文件的转储和恢复.....	257
9.9.3 Windows 的设备驱动程序.....	219	10.8 Linux 文件系统.....	257
9.9.4 Windows 的 I/O 处理	222	10.9 Windows NTFS	259
本章小结	222	10.9.1 NTFS 的重要特征.....	259
习题	223	10.9.2 NTFS 的磁盘组织.....	260
第 10 章 文件系统	225	10.9.3 NTFS 的文件组织.....	260
10.1 文件和文件系统.....	225	10.9.4 NTFS 的可恢复性.....	261
10.1.1 文件.....	225	本章小结	262
10.1.2 文件系统.....	226	习题	263
10.1.3 文件的分类.....	227	第 11 章 操作系统的安全性	265
10.1.4 文件系统的功能和基本操作.....	228	11.1 操作系统安全性概述	265
10.2 文件的逻辑结构	229	11.1.1 操作系统安全性的内容.....	266
10.2.1 文件的逻辑结构.....	229	11.1.2 操作系统安全性的特性.....	267
10.2.2 文件的组织和存取.....	230	11.2 操作系统的安全管理与保护	267
10.3 外存分配.....	233	11.3 数据的安全管理与保护	271
10.3.1 文件的物理结构.....	233	11.3.1 文件的存取控制.....	271
10.3.2 连续分配.....	234	11.3.2 数据加密技术.....	271
10.3.3 链接分配.....	234	11.3.3 认证技术.....	275
10.3.4 索引分配.....	235	11.4 计算机病毒	277
10.3.5 文件分配表 FAT	237	11.4.1 计算机病毒概述.....	277
10.4 文件目录.....	240	11.4.2 计算机病毒的分类.....	278
10.4.1 文件控制块.....	240	11.4.3 常用反病毒技术.....	279
10.4.2 索引结点.....	241	11.4.4 未来计算机病毒的发展趋势.....	280
10.4.3 单级目录结构.....	242	本章小结	281
10.4.4 二级目录结构.....	242	习题	281
10.4.5 多级目录结构.....	243	参考文献	282

第 1 章 引 论

计算机系统包括硬件和软件两个部分，操作系统（Operating System，OS）是配置在计算机硬件上的第一层软件，可以扩充硬件功能、提供软件运行环境，实现了应用软件和硬件设备的连接，在计算机系统中占据了特别重要的地位。本章从操作系统的作用、发展历史、类型和功能等方面进行讨论。

1.1 计算机系统组成

一个完整的计算机系统是由硬件系统和软件系统两大部分组成的。硬件系统是指计算机的物理设备本身，如处理器（CPU）、存储器（Memory）、I/O 设备（I/O Device）等。软件系统是与数据处理系统的操作有关的计算机程序、过程、规则和相关文档资料的总称，如众所周知的 Windows、Linux、UNIX、Word、IE 等都属于软件的范畴。

1.1.1 计算机硬件系统结构

1. 大、中、小型机的硬件组织

大、中、小型机的硬件系统由中央处理器（CPU）、存储器、输入/输出处理器（通道）、I/O 设备组成，如图 1.1 所示。

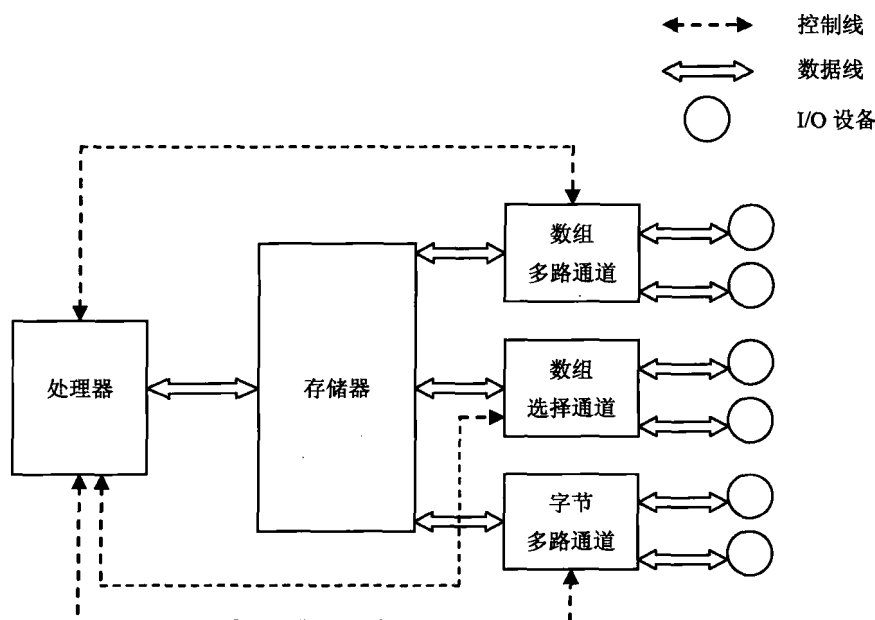


图 1.1 大、中、小型机的系统结构

由图 1.1 可以看出,这类计算机的中心是存储器。中央处理器和输入/输出处理器都与存储器相连,这些处理器执行的程序和数据都存放在存储器中,并从存储器中获取指令执行。中央处理器需要与输入/输出设备交换数据时,会命令输入/输出处理器进行管理和控制,数据传输的路线需经过存储器、输入/输出处理器,也就是说中央处理器不能直接与输入/输出设备交换数据,因为它们之间没有直接相连的数据线。

2. 微型机的硬件组织

微型计算机系统与一般计算机系统一样,由 3 个主要部分组成,即处理器、存储器和输入/输出设备,系统结构为总线型结构,如图 1.2 所示。

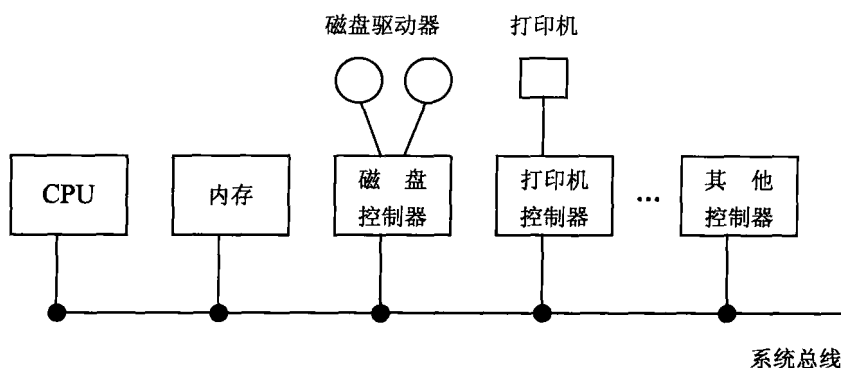


图 1.2 微型机的系统结构

3. 计算机系统的硬件资源

(1) 处理器 (CPU)

CPU 是计算机的“心脏”，由控制器和运算器两部分组成。控制器是计算机的控制中心，负责指令的执行，而运算器是计算机的数据加工和处理场所，用来完成算术和逻辑运算。

现代计算机系统大多是多道程序系统，计算机系统中 CPU 的数目往往少于正在运行的程序数目。只有一个处理器的计算机系统，称为单处理器系统；有多个处理器的系统，称为多处理器系统。

(2) 存储器

存储器分为内部存储器和外部存储器两种。内部存储器（也称内存或主存）是指 CPU 不需要经过接口电路便可直接存取其中数据的存储器。内部存储器中每个存储单元都有一个编号，称为存储地址，也称物理地址。CPU 每次访问内存时必须先指定存储单元地址。在多道程序系统中，由于内存中会存放多个程序，需要有硬件对其加以保护，将它们相互隔离。外存储器包括磁盘、光盘和 U 盘等，CPU 要经过 I/O 设备控制器才可访问。与内存相比，外存存取速度要慢，但其容量大，且每位存储量的价格低于内存。

(3) 输入/输出设备（外围设备或 I/O 设备）

I/O 设备一般由机械部分和控制电路部分组成。控制电路部分即接口电路，通常又称控制器。在微型机中，控制器做成印刷电路板卡即适配卡插在计算机内，经系统总线与 CPU 相连。适配卡与设备之间的接口通常采用标准接口（或事实上的工业标准），如 IDE、SCSI、USB、SATA 和 RS-232C 等。操作系统对设备的控制实际上是按命令和数据按一定顺序写入控制器

指定的寄存器中，包括从寄存器中读取设备的状态。

计算机系统应用于各种场合，配备各种 I/O 设备，这些设备的信息处理速度差异很大。I/O 设备大致分为两类，即块设备和字符设备。块设备是以数据块为单位存取的，数据块大小固定，且都有编号，可按块存取。字符设备则以字符为单位存取，没有块结构。

1.1.2 计算机软件系统

计算机软件通常分为两大类，即系统软件和应用软件。

系统软件又可分成 3 部分，即操作系统、语言处理系统和例行服务程序。

(1) 操作系统：操作系统是最重要的系统软件，它负责管理计算机中的各类资源，组织计算机的工作流程，为用户使用计算机提供接口，如 Windows、Linux、UNIX 等。

(2) 语言处理系统：包括各类语言的编译程序、解释程序和汇编程序，如 C 语言编译环境等。

(3) 例行服务程序：包括数据库管理系统、连接编辑程序、连接装配程序、诊断排错程序等，如 360 安全卫士、各类杀毒软件等。

应用软件是指为某一类应用需要而设计的程序，或用户为解决某个特定问题而编制的程序，如 Word、AutoCAD、Photoshop 及各种信息管理系统等。

1.1.3 计算机系统结构

一个完整的计算机系统由硬件系统和软件系统两部分构成，其结构形式如图 1.3 所示。

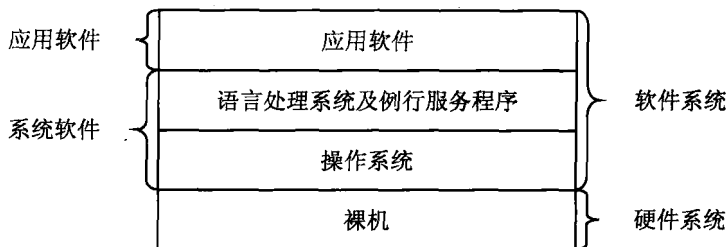


图 1.3 计算机系统的层次结构

1.2 操作系统的作用和定义

从不同的角度观察操作系统，其作用是不同的。从系统的角度观察，操作系统是资源管理者；而从用户的角度观察，操作系统则是用户与计算机硬件系统之间的接口。

1.2.1 操作系统的作用

1. 操作系统是系统软件

由 1.1.3 节可知，计算机系统是一个由硬件系统和软件系统构成的有层次结构的系统。硬件系统处于计算机系统的最底层，除实际设备外，机器语言也属于硬件设备。硬件部分通常称

为裸机。

用户直接编程来控制硬件是很麻烦的，而且容易出错。为此在硬件基础上加一层软件，用以控制和管理硬件，起到隐藏硬件复杂性的作用，呈现给用户经过“包装”的虚拟机，与裸机相比要抽象，可以使用户理解容易、使用方便。操作系统就是这层软件，操作系统是裸机的第一层扩充，是最重要的系统软件。

2. 操作系统是系统资源管理者

从资源管理的角度来看，计算机资源分为 4 大类，即处理器、存储器、I/O 设备和信息（文件）。前 3 类是硬件资源，信息是软件资源。引入操作系统的目的是为了合理地组织计算机的工作流程，管理和分配计算机系统硬件和软件资源，最大限度地提高计算机系统的利用率。其主要功能是对各类资源进行有效的管理。处理器管理解决处理器的分配和控制，存储器管理解决内存资源的分配、回收和保护，I/O 设备管理解决 I/O 设备的分配与回收，信息管理解决文件的存取、共享和保护。

作为资源管理者，操作系统在资源管理过程中要完成如下工作：

（1）监控资源状态

时刻维护系统资源的全局信息，掌握系统资源的种类、数量以及分配使用情况。

（2）分配资源

处理对资源的使用请求，协调请求中的冲突，确定资源分配算法。

（3）回收资源

用户程序对资源使用完后要释放资源，操作系统要及时回收资源，以便下次分配。

（4）保护资源

操作系统负责对资源进行有效的保护，防止资源被有意或无意地破坏。

3. 操作系统是用户与计算机硬件系统之间的接口

操作系统的一个重要任务是使计算机便于用户使用。操作系统处于用户与计算机硬件之间，用户通过操作系统来使用计算机，在操作系统的帮助下，方便、快捷、安全、可靠地操纵计算机硬件和运行自己的程序。图 1.4 表示了操作系统作为用户与计算机系统之间的接口的作用。

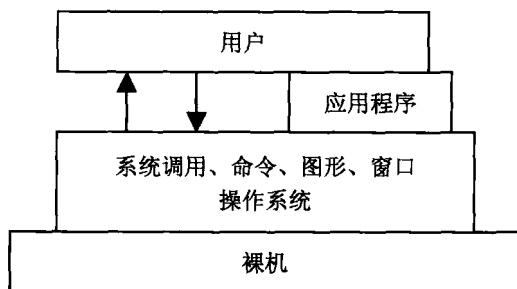


图 1.4 操作系统作为用户与计算机系统的接口

操作系统提供的用户接口有两类。

（1）作业级接口：操作系统提供一组联机命令接口，用户可以通过键盘输入有关的命令，获得操作系统的服务，并组织和控制自己的作业运行。

（2）程序级接口：操作系统提供一组系统调用，即操作系统中的某个功能模块，用户可

在应用程序中通过调用相应的系统实现与操作系统的通信，并取得它的服务。

需要指出的是，以 Windows 操作系统为代表的操作系统提供了图形、窗口的接口方式。这种方式实际上是把联机命令接口图形化，更加方便用户的使用，它仍然是一种联机命令接口。

1.2.2 操作系统的定义

综上所述，操作系统可以定义为：操作系统是直接控制和管理计算机系统硬件和软件资源，合理地组织计算机工作流程，便于用户使用的程序的集合。

从操作系统的定义可以看出，操作系统的目标有两点：首先，操作系统要方便用户使用。一个好的操作系统应提供给用户一个易于理解、简洁方便的用户界面，这里的“用户”既包括计算机系统的最终用户，又包括计算机系统的管理员，同时还包括编写应用程序的程序员。其次，操作系统应尽可能使系统中的各种资源得到最充分的利用。

可以看到，后面章节所讨论的各种实现技术、算法都围绕着这两个主要目标。

1.3 操作系统的发展过程

操作系统至今已有 50 多年的历史。回顾操作系统的发展历程，可以看到操作系统是随着计算机技术的发展和计算机的应用越来越广泛而发展的。20 世纪 50 年代中期出现了单道批处理系统，60 年代中期发展为多道批处理系统，与此同时也诞生了用于工业控制和武器控制的实时操作系统，80 年代开始到 21 世纪初，是微型机、多处理器和计算机网络高速发展的年代，也是微机操作系统、多处理机操作系统、网络操作系统和分布式操作系统大发展的年代。

1. 人工操作方式（1945—1955）

20 世纪 40 年代中期，美国科学家采用冯·诺依曼使用电子管成功地建造了第一台电子数字计算机。这个阶段一直延续到 50 年代中期，属于第一代计算机。这时的计算机由上万个电子管组成，运算速度仅为每秒数千次，但体积却十分庞大，且功耗非常高、价格昂贵，也没有操作系统。用户采用人工操作方式使用计算机，即由程序员事先将程序和数据写入纸带（卡片），装入纸带（卡片）输入机并启动，将程序和数据输入计算机，然后启动计算机运行。当一个程序运行完毕后，才能让下一个用户使用计算机。可见，这种方式不但用户独占全机，而且要 CPU 等待人工操作。

在计算机发展的早期，由于 CPU 的运算速度较慢，人机矛盾并不十分突出。但随着 CPU 运算速度的提高，这种矛盾日趋严重（例如，作业在一个每秒 1000 次的机器上运行，需要 30 分钟的时间，而手工装入和卸下作业等人工干预只需要 3 分钟。若机器速度提高 10 倍，则作业的运行时间缩短为 3 分钟，这使得一半的机时被浪费了），而且 CPU 与 I/O 设备之间速度不匹配的矛盾也更加突出。人工操作方式严重降低了计算机资源的利用率，为缓和此矛盾，急需一种方法减少人工操作的时间，操作系统应运而生。

2. 单道批处理系统（1955—1965）

20 世纪 50 年代，晶体管的发明使计算机硬件发生了革命性的变革，运算速度大幅度提高、

功耗减少、可靠性大为提高。但采用人工操作方式，人机矛盾和 CPU 与 I/O 设备速度不匹配的矛盾也更为突出。为了解决这些矛盾，出现了单道批处理系统。

单道批处理系统采用脱机输入/输出方式。所谓脱机输入/输出方式，是指输入/输出操作都是通过磁带进行的，操作员使用一台相对便宜的专门用于输入/输出的外围机将纸带（卡片）上的用户作业信息存到磁带（输入带）上，然后把磁带从外围机的磁带机上取下，并装到主机的磁带机上，再执行称为监督程序的软件，从磁带上读入第一个作业并运行，其输出写到另一盘磁带（输出带）上。一个作业结束后，监督程序自动读入下一个作业并运行。当一批作业完全结束后，操作员取下输入和输出磁带，将输入磁带换成下一批作业，把输出磁带拿到一台外围机上进行脱机输出。这样，监督程序管理着用户作业的运行，还控制磁带机的输入/输出操作。整个过程如图 1.5 所示。

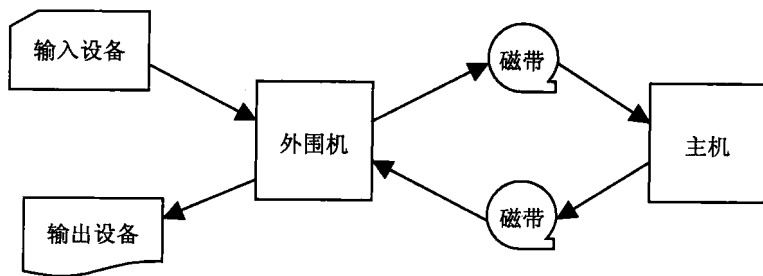


图 1.5 脱机输入/输出方式示意图

3. 多道程序系统（1965—1980）

早期的批处理系统仍是单道顺序处理作业，每次只有一个作业调入内存运行。这样可能会出现两种情况：当运行以计算为主的作业时，输入/输出量少，I/O 设备空闲时间多；而当运行以输入/输出为主的作业时，CPU 又有较多空闲。于是，多道程序设计思想应运而生。

到了 20 世纪 60 年代，计算机硬件有了新的进展，出现了通道和中断技术。通道是一种专用的 I/O 处理器，它能控制一台或多台 I/O 设备工作，执行 I/O 设备与内存之间的数据传输。一旦被启动就能独立于 CPU 运行，因此 CPU 与通道、CPU 与 I/O 设备可以并行工作。中断则是指当主机接到外部信号（如 I/O 设备操作完成信号）时，便立即停止正在执行的程序，转而处理中断事件的相应程序。处理完毕后，主机再回到原来程序被中断的位置继续执行。通道、中断技术的出现，为多道程序设计奠定了基础。

多道程序设计的主要思想是，在内存中同时存放若干道用户作业，这些作业交替地运行。当一个作业由于 I/O 操作未完成而暂时无法继续运行时，系统就把 CPU 切换到另一个作业，从而使另一个作业在系统中运行。因此，从宏观上看，若干个用户作业，或者说若干道程序是同时在系统中运行的。

多道程序方式可以使 CPU 与 I/O 设备并行工作。在实际的系统中，往往不止两道程序在系统中运行，而且有许多台设备可供用户作业使用，因此系统资源的利用率是很高的。

把批处理系统同多道程序系统相结合，就形成了多道批处理系统。这种系统从 20 世纪 60 年代初出现以来得到了迅速的发展，目前大、中型机还在使用。

第三代操作系统适于进行大型科学计算和繁忙的商务数据处理，但其本质仍然是批处理系统。在作业运行过程中，用户无法干预，许多用户十分怀念第一代计算机的联机工作方式：用

用户可以自己控制程序的运行, 调试程序十分方便。

这种需求导致了分时系统的出现, 它实际是多道程序的一种变种。在分时系统中, 一台计算机同时连接多个用户终端, 每个用户通过终端使用计算机, CPU 的时间分割成很小的时间段, 每个时间段称为一个时间片, 系统将 CPU 的时间片轮流分配给上机的各个用户。计算机内存中存放着正在上机的终端用户的程序, 它们轮流得到执行。由于时间片分割得很小, 每个用户感觉自己独占着计算机。分时操作系统是联机的多用户交互式的操作系统, 仍是当今大型计算机普遍使用的操作系统。

第三代操作系统的代表有 OS/360。在这个时期, 小型机曾盛极一时。著名的交互式的分时操作系统 UNIX 也是在这一时期出现的。

4. 现代操作系统(1980—现在)

20 世纪 80 年代以来, 随着大规模集成电路技术的发展, 微型机得到广泛应用, 工作站也逐步取代了小型机。Windows、Linux 和 UNIX 等现代操作系统成为了微机、服务器、工作站的主流操作系统。

1969 年, 第一个网络系统 ARPAnet 研制成功, 经过几十年的发展, Internet 已经深入人们生活的每一个角落。各个独立的计算机通过网络设备和线路连接起来, 实现了更大范围的通信和资源共享。网络上的计算机都运行本地的操作系统, 网络操作系统在原来操作系统的基础上增加了网络功能模块, 以实现各种网络应用和服务。常见的网络操作系统有 Windows 2005、Linux、UNIX 等。

在网络技术发展的基础上, 人们正在研制分布式计算机系统, 在用户眼里一个分布式系统像是一个传统的、单处理器的分时系统, 用户无须知道网络中资源的位置即可使用它们。遗憾的是, 到目前为止, 还没有一个成熟的分布式操作系统问世。

嵌入式操作系统被固化在嵌入式计算机的 ROM 中, 它的用户接口一般不提供操作命令, 而是通过系统调用命令向用户程序提供服务。嵌入式操作系统被广泛应用于电器设备的控制中。

1.4 操作系统的分类

当前计算机已经逐渐深入到人们生活的各个方面, 办公自动化、图像处理、工业设计、自动控制、科学计算、数据处理、网络浏览等都是其主要应用。在如此广泛的应用领域里, 人们对计算机的要求也各不相同, 因此对计算机操作系统的性能要求、使用方式也各不相同, 对操作系统的分类方法也很多。

按同时使用操作系统的用户数目, 可把操作系统分为单用户操作系统和多用户操作系统。

根据操作系统所依赖硬件的规模, 可分为大型机、中型机、小型机和微型机操作系统。

最常见的是按照操作系统所适用的环境进行分类。实际的操作系统只是适应某一特定环境的系统。在介绍操作系统历史时提到的操作系统的 3 种基本类型, 即批处理系统、分时系统和实时系统, 分别适应于不同的环境。随着计算机技术的发展, 又出现了网络操作系统、分布式操作系统、嵌入式操作系统和智能卡操作系统等。

1.4.1 批处理系统

在 20 世纪 50 年代中期出现了单道批处理系统，60 年代中期发展为多道批处理系统。批处理系统的基本特征是具有成批处理作业的能力，批处理系统的主要目标是提高系统的处理能力，即作业的吞吐量，同时也兼顾作业的周转时间。根据处理方式的不同，可将批处理系统分为单道批处理系统和多道批处理系统。

1. 单道批处理系统

单道批处理系统是最早的操作系统。它是为了解决人工操作严重降低计算机资源利用率的问题，即 CPU 等待人工操作和高速 CPU 与低速 I/O 设备间的矛盾而产生的操作系统。它采用脱机输入/输出技术，即利用一台外围机，脱离主机将低速输入设备（如卡片机、纸带机等）的数据输入到较高速、大容量的输入设备（如磁带、磁盘）上，在监督程序的控制下，根据卡片机读入的控制作业操作信息，先将一个作业读入内存并进行处理，当作业处理完毕后再读入下一个作业。计算机系统就这样自动地对一个一个作业进行处理，直至磁带（磁盘）上的所有作业全部完成。单道批处理系统运行示意图如图 1.6 所示。

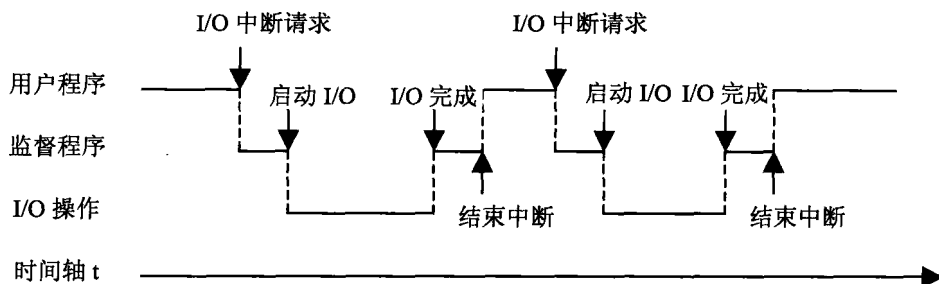


图 1.6 单道批处理系统运行示意图

从图 1.6 可以看出，由于内存中仅有一道程序，每当该程序发出 I/O 请求后，CPU 就空闲，必须在 I/O 操作完成后才能继续运行。尤其因为 I/O 设备的低速性，使得 CPU 的利用率显著降低。

这种批处理系统不能很好地利用系统资源，故现在已很少使用。

2. 多道批处理系统

多道批处理系统（Multiprogrammed Batch Processing System）是以脱机操作为标志的操作系统，特别适于处理运行时间比较长的作业。在该系统中，作业通过输入机输入到输入井中，形成后备作业队列。操作系统按一定算法从后备作业队列中选择若干个作业调入内存，这些作业并发执行，共享 CPU 和系统中的各种资源，并把计算结果输出到输出井中，形成输出结果队列，操作系统也按一定算法从输出结果队列中进行选择，通过输出机输出结果。多道批处理系统的工作原理如图 1.7 所示。

由于内存中有多个进程，所以当有一个进程请求 I/O 操作时，其他进程就可以使用 CPU 资源，从而提高了 CPU 的利用率。CPU 可以与 I/O 设备并行工作，也提高了 I/O 设备的利用率。图 1.8 是一个 4 个作业的多道程序运行示意图。

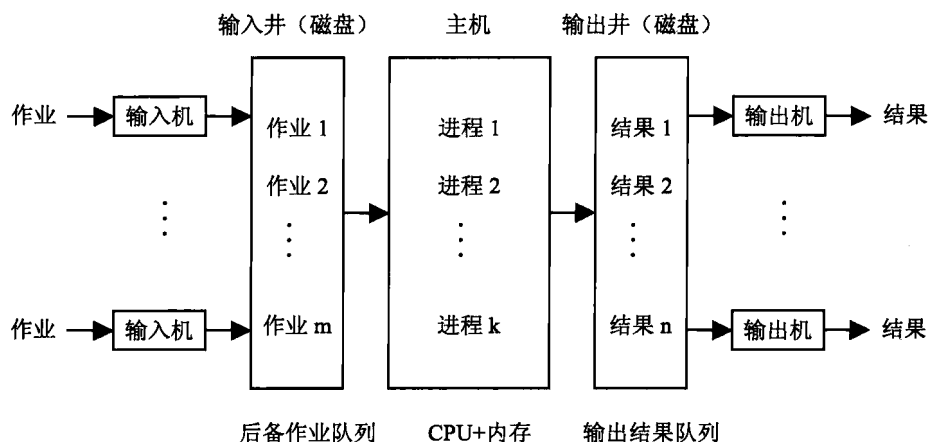


图 1.7 多道批处理系统的工作原理

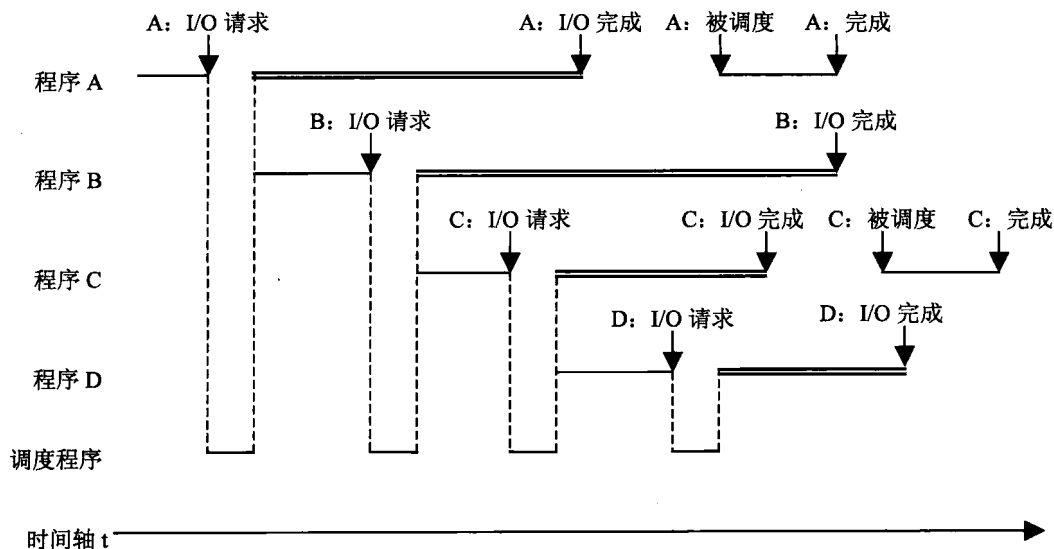


图 1.8 多道程序运行示意图

在批处理系统中采用多道程序设计技术，有以下优点：

- (1) 如果将 I/O 操作较多的作业与占用 CPU 时间较长的作业（如数值计算）搭配执行，可使 CPU 和 I/O 设备都得到充分利用。
- (2) 为了运行较大程序，系统内存空间配置较大，但大多数作业都为中、小程序，所以内存允许装入多道程序并允许它们并发执行，可提高内存空间利用率。
- (3) 多道批处理系统可使单位时间内处理作业的数量（即吞吐量）增加。

引入多道程序设计技术后，系统具有以下特征：

- (1) 内存中存放有多个尚未执行结束的程序，它们交替占用 CPU 执行。
- (2) 调入内存的一批作业完成的先后顺序与其调入内存的先后顺序之间没有对应关系。
- (3) 作业从提交系统开始直至完成，要经过两次调度。第一次是作业调度：作业要从外存的后备队列中被选中调入内存。第二次是进程调度：进入内存的作业中的一个能分配到 CPU，得以执行。