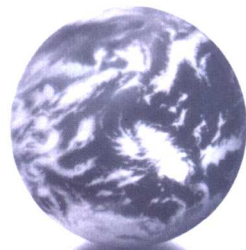
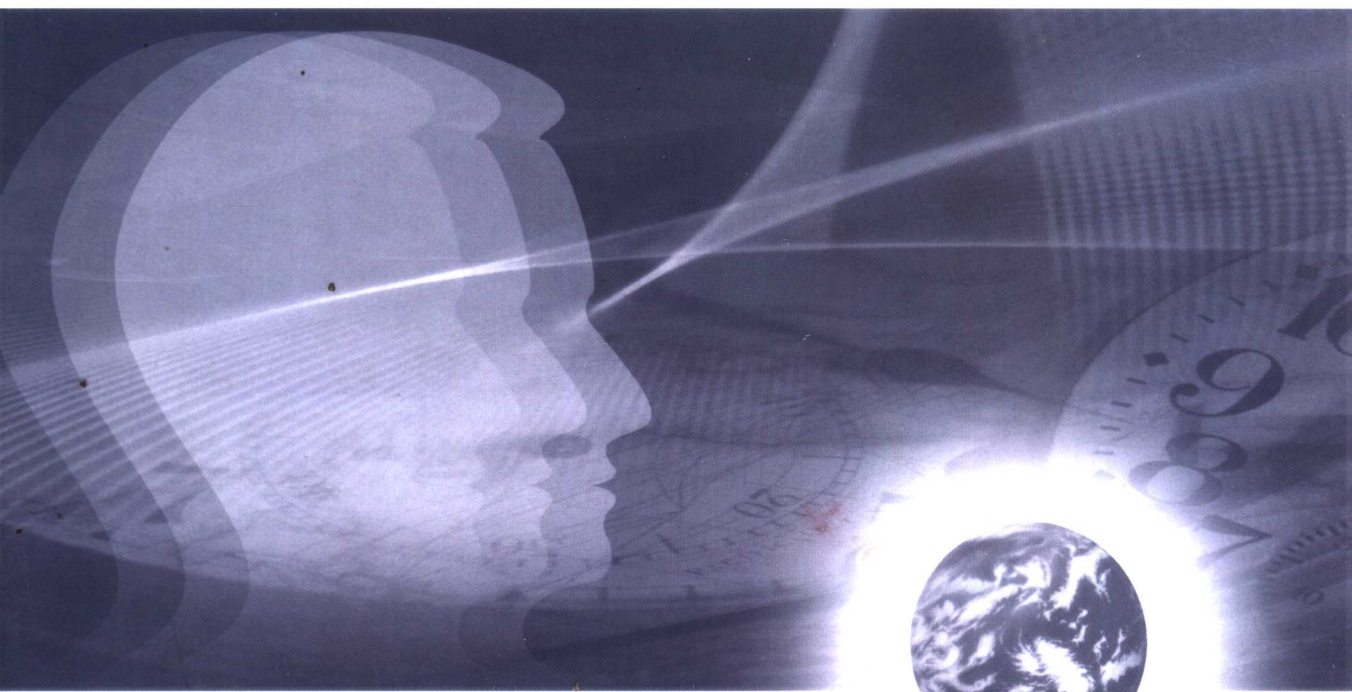


高等院校计算机科学与技术  
“十五”规划教材

# 操作系统原理 与实例分析



● 蒲晓蓉  
张伟利  
编著



 机械工业出版社  
CHINA MACHINE PRESS

高等院校计算机科学与技术“十五”规划教材

# 操作系统原理与实例分析

蒲晓蓉 张伟利 编著



机械工业出版社

操作系统是计算机系统的基础、核心组成部分，负责协调和管理计算机系统资源。本书基于操作系统的资源管理功能，将内容组织为概述、进程管理、存储管理、设备管理和文件管理系统 5 部分，深入浅出、系统全面地介绍了操作系统管理软、硬件资源的工作原理，并附有相应的实例分析。

本书可作为大学计算机专业或相关专业的本、专科学生的教材和参考书，也适合计算机专业相关人员使用。

### 图书在版编目 (CIP) 数据

操作系统原理与实例分析/蒲晓蓉, 张伟利编著. —北京: 机械工业出版社, 2004.6

(高等院校计算机科学与技术“十五”规划教材)

ISBN 7-111-14304-3

I. 操... II. ①蒲... ②张... III. 操作系统—高等学校—教材  
IV. TP316

中国版本图书馆 CIP 数据核字 (2004) 第 029112 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策 划: 胡毓坚

责任编辑: 王 颖

责任印制: 洪汉军

北京中兴印刷有限公司印刷·新华书店北京发行所发行

2004 年 6 月第 1 版·第 1 次印刷

787mm×1092mm 1/16·18.25 印张·451 千字

0001—5000 册

定价: 26.00 元

凡购本图书, 如有缺页、倒页、脱页, 由本社发行部调换

本社购书热线电话 (010) 68993821、88379646

封面无防伪标均为盗版

## 出版说明

信息技术高度普及的今天，具备一定层次的信息技术素养成为社会素质教育的一个重要目标，由此对高等院校的计算机专业教育提出了更高更新的要求。教育水平提高的关键是教学质量，那么对教学质量有直接影响的教材建设就成为了计算机专业教育的根本，为重中之重。

适逢高等院校计算机专业教育改革的关键时期，为配合相关的教材建设，机械工业出版社同全国在该领域内享誉盛名、具备雄厚师资和技术力量的高等院校，包括清华大学、上海交通大学、南京大学、电子科技大学、东南大学、西安电子科技大学、解放军理工大学、北京科技大学等重点名校，组织长期从事教学工作的骨干教师，集思广益，对当前高等院校的教学现状开展了广泛而深入的研讨，继而紧密结合当前技术发展需要并针对教学改革所提出的问题，精心编写了这套面向普通高等院校计算机专业的系列教材，并陆续出版。

本套教材内容覆盖了普通高等院校计算机专业学生的必修课程，另外还恰如其分地添加了一些选修课程，总体上分为基础、软件、硬件、网络和多媒体五大类。在编写过程中，对教学改革力度比较大、内容新颖以及各院校急需的并且适应社会经济发展的新教材，优先选择出版。

本套教材注重系统性、普及性和实用性，力求达到专业基础课教材概念清晰、深度合理的标准，并且注意与专业课教学的衔接；专业课教材覆盖面广、深浅适中，在体现相关领域最新发展的同时注重理论联系实际。全套教材体现了教育改革的最新思想，可作为高等院校计算机科学与技术专业的教学用书，同时也是培训班和自学使用的最佳教材。

机械工业出版社

# 前 言

操作系统是计算机系统中最基本、最重要的基础系统软件，是联系计算机硬件和其他软件的纽带。随着计算机技术的飞速发展，操作系统也经历了不断发展演变的过程。今天，Internet 技术，尤其是分布式技术的应用越来越广泛深入，其中涉及的主要技术——并发控制和多道程序设计技术是现代操作系统的主要工作原理。因此，对于计算机专业人员以及 IT 从业人员学习了解操作系统原理是非常必要的。

本书力求简单明了地剖析现代操作系统的工作原理，并辅以大量的实例进行说明。编者作为“操作系统原理”主讲教师，具有近 10 年的计算机专业课程讲授经验，清楚该课程的讲授和学习方法，以及学习的难点和重点，并积累了该领域大量的资料。这里有必要澄清一个问题，即很多第一次学习“操作系统原理”的人不明白其学习的目的和意义，因为能设计操作系统的专家毕竟有限。一般可以将与计算机系统打交道的人分为三类：计算机用户、应用程序员和操作系统设计员。由于计算机操作系统是计算机中最基础的系统软件，它为计算机用户提供了使用计算机的一个工作环境，专业术语称之为接口。一般的计算机用户必须至少熟悉一种计算机操作系统的使用方法，即了解某种操作系统的用户接口，比如能操作 Windows、UNIX 等操作系统，这是用户了解计算机操作系统的最低要求。有了这些基础以后，用户就可以更深一步使用操作系统，利用操作系统提供的程序接口编写应用程序，了解操作系统的实现原理，成为应用程序员。系统分析员或操作系统设计者应用系统深入地学习计算机操作系统的实现原理、了解操作系统设计思想、算法和理论，具备开发系统级程序的能力和系统分析能力。

明确了学习目的，需要再了解学习方法。“操作系统原理”主要介绍计算机操作系统的设计方法和实现技术，以及众多操作系统的设计精髓，而不针对某一特定的操作系统产品。可以说，其中涉及的内容是人们在操作系统工程实践方面的经验总结。所以，“操作系统原理”包含的内容呈现出这样几个主要特点：更新速度快、系统性差、概念抽象、原理性强，学习具有一定的难度。编者发现，采用“问答式启发学习”方法会有很好的效果。翻开书之前，首先问自己“为什么学习它？”，浏览目录了解全书概貌。本书内容有一个非常明显的特点：结构清晰，浅显易读。第 1 章概述，总体介绍操作系统概念、发展、功能、体系结构等知识，后续章节按照操作系统功能划分为进程管理、存储管理、设备管理及文件管理系统 4 部分。每一章都提出了本章涉及的主要问题，读者就这些问题问问自己是否能回答，并连同自己提出的问题在学习中寻求解答。例如，第 1 章提出并解答了为什么需要配置操作系统、如何选择适合自己或企业需要的操作系统等问题。

每一章的后面都配有实例分析，针对本章介绍的操作系统原理，分析当前流行操作系统的实例。力求使读者能利用所学的有关原理，理解、解释相应的操作系统实例。然而，大家知道，每一个操作系统实例的具体实现都是十分复杂的，且源于基本原理，又在基本原理的基础上进行了很大的发展和改进。本书注重操作系统通用原理的介绍，对实例分析部分仅仅是一个引入。有兴趣的读者还需要参考大量的文献，深入了解某一个操作系统的实现。

本书由蒲晓蓉老师规划和统稿，并负责撰写第 1、2、3、5 章的内容。张伟利老师组织

编写了第 4 章的内容，并协助完成了校稿工作。

在本书的编写过程中，得到了电子科技大学计算机学院相关领导和同事的支持，其中傅彦副教授给予了极大的支持和帮助。另外，编者的家人也理解并积极支持了编者的工作。在此，谨向所有提供过帮助和支持的朋友们表示深深的谢意。

由于编者经验和知识等方面的不足，加之时间仓促，书中难免存在错误和不足，敬请广大读者批评指正。

编 者

# 目 录

出版说明

前言

<b>第 1 章 操作系统概述</b> .....	1
1.1 计算机系统资源 .....	1
1.2 什么是操作系统 .....	2
1.3 操作系统的形成与发展 .....	4
1.3.1 推动操作系统发展的关键因素 .....	5
1.3.2 手工操作时期, 没有操作系统 .....	5
1.3.3 单道批处理系统, 早期的操作系统 .....	5
1.3.4 多道批处理系统, 现代意义的操作系统 .....	6
1.3.5 分时系统与实时系统 .....	8
1.3.6 操作系统的进一步发展 .....	9
1.4 操作系统的功能 .....	10
1.4.1 接口功能 .....	10
1.4.2 处理机管理 .....	11
1.4.3 存储器管理 .....	12
1.4.4 设备管理 .....	12
1.4.5 文件管理 .....	13
1.5 现代操作系统的特征及基本概念 .....	13
1.5.1 现代操作系统的特征 .....	14
1.5.2 基本概念 .....	15
1.6 操作系统分类 .....	17
1.7 现代主流操作系统简介 (Windows, UNIX, Linux) .....	18
1.7.1 Windows 操作系统 .....	18
1.7.2 UNIX 操作系统 .....	19
1.7.3 Linux 操作系统 .....	20
1.8 小结 .....	21
1.9 习题 .....	21
<b>第 2 章 进程管理</b> .....	22
2.1 进程的引入 .....	22
2.1.1 程序顺序执行与并发执行 .....	22
2.1.2 进程的概念与特征 .....	24
2.1.3 进程的结构 .....	25
2.2 进程的状态 .....	28
2.2.1 进程执行轨迹 .....	28
2.2.2 两状态进程模型 .....	30

2.2.3	五状态进程模型 .....	30
2.2.4	进程的挂起状态 .....	32
2.3	进程的控制 .....	34
2.3.1	执行模式 .....	34
2.3.2	操作系统内核 (Kernel) .....	35
2.3.3	进程控制 .....	36
2.4	进程调度 .....	38
2.4.1	调度的目标、原则和方式 .....	38
2.4.2	调度的类型 .....	40
2.4.3	进程调度算法 .....	41
2.4.4	实时系统与实时任务调度 .....	47
2.5	线程 .....	48
2.5.1	多线程 .....	48
2.5.2	进程与线程 .....	49
2.5.3	线程的类型 .....	50
2.6	进程互斥与同步 .....	51
2.6.1	并发控制 .....	51
2.6.2	互斥与同步的解决策略 .....	54
2.6.3	互斥与同步解决方法之一: 软件方法 .....	55
2.6.4	互斥与同步解决方法之二: 硬件方法 .....	60
2.6.5	互斥与同步解决方法之三: 信号量 (semaphores) 方法 .....	63
2.6.6	经典进程互斥与同步问题之一: 生产者/消费者问题 .....	65
2.6.7	经典进程互斥与同步问题之二: 读者/写者问题 .....	67
2.6.8	互斥与同步解决方法之四: 管程 .....	70
2.6.9	互斥与同步解决方法之五: 消息传递 .....	72
2.6.10	进程死锁 .....	77
2.6.11	经典进程互斥与同步问题之三: 哲学家进餐问题 .....	88
2.7	小结 .....	90
2.8	实例分析 .....	90
2.8.1	实例分析 1 Windows2000/XP 的进程和线程管理 .....	90
2.8.2	实例分析 2 UNIX SVR4 的进程管理 .....	97
2.8.3	实例分析 3 Linux 的进程和线程管理 .....	100
2.9	习题 .....	109
<b>第 3 章</b>	<b>存储管理 .....</b>	<b>112</b>
3.1	存储管理的任务 .....	112
3.1.1	存储分配 .....	112
3.1.2	地址映射 .....	113
3.1.3	存储保护 (Memory Protection) .....	115
3.1.4	存储共享 (Memory Sharing) .....	115



3.1.5	存储扩充	117
3.2	内存划分与分配技术	117
3.2.1	静态划分	117
3.2.2	动态划分与分配算法	119
3.2.3	伙伴系统 (Buddy System)	123
3.3	程序装入技术	125
3.3.1	可执行程序的生成步骤	125
3.3.2	可执行程序的装入	125
3.3.3	可执行程序的链接形成	126
3.4	简单存储管理技术	127
3.4.1	简单存储管理概述	127
3.4.2	连续存储管理	128
3.4.3	简单分页存储管理	129
3.4.4	简单分段存储管理	135
3.4.5	简单段页式存储管理	137
3.5	虚拟存储管理技术	139
3.5.1	虚拟存储技术概述	139
3.5.2	虚拟存储分页技术	141
3.5.3	虚拟存储分段技术	142
3.5.4	虚拟存储段页式技术	144
3.5.5	虚拟存储系统的软件策略	144
3.6	小结	153
3.7	实例分析	154
3.7.1	实例分析 1 Windows2000/XP 的存储器管理	154
3.7.2	实例分析 2 UNIX 的存储器管理	159
3.7.3	实例分析 3 Linux 存储器管理	163
3.8	习题	168
<b>第 4 章</b>	<b>设备管理</b>	<b>170</b>
4.1	设备管理概述	170
4.1.1	设备管理的主要功能	170
4.1.2	设备管理分层模型	171
4.1.3	设备管理工作流程	172
4.2	计算机 I/O 子系统的组成	173
4.2.1	I/O 系统的结构	173
4.2.2	设备的控制	176
4.2.3	I/O 控制方式	179
4.3	设备分类	180
4.3.1	输入/输出型设备与存储型设备	180
4.3.2	块型设备与字符型设备	180

4.3.3	独占型设备与共享型设备 .....	181
4.4	设备分配 .....	181
4.4.1	相关数据结构 .....	181
4.4.2	设备无关性 .....	182
4.4.3	设备分配算法 .....	182
4.4.4	独占型设备的分配 .....	183
4.4.5	共享型设备的分配 .....	183
4.5	I/O 缓冲技术 .....	185
4.5.1	缓冲技术的引入 .....	185
4.5.2	硬件缓冲和软件缓冲 .....	186
4.5.3	缓冲区的组织形式 .....	186
4.5.4	缓冲技术的实现 .....	187
4.6	虚拟设备 .....	189
4.6.1	虚拟设备的引入 .....	189
4.6.2	虚拟设备的实现 .....	189
4.6.3	SPOOLing 系统 .....	190
4.7	磁盘设备的管理 .....	191
4.7.1	磁盘设备的地位和作用 .....	191
4.7.2	磁盘设备的物理特性 .....	192
4.7.3	调整磁盘 I/O 性能 .....	194
4.7.4	磁盘容错技术 .....	197
4.7.5	RAID 技术 .....	199
4.8	小结 .....	201
4.9	实例分析 .....	202
4.9.1	实例分析 1 Windows NT/2000 的设备管理 .....	202
4.9.2	实例分析 2 UNIX 系统 V 的设备管理 .....	214
4.9.3	实例分析 3 Linux 的设备管理 .....	218
4.10	习题 .....	229
<b>第 5 章</b>	<b>文件管理系统 .....</b>	<b>231</b>
5.1	文件系统概述 .....	231
5.2	文件系统与数据库管理系统 .....	232
5.3	文件 .....	232
5.3.1	文件的概念 .....	232
5.3.2	文件的类型 .....	234
5.3.3	对文件的操作 .....	236
5.4	文件目录 .....	237
5.4.1	文件目录的内容 .....	237
5.4.2	对文件目录的操作 .....	238
5.4.3	目录结构 .....	239

5.5	文件的逻辑组织与访问 .....	242
5.5.1	有结构文件与文件系统 .....	242
5.5.2	堆文件 (pile) .....	243
5.5.3	顺序文件 (Sequential File) .....	243
5.5.4	索引顺序文件 (Indexed Sequential File) .....	244
5.5.5	索引文件 (Indexed File) .....	246
5.5.6	直接 (哈希) 文件 .....	246
5.6	文件的物理组织——存储空间的管理 .....	247
5.6.1	文件存储空间分配的有关问题 .....	247
5.6.2	文件存储空间的分配技术 .....	248
5.6.3	空闲空间的管理 .....	252
5.7	逻辑文件与物理数据块之间的转换 .....	254
5.7.1	字节流、记录与数据块之间的转换 .....	254
5.7.2	记录如何组成数据块 .....	255
5.8	文件共享 .....	257
5.8.1	文件共享的控制 .....	257
5.8.2	文件共享的实现 .....	258
5.9	文件保护与安全 .....	261
5.9.1	文件保护 .....	261
5.9.2	安全性管理 .....	261
5.10	文件系统的可靠性 .....	263
5.10.1	系统备份 .....	263
5.10.2	文件系统数据的一致性 .....	264
5.11	小结 .....	266
5.12	习题 .....	267
<b>附录</b>	.....	268
附录 A	目前几种流行的文件系统简介 .....	268
附录 B	磁盘文件操作流程 .....	272
附录 C	文件目录实例 .....	275
<b>参考文献</b>	.....	281

# 第1章 操作系统概述

初次学习“计算机操作系统原理”，很多人难免会有很多问题，诸如为什么需要专门介绍操作系统？操作系统究竟是什么？学习“操作系统原理”需要学习哪些内容？从软件设计角度看，应该如何设计操作系统？当今有哪些流行的操作系统，各自有些什么特点？作为概述，本章旨在让读者初步认识操作系统（Operating System,简称 OS），了解操作系统的发展过程，总体认识操作系统的功能、服务和特征，区别多种操作系统类型，明确将要学习的内容和本教材的布局结构，以便读者能从总体到细节，带着问题进行学习。

## 1.1 计算机系统资源

要学习操作系统原理，首先需要了解计算机的结构以及操作系统在计算机系统中的作用和地位。显然，计算机不同于电视机，仅由一台显示器和一个装满电子元件的机箱构成的计算机加电以后是无法工作的，必须安装一种称之为操作系统的软件才可以运行。当然，这时的计算机还无法完成如数据库管理、3D 绘画或收发电子邮件等工作，还必须再根据功能需要安装相应的应用程序。操作系统和其他的应用程序统称为计算机软件，那些看得见、摸得着的部分称为计算机硬件。因此，计算机系统资源包括硬件资源和软件资源两大类，二者缺一不可。硬件是计算机的基础，是软件运行的平台，但仅有硬件的计算机还不能称为完备的计算机，只有将硬件与软件结合起来才具有强大的功能。图 1-1 为计算机系统资源的结构示意图。最底层是计算机硬件。计算机的硬件系统由处理器、内存、输入/输出设备和系统总线 4 部分组成。其中，处理器控制计算机的操作，执行数据处理功能。对于单处理机系统，它通常称为中央处理器（处理机）。内存用于存储数据和程序，由于内存中的数据是暂存的，常称为随机存储器（RAM）或主存储器。输入/输出设备包括外存（磁盘、磁带、光盘等）、键盘、显示器、打印机、绘图仪以及通过 I/O 接口与外界通信的设备。处理机和内存是计算机运行的核心部件，二者合称为主机。系统总线为处理器、内存和输入/输出设备间的通信提供一些结构和机制。

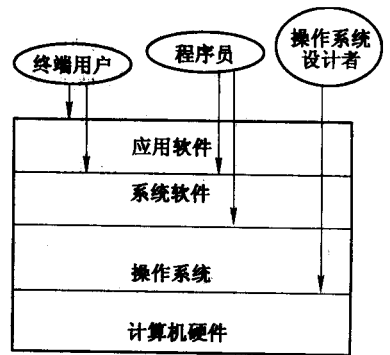


图 1-1 计算机系统结构

硬件之上是操作系统。操作系统是计算机必须配置的最基本的系统软件，是其他系统软件及应用程序的基础。其主要功能之一就是隐藏底层硬件具体操作的复杂细节，为用户提供方便使用计算机的环境及程序设计和执行环境。例如，用户可以直观地向计算机发出“保存文件（Save）”命令，而不必关心“磁头移动、查找空闲磁盘块、分配磁盘空间”等操作细节。这一切繁琐的操作细节全由操作系统控制相应部件完成，使用户轻松、方便地使用计算机。

在操作系统之上是其他系统软件，包括命令解释器（shell）、窗口（Windows）系统、编译器、编辑器及其他某些独立于应用的程序。它们不是操作系统的组成部分。操作系统专指特权保护模式下运行的软件，它受硬件保护，且不允许用户修改。该运行模式又称为核心态（kernel mode）或管态（supervisor mode）。编译器、编辑器及应用程序等运行在较低保护模式下，该模式称为用户态（user mode）。如果用户不喜欢某一个编译器，可以自己重写一个，但却不能编写一个磁盘中断处理程序，因为这是操作系统的一部分，而且硬件阻止用户对它进行修改。

系统软件之上是应用软件，这些软件可以是购买的产品，或者由用户自行开发的，它们用来解决特定的问题，如字处理、表格处理、工程计算、数据库管理、电子邮件收发或者电子游戏等。

## 1.2 什么是操作系统

从前一节的介绍不难发现，操作系统在计算机系统资源中的位置和作用。它是计算机硬件和其他软件及计算机用户之间的联系纽带。如果没有操作系统，用户几乎是无法使用计算机的。那么，操作系统究竟是什么？这是很难用一句话概括的问题。不同的人在不同的时期对操作系统会有不同的认识。对于现代操作系统，人们常用以下四种观点来描述：用户环境的观点、虚拟机的观点、资源管理的观点和作业组织的观点。

### 1. 用户环境的观点

如图 1-1 所示，计算机的用户大体可以分为终端用户、程序员和系统设计者三类。除了操作系统设计者需要直接面对计算机硬件以外，一般用户包括终端用户和程序员都工作在操作系统之上。用户环境观点认为操作系统是计算机用户使用计算机系统的接口（Interface，也称为界面），它为计算机用户提供了方便的工作环境。用户利用操作系统为他们提供的各种服务、功能接口及系统调用等来使用计算机系统及其各类系统资源，而不必知道各种系统资源的细节和控制过程。操作系统为计算机用户提供了两种接口，即用户接口（User Interface，也称为命令接口）和程序接口（即操作系统提供的系统调用的集合，也称为应用编程接口，Application Programming Interface，简称 API）。终端用户通过鼠标、键盘等使用操作系统提供的用户接口。程序员利用操作系统提供的程序接口设计应用程序，这一类程序员常称为系统程序员，他们基于操作系统进行程序设计，如基于 UNIX、Windows 或 Linux 程序设计等。还有一类程序员称为应用程序员，他们利用操作系统之上的系统工具设计应用程序，通过系统工具获得操作系统的服务。用户通常以使用者或程序员的角度评判一个操作系统，评价标准一般包括使用方便性、稳定性、安全性、性价比高低等，这些也是推动操作系统发展的关键性因素。

用户可以通过这样几种方式使用操作系统。其一，直接使用键盘命令或 Shell 命令语言；其二，利用鼠标器点击窗口中的图标，以执行相应的应用程序，如 Windows 操作系统的图形用户接口（GUI）；其三，调用操作系统内部功能模块，即系统调用接口。这些接口和各种应用程序为用户开发和运行应用软件提供了便利的环境和手段。

操作系统的发展过程也是用户界面不断改进的过程。人机界面问题一直都是用户和系统设计者十分关注的问题。操作系统用户界面已经从字符界面、菜单界面、窗口界面发展到目

前的图形界面和多媒体界面。随着计算机应用的不断深入，简单、友好、使用方便已成为操作系统界面的发展方向。

## 2. 虚拟机的观点 (Virtual Machine)

虚拟机的观点认为操作系统是建立在计算机硬件平台上的虚拟机器，它为应用软件提供了许多比计算机硬件功能更强或计算机硬件所没有的功能。虚拟机的观点是基于程序设计人员的观点，也称为扩展机器的观点。安装有操作系统的计算机极大地扩展了原计算机的功能，把用户面对的一个包含有各种硬件部件的计算机系统的操作和使用由复杂变得简单，从低级操作上升为高级操作，把基本功能扩展为多种功能。

操作系统在虚拟机中充当管理员和协调员的角色，管理计算机的处理器、存储器、时钟、磁盘、网络接口以及打印机等硬件设备，以及计算机中文件的存储、读 / 写等，并协调多任务、多进程之间的运行。设想在一台计算机上运行的三个程序同时试图在一台打印机上输出计算结果，如果控制不当，那么可能头几行是程序 1 的输出数据，接下来几行是程序 2 的输出数据，然后又是程序 3 的输出数据等，最终打印结果显然不是用户所需要的。为了避免这种混乱，操作系统可以将打印输出首先缓冲到磁盘上，然后逐个打印每个输出结果。当前一个打印任务完成以后，操作系统将启动下一个打印任务，将那些暂存在磁盘上的输出结果送到打印机打印。同时，其他程序可以继续运行，产生新的输出结果，而这些程序并不知道它们的输出有没有立即送至打印机。

操作系统虚拟机是逻辑上的概念，它不是一台单独的机器，而是与计算机硬件紧密结合在一起，在用户面前呈现的是一台比裸机功能丰富、操作简单的高级计算机。一台硬件配置完全相同的计算机，可以配置不同的操作系统，用于不同的环境。例如，一台硬件配置相同的个人计算机可以配置 Windows 操作系统及相关应用程序，用于文字处理或管理数据库；也可以配置实时操作系统，用于工业控制；或者配置多用户分时操作系统如 UNIX，使多个人可以同时使用该计算机；还可以配置网络操作系统，管理网络资源等。可见，操作系统虚拟机不仅可以扩充系统功能，甚至还能“扩充”计算机数量，多个用户同时使用一台计算机，每个用户都感觉自己独占该计算机。后一种操作系统即是本书介绍的重点，内容涉及如何实现多个任务同时执行以及实现这类操作系统时所遇到的各类问题及解决技术。

## 3. 资源管理的观点

资源管理观点认为操作系统是计算机系统中各类资源的管理者，它负责分配、回收以及控制系统中的各种软、硬件资源。资源管理的观点是目前对操作系统描述的主要观点。本书也是按照资源管理的观点介绍操作系统的原理，了解操作系统如何管理系统资源。

当一台计算机（或网络）有多个用户时，由于用户间可能相互影响，所以管理和保护存储器、I/O 设备以及其他设备的需求随之增加。而且用户往往不仅需要共享硬件，还要共享信息（文件、数据库等）。此时，操作系统的首要任务是跟踪资源的使用状况、满足资源请求、提高资源利用率，以及协调各程序和用户对资源的使用冲突。资源管理的主要功能包括：

- 1) 监视资源。监视系统资源的使用情况，包括系统中各类资源的总数，资源的分配状态（已分配及未分配），资源的使用状态（正在被使用或空闲），谁在使用，统计资源的使用历史等。

- 2) 分配资源。处理对资源的使用请求，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，根据资源分配的条件、原则和环境，决定把资源分给谁，分配多少资源，何时分配资源（立即分配，还是暂缓分配）等。同时，记录资源分配情况，更新相应的数据结构和表格的内容。

3) 回收资源。当使用者不需要某资源时，系统对该资源进行处理，如果是可重复使用资源，则进行资源回收，更新相应的数据结构和表格的内容，以备再次使用。

4) 保护资源。系统需要保护资源，防止它们被有意无意地破坏。防止非授权的操作，或非法用户的访问。例如，防止用户非法修改系统数据。在多用户系统中，要保护每个用户的数据及程序，防止彼此间的干扰和冲突。允许多个用户使用共享资源，并对共享资源进行保护。

#### 4. 作业组织的观点

现在很少有计算机用户知道操作系统中作业（JOB）这个概念，然而它却曾经被广泛地使用。随着个人计算机的出现和普及，作业逐渐被新的名词取代。目前，在巨型机和大型服务器上，作业的概念仍然存在。在这样的服务器上（称为主机）安装了能支持多个用户程序运行的操作系统，用户通过与主机相联的前端机（一般为普通工作站或 PC 机）以批文件方式提交作业，请求主机逐个运行。主机操作系统负责组织、协调各个作业的运行，报告执行结果或错误信息。专门的作业控制软件将作业批量提交给主机，有效减少了作业间切换的人工干预，大大提高了系统的工作效率。这种工作方式有利于有效利用造价高且性能强大的主机资源。

由于主机系统中存在多道程序、程序间并发执行和共享资源，操作系统必须对用户作业和任务进行管理，组织并控制系统工作的流程，建立相关的数据结构、操作模式和管理程序。随着操作系统的发展，根据系统工作的层次和并发控制程度，在操作系统中引入了任务、进程、线程及对象等概念，操作系统需要对它们进行组织和管理，我们将在后继章节详细讨论这些概念及其原理。

综上所述，操作系统对计算机用户而言是一个应用环境，一个工作平台，是人与机器交互的界面。对系统设计者来说是一种具有强大功能的系统资源管理器，一种包含有各种数据结构和算法的集成式软件系统。据此，我们可以把操作系统定义为：操作系统是计算机系统中的一个系统软件，管理和控制计算机系统中的硬件和软件资源，合理地组织计算机的工作流程，以便有效利用这些资源为用户提供一个功能强、使用方便的工作环境，从而在计算机与用户之间起到接口的作用。

### 1.3 操作系统的形成与发展

世界上的第一台计算机有操作系统吗？今天的操作系统是如何演变来的？是什么因素促进了操作系统的发展？为何现在出现了多种不同类型的操作系统？为何需要不断升级操作系统版本？操作系统还将如何发展？本节将回答这一系列的问题。

操作系统的形成与发展经历了漫长的时期。自 20 世纪 50 年代中期第一个简单批处理操作系统诞生以来，随着计算机硬件的发展，操作系统得到了迅猛的发展。20 世纪 60 年代中期产生了多道程序批处理系统，不久又出现了基于多道程序的分时系统。20 世纪 80 年代因为微机和计算机局域网的迅速发展，形成并发展了微机操作系统和网络操作系统。

### 1.3.1 推动操作系统发展的关键因素

自从第一代操作系统产生以来，操作系统的发展和进步是有目共睹的。是什么力量推动着操作系统的发展呢？这个问题可以归结到以下几个方面。

#### 1. 计算机硬件升级和新硬件的出现

计算机硬件是操作系统运行的基础，硬件的发展推动着操作系统的发展。例如，字符界面终端发展到图形用户终端，后者允许用户通过屏幕上的窗口同时查看多个应用程序，这就要求操作系统提供更高级的支持。存储管理方面，当分页（Paging）硬件机制出现以后，操作系统也需要支持分页管理技术，从而解决了分区管理中存在的问题（后面的存储管理部分将有介绍）。

#### 2. 提供新的服务，方便使用

推动操作系统发展的关键因素之一是为了满足用户不断增长的要求。方便使用是用户最广泛的需求，从批处理系统到分时系统，人机交互环境得到了极大的改善，用户使用计算机更加直接、方便。个人计算机操作系统如 Windows 系列增加了“即插即用”功能，同样是为了满足用户对多种计算机外设的使用需要。

#### 3. 提高计算机资源利用效率

在计算机发展的初期，计算机硬件资源特别昂贵，如何有效利用计算机的各种资源是研究和改进操作系统的一个非常重要的目标。所以，出现了批处理操作系统和多道程序设计技术。

#### 4. 更正软件错误

任何一种软件包括操作系统都不可避免地存在一些错误和漏洞，在使用一段时间以后发现了错误，就需要进行更正。所以，操作系统公司常常需要发布“补丁”程序或升级操作系统，甚至开发新的操作系统。

#### 5. 计算机体系结构的发展

计算机体系结构的发展也是推动操作系统发展的一个关键因素。单处理机操作系统发展到多处理机操作系统，是由于计算机体系结构由单处理机系统发展到了多处理机系统。当分布式系统和计算机网络出现以后，相应地又出现了分布式操作系统和网络操作系统。

### 1.3.2 手工操作时期，没有操作系统

早期的电子数字计算机是由成千上万个电子管和许多开关装置组成的庞然大物。这种计算机非常庞大，占地几个房间，耗能极其严重，但运算速度极慢，且价格昂贵，然而，它具有程序的概念。用户在这种计算机上操作和编程完全由手工进行，且编程只能用机器语言（二进制代码）。采用接插板或开关板控制计算机操作，没有显示设备，由氖灯或数码管显示。

直到 20 世纪 50 年代初期，卡片穿孔成为程序编制和记录的方法，不再使用插板，而形成一种可以“阅读”的程序。但计算过程仍然没有改变。程序员同时也是操作员在上机期间独占整台计算机及其他相关设备，效率非常低。可见，手工操作时期，没有操作系统，计算机资源利用率很低，工作效率也很低，而且计算机的使用非常复杂困难。

### 1.3.3 单道批处理系统，早期的操作系统

20 世纪 50 年代晶体管的诞生使计算机产生了一次革命，晶体管计算机已经很可靠了，并能批量生产。运行一个作业（一个或一组程序）时，程序员首先将命令、程序和数据用汇



编语言或 FORTRAN 语言写在纸上，然后用穿孔机制成卡片，最后将这些卡片交给操作员，作业卡片如图 1-2 所示。计算机运行完当前任务后，其计算结果从打印机上输出，操作员从打印机上取得运算结果，并送交程序员，然后，操作员再从卡片上读入另一个任务。如果需要 FORTRAN 编译器，操作员还需要先将编译器读入计算机内。这种手工切换方式使计算机长时间处于闲置等待状态，严重浪费了机器资源。

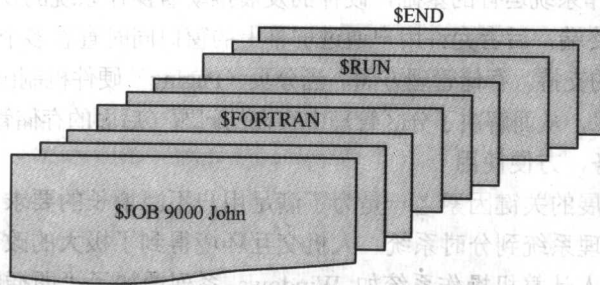


图 1-2 作业卡片示意图

为了有效利用当时价格昂贵的计算机，人们为计算机设计了批处理程序（又称为监督程序，或管理程序），用它管理用户提交的各种应用程序的运行。其基本操作步骤为：首先收集一批作业（卡），然后用一台专用计算机，称为输入/输出计算机，将作业逐个读到磁带上保存起来，再将磁带装到磁带机上。然后，由批处理程序将磁带上的第一个作业读入计算机，运算结束后将结果输出到输出磁带上，该计算机不负责打印输出。每当一个作业运行结束，批处理程序自动从输入磁带读入下一个作业并运行。当收集的这一批作业全部执行结束后，操作员取下输入磁带和输出磁带，用输入磁带录入下一批作业，将输出磁带送到专用输出计算机，进行脱机打印。

这种批处理控制程序解决了作业间的自动转接问题，减少了机器时间的浪费，但它还未真正实现对作业的控制和管理，存在较多的问题和不足。其一，不管作业大小，一旦它占用处理机开始执行，则它必须一直占据处理机，直到运行完毕。其二，资源利用率低。作业运行过程中通常都有一些计算和一些输入/输出操作，计算时，I/O 设备空闲；输入/输出数据时，处理机空闲。内存仅装入一个作业，也会浪费很多内存空间。其三，对短作业不公平，因为它们等待执行的时间可能远远超过它们实际执行的时间。其四，交互性差。作业由批处理程序控制运行，用户无法实时控制。如果运行期间出现故障，也只能停下来，重新运行。人们将这一时期采用批处理程序控制的计算机系统称为批处理系统，把这种批处理控制程序称为早期操作系统。

### 1.3.4 多道批处理系统，现代意义的操作系统

在单道批处理系统中，任意时刻只允许一道作业在内存中运行。由于作业运行期间可能需要一段时间计算，一段时间进行输入/输出，而相对于处理机的计算速度，输入/输出的速度是非常慢的。这就意味着在一个作业的运行过程中，处理机可能会长时间空闲。为了提高系统资源利用率和系统吞吐量，在 20 世纪 60 年代中期出现了多道程序设计技术，由此形成了多道批处理系统。