



内含光盘

操作系统原理

UNIX 篇

李林英 贺敏伟 编著
张昆藏 审

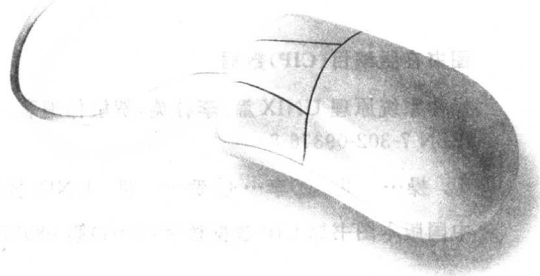


清华大学出版社

TP316.81
70D

操作系统原理 UNIX 篇

李林英 贺敏伟 编著
张昆藏 审



北方工业大学图书馆



00564768

RJS120/08

清华大学出版社
北京

内 容 简 介

本书是《操作系统原理 DOS 篇》(第二版)的姊妹篇。书中以原理为主,典型实例以 UNIX 为背景,把 MS-DOS 有关内容浓缩到最后一章里,讲述操作系统的基本概念、原理及实现技术。

全书共 9 章:分别为操作系统引论;UNIX 系统组成与功能服务;存储管理;进程管理;设备管理;文件管理;作业管理;UNIX 系统网络应用接口;微机操作系统功能的实现。各章结合原理,相应地介绍了 UNIX 操作系统的内核、网络与 Shell(命令解释器),使读者能比较全面地学习原理和认识 UNIX。本书带有配书光盘。同时配套出版《操作系统原理 UNIX 篇习题与实验指导》,该书也带有配书光盘。

本书读者对象是大学本科计算机专业的学生,本书配书光盘是操作系统多媒体教学光盘,有多种版本的电子教案(72 学时和 54 学时)、网页教案、点听系统、OS_Web 动画集、个人主页方式教学资源文件夹(为适应信息时代的全新教学方式提供了可利用的教学资源),课外阅读文件夹(其中提供了部分学生用 Flash 学习 OS 的作品),内容丰富多彩,生动有趣,可供教师选择组合。本书适应面宽,可作为本科、专科、成教、高职及各种教学需求的培训教材。对从事微机应用开发的广大科技人员,也是具有很强实用性、理论性的参考书,且适合自学。

版权所有,翻印必究。举报电话:010-62782989 13901104297 13801310933

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。

本书防伪标签采用清华大学核研院专有核径迹膜防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

操作系统原理 UNIX 篇/李林英,贺敏伟编著. —北京:清华大学出版社,2004.11

ISBN 7-302-09376-8

I. 操… II. ①李… ②贺… III. UNIX 操作系统—教材 IV. TP316.81

中国版本图书馆 CIP 数据核字(2004)第 089274 号

出 版 者:清华大学出版社

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

社 总 机:010-62770175

组稿编辑:李幼哲

文稿编辑:赵彤伟

版式设计:肖 米

印 刷 者:北京鑫丰华彩印有限公司

装 订 者:三河市金元装订厂

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:22.25 字数:521 千字

版 次:2004 年 11 月第 1 版 2004 年 11 月第 1 次印刷

书 号:ISBN 7-302-09376-8/TP·6547

印 数:1~4000

定 价:35.00 元(含光盘)

地 址:北京清华大学学研大厦

邮 编:100084

客户服务:010-62776969

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770175-3103 或(010)62795704

前言

FOREWORD

《操作系统原理 UNIX 篇》(以下简称《UNIX 篇》)是《操作系统原理 DOS 篇》(第二版)(以下简称《DOS 篇》)的姊妹篇。随着信息高速公路的飞速发展,因特网成为计算机界甚至许多家庭的热门话题,而 UNIX 成为因特网的主要开发基础。因此本书重点突出 UNIX,以原理为主,结合原理介绍 UNIX 的相关技术,而把 MS-DOS 的资源管理部分浓缩到最后一章中。

全书共 9 章:第 1 章操作系统引论;第 2 章 UNIX 系统组成与功能服务;第 3 章存储管理;第 4 章进程管理;第 5 章设备管理;第 6 章文件管理;第 7 章作业管理;第 8 章 UNIX 系统网络应用接口;第 9 章微机操作系统功能的实现。各章结合原理,相应地介绍了 UNIX 系统的功能。UNIX 本身没有作业管理模块,其作业功能是通过 Shell(命令解释器)实现的,所以本书在结构上把 Shell 的内容作为第 7 章中的一个独立部分。本书对 UNIX 技术的介绍着重于其内核原理和网络技术。为了使读者能同时熟悉 UNIX 的使用,第 7 章较详细地介绍了 UNIX Shell 的命令解释和过程设计两种功能。

本书的配书光盘是操作系统多媒体教学光盘,简称《UNIX 篇》配书光盘,包含 UNIX 篇教案、UNIX 篇点听系统、OS_Web 动画集。该光盘把课堂教学要点全部用电子教案、网页教案、矢量动画展示出来,围绕操作系统的五大模块原理和操作系统的三种典型、UNIX 相关的技术、微机应用技术等内容,配合教材中的插图设计制作了 100 多个矢量动画演示。OS_Web 动画集通过视觉动画界面进入,生动有趣。UNIX 篇点听系统对课程中原理的重点、难点配置了 100 多个录音讲解,为学生提供了课外学习的辅导资料。光盘上设置了个人主页方式教学资源文件夹,为课堂教学与网络环境下学习相结合的教学方式提供了可利用的教学资源,以顺应信息时代教学改革的需求。

本书同时配套出版《操作系统原理 UNIX 篇习题与实验指导》(简称《UNIX 实验篇》),该书也有配书光盘。《UNIX 实验篇》每章习题都有答案,实验内容丰富,有原理训练和课程设计,有详尽的实验指导,并提供了学生自己设计实验的案例。这两本书的配书光盘经由部分高校试用,普遍受到老师和学生的欢迎,为深化本门课程的教学改革提供了良好的基础。

本书的编写参考了若干有关著作及教材,以及因特网上的有关内容。本书的读者对象是大学计算机专业的学生,由于配书光盘有多种版本的教案(72 学时和 54 学时),并配有多媒体辅助教材,也适合本科、专科、成教、高职等各种需求的教学。

本书的写作在张昆藏教授的指导下,延续《DOS 篇》第二版有关内容和理论联系实

际的风格。张昆藏教授从始至终地关心着本书的编写,并提供了详细的指导意见和宝贵资料。第8章 UNIX 网络应用接口、线程的应用由贺敏伟教授编写,吕耀兴副教授对本书的编写提出了宝贵意见,并编写了有关缓冲池的实例,其余部分由李林英编写。全书由李林英、贺敏伟统筹组织。配书盘由李林英、贺敏伟策划,袁承武副教授为多媒体技术指导,朱蕙老师为美术顾问。颜小洋博士、吕剑虹硕士在光盘的研制初期做了许多有益的工作,白明副教授为本书的配书光盘提供了课外阅读学习线程的应用实例,在这里对他们表示感谢!许多青年教师利用休息时间满腔热情地为《UNIX 篇》配书盘的制作作出了积极的贡献,还有多届计算机专业学生的毕业设计的积极参与,才有今日此书配书光盘的诞生。本配书光盘中的电子教案由刘兴林老师制作。全书和光盘由李林英统编,由张昆藏教授审阅。

本书和配书光盘从编写到出版凝聚了集体的智慧和力量,体现了创新精神。在此,对全国高等学校计算机基础教育研究会常务理事北京大学谢柏青教授、广东华南理工大学钟慕良教授的热情指导和帮助,对全国教育科学“十五”规划重点课题《信息化进程中的教育技术发展研究》的教育专家李龙教授和桑新民教授的热情指导和帮助,表示深深地感谢!对试用本教学光盘的教师和学生所提供的宝贵经验与建议,对清华大学出版社的大力支持与鼓励,对五邑大学、青岛大学、清华大学、北京大学、华南理工大学、华南师范大学的许多老师和学生的关心与支持,一并表示热忱的谢意!

本教材突出了操作系统原理的基本概念和 UNIX 操作系统的相关技术,力求通俗易懂。但由于作者水平所限,书中难免存在不当之处,恳请广大读者提出宝贵意见。

作者

于广东江门五邑大学

目 录

CONTENTS

前言	I
第1章 操作系统引论	1
1.1 什么是操作系统	1
1.1.1 操作系统的地位	1
1.1.2 操作系统的功能	3
1.2 操作系统分类	7
1.2.1 批处理系统	7
1.2.2 分时系统	8
1.2.3 实时系统	9
1.2.4 通用操作系统	11
1.2.5 网络操作系统	11
1.3 UNIX 操作系统简介	12
1.3.1 UNIX 概述	12
1.3.2 UNIX 的发展历史	12
1.3.3 UNIX 的主要特点	15
1.3.4 UNIX 的现状	17
习题	17
第2章 UNIX 系统组成与功能服务	20
2.1 UNIX 系统的功能	20
2.2 UNIX 系统的结构	20
2.3 UNIX 中断与陷入	24
2.3.1 中断的概念	24
2.3.2 中断优先级	25
2.3.3 中断源的分类	25
2.3.4 UNIX 中断处理	26
2.4 UNIX 系统的生成	31
2.4.1 UNIX 系统的启停	31

2.4.2	UNIX 系统初启的全过程	31
2.4.3	UNIX 系统的管理	33
2.4.4	UNIX 操作系统的生成与再生	38
习题		40
第 3 章	存储管理	42
3.1	存储管理概述	42
3.1.1	程序的装入和链接	42
3.1.2	分区存储管理	43
3.1.3	分布存储管理	49
3.1.4	分段存储管理	52
3.2	虚拟存储技术	57
3.2.1	虚拟存储器的概念	58
3.2.2	请求页式管理	60
3.2.3	请求分段	62
3.3	UNIX 存储管理技术	65
3.3.1	UNIX 交换技术	65
3.3.2	UNIX 请求调页	69
习题		74
第 4 章	进程管理	79
4.1	进程的概念及实体	79
4.1.1	进程的引入	79
4.1.2	进程的基本概念	87
4.2	进程的控制	91
4.2.1	创建和撤销	91
4.2.2	阻塞与唤醒	92
4.2.3	挂起与激活	93
4.3	进程的同步与通信	94
4.3.1	基本概念	94
4.3.2	加锁与开锁	96
4.3.3	信号量机制	97
4.3.4	信号量上的 P,V 操作的应用	100
4.3.5	经典进程同步问题	103
4.3.6	监督机制	106
4.3.7	进程通信	107
4.4	进程调度与死锁	110
4.4.1	进程调度的基本概念	111

4.4.2	进程调度算法	111
4.4.3	死锁的基本概念	113
4.4.4	死锁的对策	116
4.4.5	银行家算法	117
4.5	UNIX 进程管理技术	119
4.5.1	UNIX 进程的数据结构	119
4.5.2	UNIX 进程控制	121
4.5.3	UNIX 进程调度	129
4.5.4	UNIX 进程通信	129
4.5.5	UNIX 共享内存的系统调用	134
4.6	线程的基本概念	135
4.6.1	线程的引入	135
4.6.2	线程与进程的比较	137
4.6.3	以库函数为基础的线程和以内核支持为基础的线程	137
4.6.4	线程的应用	138
	习题	143
第 5 章	设备管理	149
5.1	概述	149
5.1.1	设备的分类	149
5.1.2	设备管理的目标 and 功能	150
5.2	通道技术	151
5.2.1	I/O 控制方式	151
5.2.2	通道	153
5.3	缓冲技术	155
5.3.1	缓冲的引入	155
5.3.2	缓冲区的分类	156
5.3.3	缓冲池	157
5.4	设备分配	160
5.4.1	设备分配策略	160
5.4.2	设备分配程序	163
5.5	UNIX 设备管理技术	165
5.5.1	概述	165
5.5.2	UNIX 块设备管理	166
5.5.3	UNIX 字符设备管理	172
	习题	174
第 6 章	文件管理	176
6.1	概述	176



6.1.1	文件和文件系统	176
6.1.2	文件的分类	177
6.2	文件结构和存取方法	178
6.2.1	文件的逻辑结构	179
6.2.2	文件的物理结构	179
6.2.3	文件的存取方法	181
6.2.4	文件存储空间管理	182
6.3	文件目录	183
6.3.1	单级目录和二级目录	183
6.3.2	多级目录	184
6.3.3	检索目录	184
6.4	文件的使用	185
6.4.1	文件操作的系统调用	185
6.4.2	文件共享、保护和保密	186
6.5	UNIX 文件管理技术	189
6.5.1	UNIX 文件结构和存取方法	189
6.5.2	UNIX 文件目录管理	193
6.5.3	UNIX 文件的操作	199
6.5.4	UNIX 文件的使用	205
	习题	209
第 7 章	作业管理	212
7.1	操作系统与用户的接口	212
7.1.1	系统调用	212
7.1.2	作业控制命令和键盘操作命令	213
7.2	作业调度	215
7.2.1	作业的状态及其转换	215
7.2.2	作业调度程序	216
7.3	作业控制	216
7.3.1	脱机控制方式	217
7.3.2	联机控制方式	218
7.4	作业控制语言概述	220
7.5	UNIX 的界面 Shell	221
7.5.1	Shell 控制语言的组成	221
7.5.2	Shell 程序设计	228
7.5.3	Shell 命令解释程序	236
	习题	245

第 8 章 UNIX 系统网络应用接口	247
8.1 UNIX 网络通信概述	248
8.1.1 协议软件的层次结构	249
8.1.2 OSI 参考模型	250
8.1.3 TCP/IP 参考模型	252
8.1.4 AT&T UNIX 的流机制和 RFS	257
8.1.5 Berkeley UNIX 的套接字机制和 NFS	261
8.2 Socket API	264
8.2.1 Socket API 的风格	264
8.2.2 Socket API 的过程	265
8.3 客户-服务器网络编程应用实例	269
8.3.1 客户-服务器模型	269
8.3.2 客户-服务器编程实例	271
8.4 标准网络应用程序	279
8.4.1 远程登录	280
8.4.2 文件传送与存取	283
8.4.3 电子邮件	289
习题	293
第 9 章 微机操作系统功能的实现	296
9.1 多种流行的微机操作系统简介	296
9.1.1 MS-DOS 操作系统	296
9.1.2 Windows 操作系统	297
9.1.3 Linux 操作系统	301
9.2 MS-DOS 操作系统功能的实现	304
9.2.1 MS-DOS 的组成与启动	304
9.2.2 MS-DOS 的进程管理	306
9.2.3 MS-DOS 的 TSR 驻留技术	310
9.3 微机操作系统功能浏览	318
习题	318
附录 操作系统原理 UNIX 篇教学光盘资料	322
后记	341
参考文献	342

第1章

操作系统引论

操作系统是管理计算机软、硬件资源的系统软件。“操作系统”被称为计算机的灵魂。没有它,计算机硬件资源的优势就无法发挥。操作系统是配置在计算机硬件上的第一层软件,所有其他软件——如各种语言、实用程序、大量的应用软件都必须在操作系统的支持下工作。本章通过对批处理系统、分时系统和实时系统的介绍,使读者了解操作系统所要解决的主要问题以及解决这些问题的方法,并对操作系统的一些基本概念有一个初步了解。风靡全世界的互联网,其 90% 的内容都是在 UNIX 系统下开发的。本章在最后一节对 UNIX 操作系统作了简要介绍。

1.1 什么是操作系统

1.1.1 操作系统的地位

1. 操作系统的由来

世界上第一台机械计算机——加法机,是 1642 年由法国数学家布累斯·帕斯卡设计成功的。1944 年,霍华德·艾肯教授在国际商用机器公司(IBM)的支持下,以继电器为基本部件完成一台“自动序列受控计算机”。从用机械方式完成计算发展到用电子元件完成计算,促进了电子技术与计算技术的发展。

世界上第一台电子计算机,是 1946 年由美国宾夕法尼亚大学附属莫尔电工学校物理教员穆奇里博士设计的。由于人工准备时间的慢速与计算机运算的高速产生了矛盾,促使人们去思考怎样把算题的数据、计算公式保存到计算机中,由计算机自动识别。美籍匈牙利人冯·诺依曼教授参加了第一台电子计算机的研究,他的“存储程序”思想是计算机设计的一大革命。

如果在慢速的机器上,人工干预 3 分钟,是无所所谓的。但在高速的机器上,人工干预 3 分钟,则会浪费处理机的宝贵时间。将人工干预转为自动转换,于是出现了早期单道批处理系统。

20 世纪 60 年代,硬件有两大发展:通道和中断技术出现了,中断技术具有中断主机的能力,进而出现多道程序设计系统。

为了设法使计算机自己管理自己,需要以高速处理问题的复杂的系统软件去对付高速运转的主机和各个部件,以组织严密的管理体制来解决计算机中的各种复杂问题。具



有这种功能的大型软件就是操作系统。

计算机的发展已经历了四代：

第一代计算机由电子管组成。

第二代计算机由晶体管组成。

第三代计算机由集成电路组成。

第四代计算机由超大规模集成电路组成。

操作系统的形成是第三代计算机关键性的特征。人工智能计算机和人工神经网络计算机的出现,以及运用量子力学奇异规则的高速量子计算机的出现,又使计算机的研究发展进入到更新的一代,这将引导操作系统的发展产生新的变革。

2. 大型机硬件结构

系统硬件是计算机系统的基础,操作系统及其他软件最终还是要以机器指令形式来访问和控制各种硬件资源。硬件系统的组织结构也在不断地改进。图 1.1 给出的是目前大、中、小型计算机经常采用的一种硬件系统组织方式。这种结构以主存为中心,CPU 和 I/O 系统充分地并行工作,并且通过这个双端口的主存相互通信。I/O 通道实际上是一台专用的 I/O 处理机,它接收 CPU 的委托,独立地执行自己的通道程序,以字节多路通道、成组多路通道和选择通道分别控制低速、中速和高速外围设备的工作。I/O 通道使 CPU 摆脱了对各种 I/O 设备的繁杂控制,而且还可使各种外围设备之间并行工作。即这种大、中、小型计算机系统是以 CPU 和 I/O 系统之间以及各 I/O 设备之间尽可能地并行工作来组织硬件系统的,目的是换取更高的性能/价格比。

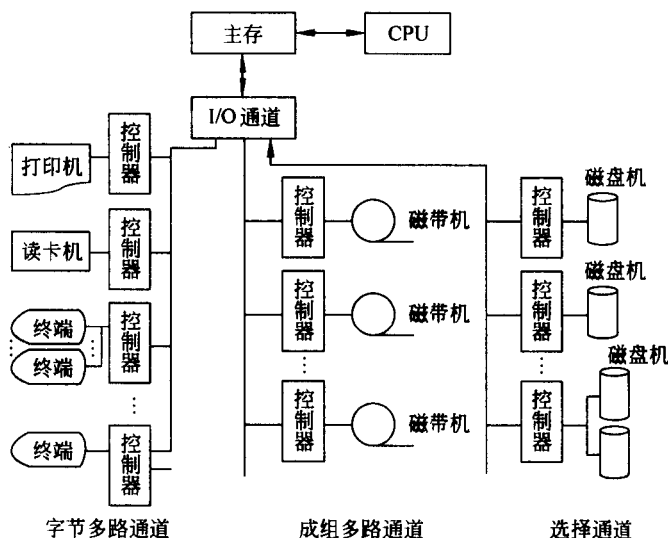


图 1.1 大、中、小型计算机系统

3. 微机硬件结构

微型计算机系统普遍采用单总线结构,以便扩充灵活、维护方便。CPU 通过系统总

以及与用户开发应用软件之间的关系。现在再从资源管理的观点,进一步介绍操作系统应具备的功能。

为了实现上述目标,操作系统通常具备以下五方面的功能。

(1) 进程管理

进程管理主要是对处理机进行管理。CPU 是计算机系统中最宝贵的硬件资源,为了提高它的利用率,采用了多道程序技术。如果一个程序因等待某一条件而不能继续运行时,就把处理机占用权转交给另一个可运行的程序。或者出现了一个比当前运行的程序更重要的可运行的程序时,后者应能抢占 CPU。为了描述多道程序的并发执行,就要引入进程的概念。通过进程管理协调多道程序之间的关系,以使 CPU 资源得到最充分的利用。

(2) 存储管理

内存(又称主存)是另一宝贵的硬件资源。虽然 RAM 芯片的集成度不断地提高、价格不断地下降,但需求量大,内存整体的价格仍较昂贵,而且受 CPU 寻址能力的限制,内存的容量也有限。因此,当多个程序共享有限的内存资源时,如何为它们分配内存空间,使它们既彼此隔离、互不侵扰,又能保证在一定条件下被调度。尤其是当内存不够用时,如何把当前未运行的程序及数据及时调出内存,要运行时再将其从外存调入内存等,这都是存储管理的任务。

(3) 设备管理

设备管理是指计算机系统中除了 CPU 和内存以外的所有输入、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。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、磁鼓、磁带)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、外围设备以外的另一类资源,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。

对上述四种资源的管理,并非是完全独立的,它们之间存在着相互依赖的关系。对每一种资源的管理,操作系统都要做到记录资源的使用情况,以某种策略分配资源,回收资源。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。除此之外,操作系统还应该向用户直接提供使用其本身的手段,这就是操作系统的作业管理功能。

(5) 作业管理

所谓作业,就是用户在一次计算或一次事务处理过程中,要求计算机系统所做工作的集合。例如,我们用 FORTRAN 语言编制了一个算题程序,那么要经过下列步骤:把源程序读入系统,对源程序进行编译,运行目标程序,打印输出结果。上述每一步骤都称为



作业步,它们的总和完成了一个作业。用户应该如何向系统提交作业,操作系统应该如何组织和调度作业的运行来提高整个系统的运行效率,这就是作业管理的任务。

2. OS 的层次模型

可以把整个计算机系统按功能划分为四个层次,即硬件、操作系统、系统实用软件和应用软件。这四个层次表现为一种单向服务关系,即外层可以使用内层提供的服务,反之则不行。如图 1.3 所示。

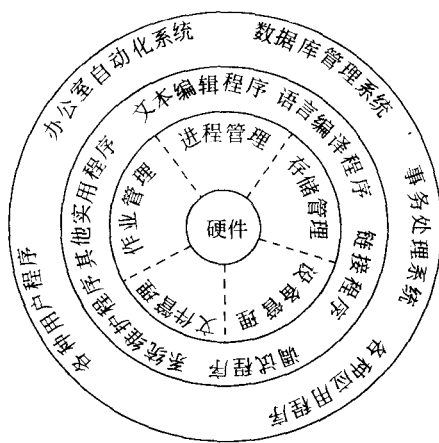


图 1.3 计算机系统层次结构

在计算机系统的层次结构中,包围着系统硬件的一层就是操作系统。它是最基本的系统软件,控制和管理着系统硬件(处理机、内存和外围设备),为上层的实用程序和用户应用程序提供一个屏蔽硬件工作细节的良好使用环境。另一方面,因为计算机中的程序、数据大多以文件形式存放在外存储器中,而构成文件系统,接受操作系统的管理。所以,尽管操作系统处于系统软件的最底层,但却是其他所有软件的管理者。我们看到,操作系统层在计算机系统层次结构中是特殊的、极为重要的一层,它不仅接受硬件层提供的服务并为上层的系统实用软件层、应用软件层提供服务,而且还管理着全系统的硬件和软件资源。系统实用软件层由一组系统实用程序(utility)组成,如语言编译程序(compiler)、文本编辑程序(text editor)、调试程序(debugging utility)、连接程序(linker)和系统维护程序(maintenance program)等。系统实用程序的功能是为应用软件以及用户加工自己的程序和数据提供服务,并为管理员对系统进行日常维护提供手段。系统实用软件和操作系统一起,应由计算机厂商提供。但使用时与操作系统不同,系统启动时操作系统即由外存储器调入内存并常驻内存,而系统实用程序则在需要时才从外存储器装入内存。

应用软件层中的各类软件,如数据库管理系统、办公自动化系统、事务处理系统等,可能是商售软件,也可能是用户自己编制的软件。

从操作系统的作用和它在计算机结构层次中所处的地位而言,它本身应具备三个层次。其最里层是 OS 对象(CPU、存储器、设备、文件和作业),最外层是面向用户的接口层,中间一层便是体现操作系统功能的、面向 OS 对象操作和管理的软件集成,这一层中



包括了进程管理、存储管理、设备管理、文件管理和作业管理。OS 面向用户的接口层包括了命令接口、程序接口、图形用户接口。用户从键盘输入命令,通过命令接口来获取 OS 的服务;用户在程序中进行系统调用来获取 OS 的服务;用户还可以操作鼠标点击图标通过图形用户接口来获取 OS 的服务。

3. OS 的定义

从系统层次结构和资源管理两个角度考察操作系统后,就可以完整地回答什么是操作系统了。

操作系统是直接控制和管理计算机硬件、软件资源的最基本的系统软件,用以方便用户充分、有效地利用这些资源并增强计算机的整体处理能力。

4. OS 的目标

配置操作系统,使用户可以通过计算机的各种命令使用计算机,操作系统起到了用户与计算机之间接口的作用,用户在操作系统基础上很容易进行软件的再次开发,从而方便了用户。由于计算机软、硬件资源的分配与回收由操作系统统一地管理,大大提高了资源的使用效率。引入操作系统的目的有如下两方面:

一是操作系统要方便用户使用。一个好的操作系统应给用户使用计算机提供一个良好的界面,使用户不必了解硬件和系统软件的细节就可方便地使用计算机。这里的“用户”是一个广义的概念,不仅包括系统的一般用户、系统管理员,还应包括系统实用软件的设计者。

二是操作系统应最大限度地发挥计算机系统资源的使用效率。这里的系统资源既包括 CPU、内存、外围设备等硬件资源,也包括程序和数据等软件资源(亦称信息资源)。应合理地组织计算机的工作流程,使这些硬件和软件资源为多个用户所共享。

方便性和高效性正是操作系统追求的最重要的目标。

5. 软件层次与虚拟机的概念

计算机的资源管理分为硬件资源管理和软件资源管理。软件又可分为两大类:系统软件和应用软件。系统软件本身又分为操作系统、语言处理系统和常用的例行服务程序。软件是在硬件的基础上对硬件性能的扩充和完善。

在操作系统的设计中,常常会出现“虚拟”的概念,这是操作系统中极重要的概念,它以硬件为基础,用软件模拟硬件的功能,成为“虚拟的硬件”。一台裸机每加上一层软件,就变成了功能更强的机器,这样的机器成为“虚拟机”。对于多道操作系统,一个实际的处理器被软件扩充为多个虚拟机器,使每个终端上的用户都拥有一台虚拟计算机,由于“虚拟机”屏蔽了物理细节,实现了多虚到一实的映射,又经软件扩充了功能,因此方便了用户、提高了效率。在后面的章节中会遇到“虚拟存储器”、“虚拟设备”等概念,都是用软件对硬件扩充的方法实现的。

从计算机系统结构中可以看到操作系统是最接近硬件的一层,是所有软件的基础。由此可见,研究操作系统的设计原理是很重要的,操作系统设计得越好,硬件的功能就发

挥得越好,就越有经济价值。学习操作系统设计的原理,了解怎样设计才能使裸机硬件资源的作用发挥得更好,怎样才能得到更高的使用效率和经济效益,为今后从事计算机应用系统的软件开发打下良好的基础。

1.2 操作系统分类

不同的硬件结构,尤其是不同的应用环境,应有不同类型的操作系统,以实现不同的目标追求。通常,将操作系统分成如下三类:

- 批处理系统(batch processing system)
- 分时系统(time sharing system)
- 实时系统(realtime system)

1.2.1 批处理系统

批处理系统的设计目标是提高系统资源的使用效率和作业吞吐量(单位时间里处理作业的个数)。这类操作系统要求用户使用系统提供的作业控制语言,描述自己对作业运行的控制意图,并将这些控制信息连同自己的程序和数据一起作为一个作业提交给操作员。操作员启动有关程序将一批作业输入到计算机外存,由操作系统去控制、调度各作业的运行并输出结果。由于作业进入计算机系统后,用户不再对作业的运行进行人工干预,从而提高了系统的运行效率。

早期的批处理系统是单道的,一个作业调入内存后要到正常运行结束或因某种原因运行不下去而中途退出后,系统才去调入下一个作业。作业总是一个接着一个地顺序运行,当一个作业运行时便独占了所有资源。显然,尽管这种单道批处理系统减少了作业交接时的时机浪费,但整个系统资源利用率和运行效率还是很低的。其主要原因是处理机和输入/输出设备的串行工作方式,当作业进行输入/输出操作时处理机处于空闲状态,等待输入/输出操作的完成,这就浪费了大量的处理机时间。

大型计算机的批处理系统见图 1.4。为了提高处理机资源的利用率,应尽可能地使处理机与输入/输出设备并行工作,这样便相应地出现了多道技术。多道(multiprogramming)是指在内存中同时存放若干个用户作业,允许这些作业在系统中交替地运行。如果这些作业搭配得合理,就能使处理机与输入/输出设备高度并行工作,当然,这还需要通道和中断机构等硬件的支持。多道技术看似简单,实践起来并不容易,如何使内存中的多道作业有条不紊地运行,如何为它们分配资源,这促进了操作系统的重大进步。对此,第3章将作进一步阐述。

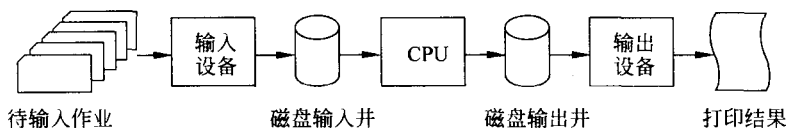


图 1.4 大型计算机的批处理系统