



高等学校电子与通信类专业“十一五”规划教材

计算机操作系统原理 —Linux实例分析

主编 肖竞华 陈建勋
主审 吴 谨



西安电子科技大学出版社
<http://www.xduph.com>

高等学校电子与通信类专业“十一五”规划教材

计算机操作系统原理 ——Linux 实例分析

主编 肖竞华 陈建勋
主审 吴 谨

西安电子科技大学出版社

2008

内 容 简 介

本书全面系统地介绍了操作系统的经典内容和最新发展,以及目前主流操作系统 Linux 的工作原理及实现技巧。

本书内容共 9 章,分为操作系统概述、进程管理、作业管理及用户接口、存储管理、文件管理、设备管理、死锁、操作系统的安全性、嵌入式操作系统。其中包括 Linux 系统介绍、Linux 进程管理、Linux 内存管理、Linux 文件管理、Linux 设备管理及 Linux 系统调用实现。从实用的角度把理论与实践紧密结合起来。

本书可作为高等院校计算机专业、信息管理及信息系统等相关专业的操作系统原理课程教材,也可作为操作系统与软件开发人员的参考书。

★本书配有电子教案,需要的老师可登录出版社网站,免费下载。

图书在版编目(CIP)数据

计算机操作系统原理: Linux 实例分析 / 肖竞华, 陈建勋主编.

—西安: 西安电子科技大学出版社, 2008.9

高等学校电子与通信类专业“十一五”规划教材

ISBN 978-7-5606-2119-7

I. 计… II. ①肖… ②陈… III. Linux 操作系统—高等学校—教材 IV. TP316.89

中国版本图书馆 CIP 数据核字(2008)第 133594 号

策 划 寇向宏

责任编辑 寇向宏

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467 邮 编 710071

http://www.xduph.com E-mail: xdupfb001@163.com

经 销 新华书店

印刷单位 陕西华沐印刷科技有限责任公司

版 次 2008 年 9 月第 1 版 2008 年 9 月第 1 次印刷

开 本 787 毫米×1092 毫米 1/16 印 张 18

字 数 423 千字

印 数 1~4000 册

定 价 25.00 元

ISBN 978 - 7 - 5606 - 2119 - 7/TP · 1086

XDUP 2411001-1

如有印装问题可调换

本社图书封面为激光防伪覆膜, 谨防盗版。

西安电子科技大学出版社
高等学校电子与通信类专业“十一五”规划教材
编审专家委员会名单

主任: 杨 震 (南京邮电大学校长、教授)

副主任: 张德民 (重庆邮电大学通信与信息工程学院副院长、教授)

秦会斌 (杭州电子科技大学电子信息学院院长、教授)

通信工程组

组 长: 张德民 (兼)

成 员: (成员按姓氏笔画排列)

王 晖 (深圳大学信息工程学院副院长、教授)

巨永锋 (长安大学信息工程学院副院长、教授)

成际镇 (南京邮电大学通信与信息工程学院副院长、副教授)

刘顺兰 (杭州电子科技大学通信工程学院副院长、教授)

李白萍 (西安科技大学通信与信息工程学院副院长、教授)

张邦宁 (解放军理工大学通信工程学院卫星系系主任、教授)

张瑞林 (浙江理工大学信息电子学院院长、教授)

张常年 (北方工业大学信息工程学院院长、教授)

范九伦 (西安邮电学院信息与控制系系主任、教授)

姜 兴 (桂林电子科技大学信息与通信学院副院长、教授)

姚远程 (西南科技大学信息工程学院副院长、教授)

康 健 (吉林大学通信工程学院副院长、教授)

葛利嘉 (中国人民解放军重庆通信学院军事信息工程系系主任、教授)

电子信息工程组

组 长: 秦会斌 (兼)

成 员: (成员按姓氏笔画排列)

王 荣 (解放军理工大学通信工程学院电信工程系系主任、教授)

朱宁一 (解放军理工大学理学院基础电子学系系主任、工程师)

李国民 (西安科技大学通信与信息工程学院院长、教授)

李邓化 (北京信息工程学院信息与通信工程系系主任、教授)

吴 谨 (武汉科技大学信息科学与工程学院电子系系主任、教授)

杨马英 (浙江工业大学信息工程学院副院长、教授)

杨瑞霞 (河北工业大学信息工程学院院长、教授)

张雪英 (太原理工大学信息工程学院副院长、教授)

张 彤 (吉林大学电子科学与工程学院副院长、教授)

张焕君 (沈阳理工大学信息科学与工程学院副院长、副教授)

陈鹤鸣 (南京邮电大学光电学院院长、教授)

周 杰 (南京信息工程大学电子与信息工程学院副院长、教授)

欧阳征标 (深圳大学电子科学与技术学院副院长、教授)

雷 加 (桂林电子科技大学电子工程学院副院长、教授)

项目策划: 毛红兵

策 划: 曹 晔 寇向宏 杨 英 郭 景

前 言

操作系统是计算机系统中最重要系统软件，是计算机科学与技术及相关专业的核心课程。操作系统中所涉及的概念、原理、算法等是从事计算机软硬件开发以及相关工作的工程技术人员所必不可少的基础知识。通过深入浅出的讲解，让学生熟悉和掌握操作系统的基本概念、原理等知识，同时深入理解目前使用广泛且易于开发的操作系统 Linux，是我们编写本书的目的。

本书既有操作系统原理的阐述，又结合了现代典型操作系统 Linux，将其作为实例，介绍了 Linux 操作系统的特点、主要功能及对计算机系统各种资源管理的实现技巧。

本书是作者融合操作系统教学经验，在广泛吸取国内外操作系统优秀教材的特色和操作系统研究成果的基础上编写的，适用于教学及自学。书中使用了图表和实例来阐明容易混淆的概念，从原理、核心技术和 Linux 实例三方面讲解操作系统。每章之后配有习题，不仅可以使学生加深和巩固对操作系统概念的理解，而且有助于培养学生创造性思维能力。本书具有以下特色：

(1) 简明扼要。本书以简洁的文字、清晰的结构，由浅入深地介绍了操作系统的基本原理及 Linux 实现技巧，便于教学。

(2) 系统性强。本书以操作系统的基本原理为主线，系统地讲述了 Linux 操作系统的相关基本概念、原理和实现技术，全面展现了当代操作系统的本质和特点。每章在介绍完操作系统相关原理基础上，最后讲解 Linux 系统的具体实现方法，使抽象的概念和原理在具体的系统中得到深化。

(3) 理论与实例结合。本书把操作系统的理论知识、设计原理与目前主流操作系统 Linux 的相关实现技术结合起来进行阐述，具体贴切，立足于培养学生对于系统软件的设计思想、方法和技巧。附录提供了操作系统部分实验内容及 Linux 常用命令。本书是一本既注重基本原理，又结合实际的实用教材。

对于西安电子科技大学出版社对本书的大力支持和帮助，以及武汉科技大学吴谨教授对本书提出的意见与建议，在此致以衷心的感谢！

由于编者水平有限，书中难免存在不足，恳请读者予以指正。

编 者

2008 年 8 月于武汉

目 录

第 1 章 操作系统概述	1	2.1.6 PCB 的组织方式	24
1.1 操作系统概念	1	2.1.7 进程的状态及其转换	26
1.2 操作系统的功能	2	2.2 进程控制	28
1.2.1 处理机管理	2	2.2.1 原语	28
1.2.2 存储管理	3	2.2.2 进程控制原语	28
1.2.3 设备管理	3	2.3 进程调度	30
1.2.4 文件管理	4	2.3.1 进程调度的功能	31
1.2.5 作业管理	4	2.3.2 进程调度方式	31
1.3 操作系统的发展	5	2.3.3 进程调度的原则	32
1.3.1 无操作系统的计算机系统	5	2.3.4 进程调度算法	32
1.3.2 早期简单批处理操作系统	6	2.4 进程的同步与互斥	36
1.3.3 多道批处理操作系统	6	2.4.1 临界资源和临界区	36
1.4 操作系统的类型	7	2.4.2 进程的同步	36
1.4.1 单用户操作系统	8	2.4.3 进程的互斥	37
1.4.2 批处理操作系统	8	2.4.4 用锁操作原语实现进程的互斥	37
1.4.3 分时操作系统	9	2.4.5 信号量及 P、V 操作原语	38
1.4.4 实时操作系统	10	2.4.6 利用信号量及 P、V 操作原语实现 进程的同步与互斥	39
1.4.5 通用操作系统	11	2.4.7 经典的进程同步互斥问题	42
1.4.6 网络操作系统	12	2.5 进程通信	46
1.4.7 分布式操作系统	12	2.5.1 进程通信机制	46
1.5 操作系统的特征	13	2.5.2 直接通信——消息缓冲通信	47
1.6 Linux 概述	14	2.5.3 间接通信——信箱通信	49
1.6.1 Linux 的优越性能	15	2.6 线程	50
1.6.2 Linux 内核的组成及功能	16	2.6.1 线程引入的目的	50
习题一	17	2.6.2 线程的概念	50
第 2 章 进程管理	18	2.6.3 线程属性	50
2.1 进程的概念	18	2.6.4 线程和进程的关系	51
2.1.1 程序的顺序执行及其特性	18	2.6.5 线程的功能特征	52
2.1.2 程序的并发执行及其特性	19	2.6.6 线程的分类及比较	52
2.1.3 并行与并发的概念差别	21	2.7 Linux 进程管理	53
2.1.4 进程的概念及其特性	22	2.7.1 Linux 进程的组成	53
2.1.5 进程的组成	23		

2.7.2 Linux 进程的状态及转换	54	4.3.2 逻辑地址的形式	105
2.7.3 进程在处理机上执行的核心态 和用户态	56	4.3.3 相关数据结构	106
2.7.4 进程空间和系统空间	56	4.3.4 页式地址转换	106
2.7.5 Linux 的进程调度	57	4.3.5 快表	107
2.7.6 Linux 进程控制	65	4.3.6 页的共享和保护	108
习题二	67	4.4 段式存储管理	110
第 3 章 作业管理及用户接口	69	4.4.1 段的划分技术	111
3.1 操作系统的用户接口	69	4.4.2 逻辑地址形式	111
3.1.1 程序级接口	69	4.4.3 内存分配	112
3.1.2 作业控制级接口	71	4.4.4 段表和段表控制寄存器	112
3.2 作业管理	72	4.4.5 地址转换	113
3.2.1 作业与作业步	72	4.4.6 段的共享与保护	113
3.2.2 作业控制方式	73	4.5 虚拟存储技术	114
3.3 批处理作业的管理	74	4.5.1 虚拟存储的基本思想	115
3.3.1 批处理作业的组织	74	4.5.2 页式虚拟存储管理	117
3.3.2 批处理作业的输入	75	4.5.3 段式虚拟存储管理	122
3.3.3 作业控制块	75	4.5.4 段页式存储管理	123
3.3.4 作业状态	76	4.5.5 页式管理与段式管理的比较	125
3.4 处理机调度级别	76	4.6 Linux 内存管理	127
3.4.1 作业调度	77	4.6.1 进程的地址空间	127
3.4.2 进程调度	83	4.6.2 Linux 的分段和分页机制	128
3.5 Linux 系统调用的实现	85	4.6.3 Linux 虚存管理的数据结构模型	130
3.5.1 系统调用	85	4.6.4 物理内存空间的管理	132
3.5.2 系统调用的实现	85	习题四	134
3.5.3 system_call 的执行过程	87	第 5 章 文件管理	136
习题三	88	5.1 文件与文件系统	136
第 4 章 存储管理	90	5.1.1 文件	136
4.1 存储管理的功能	90	5.1.2 文件的分类	137
4.1.1 内存分配与回收	90	5.1.3 文件系统及其功能	138
4.1.2 地址转换	91	5.2 文件的组织结构	139
4.1.3 存储保护	94	5.2.1 文件的逻辑结构	139
4.1.4 存储扩充	95	5.2.2 文件的物理结构	140
4.2 分区存储管理	95	5.2.3 记录的成组与分解	144
4.2.1 单一连续区分配方案	95	5.3 文件的存取方法	144
4.2.2 固定分区管理	96	5.3.1 顺序存取	144
4.2.3 可变分区管理	97	5.3.2 随机存取	145
4.3 页式存储管理	104	5.4 文件目录管理	145
4.3.1 页的划分技术	104	5.4.1 文件的组成	145
		5.4.2 文件控制块和文件目录	146

5.4.3 文件目录的结构	147	6.4.3 中断处理	192
5.5 文件存储空间的管理	149	6.4.4 中断返回	192
5.5.1 空闲区表法	149	6.4.5 中断的处理全过程	192
5.5.2 空闲块链法	150	6.4.6 多中断的处理	192
5.5.3 位示图法	152	6.5 设备分配	193
5.6 文件的共享与保护	152	6.5.1 设备分配的数据结构	193
5.6.1 文件的共享	152	6.5.2 设备分配技术	195
5.6.2 文件的保护	154	6.5.3 假脱机技术(Spooling)	196
5.7 文件操作	156	6.5.4 设备独立性	197
5.7.1 建立文件	157	6.5.5 设备分配算法	198
5.7.2 打开文件	157	6.6 I/O 软件	199
5.7.3 读文件	158	6.6.1 I/O 软件的目标	199
5.7.4 写文件	158	6.6.2 I/O 软件的组成	199
5.7.5 关闭文件	158	6.6.3 I/O 系统各层的主要功能	200
5.7.6 删除文件	159	6.7 设备驱动程序	201
5.8 Linux 文件系统	159	6.7.1 设备驱动程序的功能	201
5.8.1 虚拟文件系统	160	6.7.2 设备驱动程序的特点	201
5.8.2 文件系统的安装与卸载	168	6.8 磁盘存储管理	202
5.8.3 VFS 的系统调用	172	6.8.1 磁盘结构	202
5.8.4 跨文件系统的文件操作的基本原理	172	6.8.2 磁盘系统与磁盘分类	203
5.8.5 ext2 文件系统	172	6.8.3 磁盘调度	203
习题五	178	6.9 Linux 设备管理	207
第 6 章 设备管理	179	6.9.1 Linux 设备分类	207
6.1 设备及设备管理的功能	179	6.9.2 设备的识别	209
6.1.1 设备的分类	179	6.9.3 设备文件	209
6.1.2 设备管理的设计目标	180	6.9.4 Linux 设备驱动程序	210
6.1.3 设备管理的功能	180	6.9.5 Linux 逻辑 I/O 与设备驱动程序的接口	211
6.2 I/O 控制方式	181	6.9.6 Linux 系统调用与设备驱动程序的接口	214
6.2.1 查询等待方式	181	6.9.7 Linux 对中断的处理	215
6.2.2 中断方式	181	6.9.8 Linux 设备的 I/O 操作	216
6.2.3 DMA 方式	182	习题六	220
6.2.4 通道方式	183	第 7 章 死锁	222
6.3 缓冲技术	188	7.1 死锁的概念	222
6.3.1 缓冲技术的引入	188	7.2 死锁产生的原因	223
6.3.2 缓冲的种类	188	7.3 死锁的特征	224
6.4 中断技术	189	7.3.1 死锁产生的必要条件	224
6.4.1 中断的有关概念	190	7.3.2 资源分配图	225
6.4.2 中断的类型	191		

7.4 处理死锁的基本方法	226	9.1.1 嵌入式系统	255
7.4.1 死锁的预防	227	9.1.2 嵌入式操作系统	256
7.4.2 死锁的避免	228	9.2 嵌入式操作系统的发展历程	256
7.4.3 死锁的检测和解除	233	9.2.1 嵌入式技术的发展	256
习题七	236	9.2.2 嵌入式操作系统的发展	257
第8章 操作系统的安全性	238	9.3 嵌入式操作系统的特点	257
8.1 安全操作系统的必要性	238	9.3.1 实时性	257
8.1.1 针对操作系统的典型威胁	238	9.3.2 可剪裁性	258
8.1.2 现有防护方法及其不足	239	9.3.3 可靠性	258
8.2 我国安全操作系统所面临的问题	240	9.4 嵌入式操作系统的种类	259
8.2.1 目前常见的攻击者	240	9.4.1 非实时操作系统	259
8.2.2 病毒常见的攻击目标及对象	241	9.4.2 实时操作系统	259
8.2.3 计算机网络攻击的常见手法	243	9.5 几种代表性嵌入式操作系统	260
8.3 恶意程序防御机制	244	9.5.1 VxWorks	260
8.3.1 病毒防御机制	244	9.5.2 Windows CE	260
8.3.2 病毒检测与消除	245	9.5.3 嵌入式 Linux	261
8.4 安全操作系统设计	246	9.5.4 μ C/OS-II	261
8.4.1 建立安全模型	246	9.6 嵌入式操作系统的发展趋势	262
8.4.2 安全操作系统设计	247	习题九	262
8.4.3 安全操作系统的可信度验证	248	附录 A 实验	263
8.5 安全操作系统的关键问题	249	实验一 作业管理实例: UNIX/Linux shell	263
8.6 安全发展对策	250	实验二 Linux 内核添加新的系统调用	264
8.7 Linux 系统下安全机制的管理	251	实验三 文件权限管理设置实验	267
8.7.1 PAM 机制	251	附录 B Linux 常用命令	270
8.7.2 入侵检测系统	251	B.1 文件操作命令	270
8.7.3 加密文件系统	252	B.2 Linux 目录的创建与删除命令	272
8.7.4 安全审计	252	B.3 备份与压缩命令	273
8.7.5 强制访问控制	253	B.4 在 Linux 环境下运行 DOS 命令	274
8.7.6 防火墙	253	B.5 系统管理命令	274
习题八	254	参考文献	278
第9章 嵌入式操作系统	255		
9.1 嵌入式系统及嵌入式操作系统的概念	255		

从图 1.1 中可以看到, 操作系统对于用户来说是搭建在硬件平台(裸机)上的第一层软件, 它是对计算机硬件功能的首次扩展, 负责把系统资源管理起来, 以便充分发挥它们的作用, 使应用软件(用户)与机器硬件隔开, 系统内部的复杂性不显露在用户面前。用户可以使用操作系统提供的各种手段, 简单、方便地把自己的意图告诉计算机系统, 以实现所需要完成的工作。正是由于操作系统的卓越工作, 才能充分利用系统的资源, 并使用户方便地使用计算机。

综上所述, 操作系统是一组控制与管理计算机系统资源的程序集合, 它控制与协调并发活动, 为用户提供良好的接口。操作系统使整个计算机系统实现了高效率 and 高度自动化。

以上所说的系统资源是指计算机系统所具备的所有硬件设备和软件设施, 它包括中央处理器(CPU)、存储器(内存)、I/O 设备以及文件。

操作系统是必不可少的系统软件, 只有配置了操作系统, 才能使计算机系统体现出完整性和可利用性, 当用户要求计算机帮助完成其计算任务时, 程序才能有条不紊地运行。

1.2 操作系统的功能

操作系统是最基本的系统软件, 它在计算机系统中处于核心地位, 其功能可以从不同的角度进行说明。操作系统在计算机系统中的作用, 大致可以分为两方面: 对内, 操作系统管理计算机系统的各种资源, 扩充硬件的功能; 对外, 操作系统提供良好的人机界面, 方便用户使用计算机。操作系统在整个计算机系统中承上启下的地位通常体现在以下五个方面。在操作系统设计时, 通常也是基于这五个方面来划分操作系统的程序模块的。

1.2.1 处理机管理

计算机系统最重要的硬件资源就是中央处理机(CPU)。任何程序的运行, 都是由 CPU 执行其指令完成的。在单处理机多用户系统中, 处理机只有一个, 而内存又同时存在多道用户程序, 它们都要竞争得到处理机后才可能运行, 那么, 对于唯一的一个 CPU, 又要满足多个用户同时使用的要求, 这就涉及到操作系统对处理机时间分配的问题。也就是需要解决 CPU 分配给哪个用户程序使用, 它占用多长时间, 下一个又该轮到哪个程序运行等问题。为了能够清楚地描述多个程序的同时运行, 操作系统中引进了进程的概念。操作系统通过对进程的管理来协调多道程序之间的关系, 以达到充分利用处理机资源的目的。所以, 处理机的管理便体现为进程管理。进程管理的主要功能包括:

(1) 进程控制。当用户作业要运行时, 应为之建立一个或多个进程, 并为它分配除处理机以外的所有资源, 将它放入进程就绪队列。当进程运行完成时, 立即撤消该进程, 以便及时释放其所占有的资源。进程控制的基本功能就是创建和撤消进程以及控制进程的状态转换。

(2) 进程同步。所谓进程同步, 是指系统对并发执行的进程进行协调。最基本的进程同步方式是使诸进程以互斥方式访问临界资源。此外, 对于彼此相互合作、去完成共同任务的进程, 应由系统对它们的运行速度加以协调。

(3) 进程调度。当一个正在执行的进程已经完成或因某事件而无法继续执行时, 系统应进行进程调度, 重新分配处理机。进程调度是指按一定算法, 如最高优先算法, 从进程就

绪队列中选出一个进程，把处理机分配给它，为该进程设置运行现场，并使之投入运行。

(4) 进程通信。对于相互合作的进程，在它们运行时，相互之间往往要交换一定的信息，这种进程间所进行的信息交换称为进程通信。

1.2.2 存储管理

任何一个程序运行的必要条件是它必须首先进驻内存。所以，在计算机系统中，内存是仅次于 CPU 的重要资源。由于受处理机寻址能力的限制，内存资源也相当紧张。在多道程序系统中内存的存储调度往往和处理机的调度结合起来，即只有当程序在内存时，它才有可能到处理机上运行，或者仅当它可以到处理机上运行时(已获得其他资源)，才能把它调入内存，这种调度能实现对内存最有效的利用。操作系统的存储管理程序就是管理有限的内存空间，其通常包括以下功能：

(1) 存储分配。存储分配即为系统中需要运行的多个用户程序以及程序运行所需的数据分配内存空间。

(2) 地址转换。地址转换即将用户程序中的程序地址转换为物理地址。

(3) 存储保护。整个内存空间被划分为系统区与用户区两部分，其中用户区存放着正在运行着的多个用户的程序和数据，系统区是操作系统本身的程序和数据驻留区，它们使用不同的存储区域。由于操作系统和多个用户程序同时在系统中运行，因此在同一个内存存储器中存放的系统及用户的程序和数据的安全就显得十分重要。如果用户程序可以自由地访问内存的任何区域，必将危及系统和其他用户程序和数据的安全。存储保护就是保证内存中的系统存储区域不被用户程序非法访问，某个用户的存储区域也不能被其他用户程序访问，即每个用户程序只在自己的存储区域内活动，互不干扰。

(4) 存储扩充。内存资源是有限的，是计算机中较为缺乏的资源之一，尤其是在多道程序运行的环境中，内存资源变得更加紧张，这就需要扩充内存的容量。这里说的扩充不是从硬件上增加内存芯片的数量来扩充内存容量，而是操作系统通过软件的方法为用户程序提供一个比实际内存空间大得多的存储空间，这就是“虚拟存储器”技术。

1.2.3 设备管理

设备管理是操作系统中最庞杂、最琐碎的部分，其原因是：

- 这部分要涉及很多实际的物理设备，它们品种繁多、用法各异。
- 各种外部设备都能和主机并行工作，而且有的设备可被多个程序所共享。
- 主机与外部设备，以及各类外部设备之间的速度极不匹配，级差很大。

基于这些原因，设备管理主要解决以下问题：

(1) 设备分配。每个用户程序在其运行期间随时可能需要使用外部设备，使用之前就向操作系统提出设备请求。一般情况下，外设的种类与台数是有限的(每一类设备的台数往往少于用户的个数)，这些设备如何正确分配是很重要的。设备管理程序根据设备的状况，把设备分配给该任务使用。此外，当有多个任务竞争使用某些设备时，设备管理要按照一定的分配策略合理地分配设备。

(2) 设备的传输控制。设备的传输控制实现物理的输入/输出操作，即组织使用设备的

有关信息、启动设备、中断处理、结束处理等。这些工作是由设备管理提供的设备驱动程序完成的。通过设备驱动程序控制和管理设备,使用户不必了解设备的物理特性和技术细节,就可以方便地使用和操作这些设备。

(3) 为用户使用设备提供统一的操作接口。计算机系统配备的设备种类繁多,它们的操作特性完全不同。不同计算机系统,配备的设备种类和数量也可能完全不同。为此,设备管理对各种设备操作提供了统一的操作接口。用户向系统申请和使用设备时与实际操作的设备无关,即在用户程序中或在资源申请命令中使用设备的逻辑名,只需给出设备的类型,指出操作方式(读或写),通过统一接口就可以简单地使用设备。这样不仅为用户使用设备提供了方便,而且也提高了设备的利用率。

(4) 缓冲管理。几乎所有的外围设备与处理机交换信息时,都要利用缓冲来缓和 CPU 和 I/O 设备间速度不匹配的矛盾,提高 CPU 与设备、设备与设备间操作的并行程度,以提高 CPU 和 I/O 设备的利用率。

(5) 虚拟设备。通常,把一次仅允许一个进程使用的设备称为独占设备。系统可通过某种技术使该设备成为能被多个用户共享的设备,以提高设备利用率及加速程序的执行过程,使每个用户都感觉到自己在独占该设备。

1.2.4 文件管理

文件是计算机系统的软件资源,是各种程序和数据集合,它们以文件形式可长期保存在计算机的辅助存储器上。文件管理就是对这些软件资源的管理。其主要功能如下:

(1) 文件的组织。为了有效地对文件进行管理,操作系统必须提供文件驻留在存储介质上的组织方式,称为文件的物理组织。此外,为了便于系统和用户使用文件,操作系统还必须提供面向文件存取的组织方式,称为文件的逻辑组织。按照文件的组织方式,系统可以有效地分配和回收文件的存储空间,在存取文件时准确地确定文件的存放位置。

(2) 提供文件操作的统一接口。为了便于用户对文件的操作,操作系统提供了统一的接口。当用户使用文件时,不必知道文件在存储介质上的具体位置,也不需了解存储设备的物理特性和操作细节,只需给出文件的名称就可以对文件进行读取、写入、检索、修改等操作。

(3) 文件的保护和共享。在外存上同时存放着操作系统和多个用户的文件,为了系统和用户的信息安全和保密,操作系统必须对文件采取严格的保护措施,杜绝文件的非法访问,以免造成文件的破坏。与此同时,为了充分利用系统的信息资源,必须允许多个用户可以同时访问某些文件,如用户可以使用系统提供的共享程序库,多个协同工作的用户可以共享某些程序和数据,操作系统在提供文件保护的同时还要实现在一定条件下的文件共享。

1.2.5 作业管理

作业是指用户在一次算题过程中或一次事务处理过程中,要求计算机系统所做工作的集合。作业管理实现作业调度和控制作业的执行。操作系统必须合理安排和组织作业的运行,并为用户运行作业提供最大的便利。作业管理是根据系统条件和用户需要,对作业的运行进行合理的组织、调度及相应的控制,其主要功能如下:

(1) 作业的组织与管理。作业管理的一个主要功能是对系统中所有的用户作业进行统一的组织和管理,以提高整个系统的运行效率。当有大量用户作业要求计算机运行时,作业管理需要按照用户的要求和作业的性质调度其中某些作业进入计算机系统内运行。

(2) 作业调度。作业调度是指根据系统的能力和当前作业的运行情况,按一定策略,从后备作业队列中选出一批作业,为它们分配所需的 I/O 设备和存储空间,将它们调入内存并为之建立相应的进程,使之成为具有获得处理机资格的候选进程。

(3) 作业控制。作业控制是指作业从进入系统开始,直到运行完成的整个过程中,用户可通过某种形式向系统发出各种命令,以对自己的作业进行控制和管理。

为了使用户能够方便地使用计算机运行作业,操作系统提供了两种接口:一种是作业级的用户接口,它向用户提供作业控制语言和操作命令,用户利用作业级接口来组织作业的工作流程和控制作业的运行;另一种是程序级接口,它是系统提供的一组系统调用,又称系统调用指令。当用户在程序中需要使用系统资源时,如存取文件、使用设备、占用内存空间等,可以使用系统调用指令向操作系统提出申请,由操作系统完成用户的请求。

1.3 操作系统的发展

操作系统(Operation System, OS)是随着计算机技术本身及其计算机应用的日益发展而逐渐发展和不断完善的。它的功能由弱到强,现已成为计算机系统的核心。

操作系统的发展大致经历了手工操作、早期批处理、多道程序系统、分时系统、实时系统、通用操作系统等阶段。

进入 20 世纪 80 年代,硬件技术的飞速发展以及微处理机的出现和发展,使操作系统有了进一步的发展,如单用户操作系统、网络操作系统、分布式操作系统及智能化操作系统。

1946 年诞生第一台电子计算机以来,计算机的每一代进化都以减少成本、缩小体积、降低功耗、增大容量和提高性能为目标。随着计算机硬件的发展,同时也加速了操作系统的形成和发展。

1.3.1 无操作系统的计算机系统

从第一台计算机诞生到 20 世纪 50 年代中期的计算机,属于第一代计算机,其中央处理机的基本元器件是电子真空管,存储器用继电器和磁芯充当,这时还没有操作系统的概念。计算机采用人工操作方式,程序员全部采用机器语言编程,计算机还远远不是真正意义上的自动运算,计算机的使用也非常不方便,这主要体现在以下三个方面:

(1) 程序员必须懂机器语言,因为程序员必须直接面对计算机硬件,所以要求他们必须具备较强的专业基础知识和熟练的基本技能。

(2) 程序准备繁琐,常常为了几分钟的数字计算,准备工作就要用去几小时甚至更长的时间,使用很不方便。

(3) 使用计算机必须提前预约,要运行某个程序必须提前跟机房预约机器时间。批准后要准时到机房执行计算,在分配运行的这段时间内程序和机器都不能出现问题,如果出现

问题, 就会被迫强行停止, 下机修改程序, 准备下次重复这一过程。

人工操作方式具有以下特点:

- (1) 独占性。用户独占全机, CPU 等待人工操作。
- (2) 串行性。一个用户使用完后, 下一个用户接着使用。

人工操作方式的缺点是, 计算机的有效机时严重浪费, 效率低。

1.3.2 早期简单批处理操作系统

第二代计算机从 1955 年到 1965 年, 持续了十年时间。这一阶段在硬件上使用分离的晶体管代替易损坏的真空管, 使用磁芯存储器代替继电器存储器, 从而增强了设备的可靠性, 降低了硬件的故障发生率; 在软件上为了避免程序制作时直接面对机器语言, 已经开发了 FORTRAN 高级语言和汇编语言以及 FORTRAN 编译器和汇编程序, 从而出现了真正意义上的程序员。

虽然有了上述改进, 但高级编程语言(例如 FORTRAN 和汇编语言)出现之后又有了新问题: 处理机执行一个作业的时间是离散的。由于用户程序是高级语言编写的, 在执行之前先要把它翻译成机器语言, 因此一个作业除了包括加载高级语言程序之外, 还要加载编译器, 保存编译后的目标程序, 并且要将目标程序和公用函数链接在一起。其中的每一步操作都需要操作员参与完成, 这样使用计算机一是不方便, 二是在程序运行期间由于操作员的参与浪费了许多宝贵的处理机时间。

20 世纪 50 年代中期, 由 General Motors 开发了第一个简单批处理操作系统, 它也是人类历史上的第一个操作系统, 运行在 IBM701 机上。从那时开始, 许多厂商不断对简单批处理操作系统进行改进和完善, 开发出了适合自家产品的简单批处理操作系统。

简单批处理操作系统的设计思路是, 计算机上电后先运行一个常驻内存的所谓的监控程序, 由计算机的操作人员把用户提交的若干作业组织成一组, 称为一批作业, 将它们拷贝在磁带或穿孔卡片上, 然后安装在计算机的输入设备上, 由监控程序按顺序读入每个作业并执行, 每个作业执行结束后将结果输出到磁带上, 再返回到监控程序读入下一个作业。待所有作业完全结束后, 操作员再把输出磁带拿去打印。由于这个监控程序一次处理一批作业, 所以称它为批处理操作系统, 后来为了与多道程序批处理系统相区分, 又称它为简单批处理操作系统或单道批处理操作系统。

1.3.3 多道批处理操作系统

第三代计算机从 1965 年开始到 1980 年为止, 持续了十五年。这一时期的产品有 IBM 生产的 System/360-370 系列和 CDC 6600-7600 系列等。其硬件特征是采用了小规模集成电路、微程序控制、流水线和高速缓存, 与第二代计算机相比, 其性价比有了很大提高。

第三代计算机不仅能执行大型科学计算, 而且能够进行商务数据处理。而这两种应用有着很大的差别, 主要表现在处理机执行科学计算时, I/O 操作较少, 而进行商务数据处理时, I/O 操作却较为频繁。因此为了在一台机器上满足各种用户的不同需求, 同时又不降低 CPU 的利用率。特别是随着处理机速度的提高, 处理机的高速度与输入/输出设备的低速度不匹配的矛盾就越来越突出。为了解决这一矛盾, 于是产生了多道程序设计技术, 出现了

多道程序系统。多道程序设计技术是指在计算机内存中同时存放几道相互独立的程序，使它们在管理程序的控制之下，相互穿插运行，并使计算机资源在多道程序切换中得到共享，同时能够最大限度地提高 CPU 的利用率。

多道程序运行的特征如下：

(1) 多道：计算机内存中同时存放几道相互独立的程序。

(2) 宏观上并行：同时进入系统的几道程序都处于运行过程中，即它们先后开始了各自的运行，但都未运行完毕。

(3) 微观上串行：从微观上看，内存中的多道程序轮流地或分时地占有处理机，交替执行。它的优点是系统的吞吐量大，其缺点是对用户的响应时间较长。

在多道程序系统中，多个程序共享系统资源，使得系统管理更加复杂，从而对软硬件提出了更多新课题，也促进了它们的进一步发展。

随着应用的普及和各行各业不断提出的新的要求，操作系统的种类也越来越多，其功能也更加强大。

1.4 操作系统的类型

当前计算机已广泛地、深入地应用于人类生活的各个领域。在不同的领域中由于使用计算机的规模和计算机完成任务的性质及要求不尽相同，因此人们开发出了适应不同需要的操作系统。

操作系统可按照不同方式进行分类。例如，按用户数目的多少，可分为单用户和多用户系统；根据操作系统所依赖的硬件规模，可分为大型机、中型机、小型机和微型机操作系统；根据操作系统提供给用户的工作环境，可分为单用户操作系统、多道批处理操作系统、分时操作系统、实时操作系统、网络操作系统和分布式操作系统等。

不同的操作系统，运行环境不同，设计目标也不同。例如，大型计算机系统资源丰富、性能强大、价格昂贵，其设计目标是要求充分发挥系统资源利用率，提高机器的作业吞吐量。与此相比，微型机系统资源简单，价格相对便宜，所以对资源的利用率就不是操作系统设计的主要目标，而以操作便利、简单实用、适应性强作为设计目标。小型机处于大型机与微型机之间，其操作系统的设计目标也处于两者之间，既要求有较高的系统资源利用率，又要求操作方便、具有较高的通用性。

随着计算机技术和软件技术长期发展，已形成了各种类型的操作系统，以满足不同的应用要求。按其使用环境和对作业的处理方式来分，操作系统的基本类型有：

- (1) 单用户操作系统。
- (2) 批处理操作系统。
- (3) 分时操作系统。
- (4) 实时操作系统。
- (5) 通用操作系统。
- (6) 网络操作系统。
- (7) 分布式操作系统。

下面对它们作一概要的说明。

1.4.1 单用户操作系统

单用户操作系统一次只能支持一个用户程序的运行。单用户操作系统向用户提供联机交互式的工作环境，比如 MS-DOS 就是一个典型的单用户操作系统。

单用户操作系统的特点是：单用户，不要求高利用率，具有良好的交互性。

1.4.2 批处理操作系统

批处理操作系统简称批处理系统。顾名思义，批处理系统就是对多个用户的作业进行成批处理的操作系统。批处理系统的设计目标是提高系统资源的使用率和作业吞吐量(单位时间内完成运行作业的个数)。例如，OS/360 MTV 就是一个典型的多道批处理操作系统。

批处理系统可以分为单道批处理系统和多道批处理系统。多道批处理系统是多道程序设计技术与批处理系统的结合，基本思想是每次把一批经过合理搭配的作业通过输入设备提交给操作系统，并暂时存入外存，等待运行。当系统需要调入新的作业时，根据当时的运行情况和用户要求，按某种调度原则，从外存中挑选一个或几个作业装入内存运行。用户用控制命令描述对作业每一步运行的具体安排，并将这些控制连同程序和数据一起作为一个作业交给操作系统，因此，在系统运行过程中不允许用户与其作业交互作用，即用户不能直接干预自己作业的运行，直到作业运行完毕。

批处理操作系统是一种早期的大型机用操作系统。不过，现代操作系统大都具有批处理功能。

1. 单道批处理系统

单道批处理系统是指在内存中除操作系统外，只存放一道用户的作业，并且所有的系统资源仅为这一道作业服务。这是一种早期出现的批处理系统。

在单道批处理操作系统中，对于多个用户的作业，有卫星机将这些作业输入磁带中，然后逐级再从该磁带中将作业一个一个地读入内存进行处理。作业运行完毕后，将结果也都输出到另一磁带中，当这批作业全部运行完后再由卫星机把磁带上的结果通过响应的输出设备输出，处理完一批作业后再处理下一批作业。

2. 多道批处理系统

在单道批处理操作系统中，虽然解决了人工干预对机器运行效率的影响，但整个系统的运行效率还是很低。这是因为当作业在运行过程中需要使用设备进行输入/输出操作时，处理机需要等待输入/输出操作的完成。这就是说，在设备进行输入/输出操作期间，处理机处于闲置状态。所以单道批处理系统虽然通过减少人工干预改善了计算机的效率，但是由于处理机与设备的串行工作方式仍浪费了大量的处理机时间。

在多道批处理系统中同时有多道作业在运行，作业可随时(不必集中成批)被接收进入系统，并存放于磁盘输入井中形成作业队列。然后，操作系统按一定的原则从作业队列中调入一个或多个作业进入内存运行。由于引进了处理机与设备并行工作方式，使得处理机可以同时运行多道作业，这些作业轮流交替占用处理机运行。在多道批处理系统中，不仅处理机，其他所有的系统资源均由多个作业共享，并且作业的批处理方式是通过一定的调度