

高等学校教材

计算机操作系统

(第二版)

汤子瀛 杨成忠 哲凤屏



西安电子科技大学出版社

高等学校教材

计算机操作系统

(第二版)

汤子瀛 杨成忠 哲凤屏

西安电子科技大学出版社

1992

(陕) 新登字 010 号

内 容 简 介

本教材介绍计算机系统中的一个重要系统软件——操作系统。全书共分两部分。第一部分介绍操作系统的基本原理,共八章。第一章概述了操作系统的形成、类型、特征和功能等,第二章系统地阐述了进程的概念、进程同步和通信、进程调度和死锁等,第三~六章分别介绍了作业管理、存储管理、设备管理及文件管理,第七、八章为多处理机操作系统和网络操作系统。第二部分介绍一个实例——UNIX 系统 V 的内部结构和 SHELL。

本教材是原 1984 年出版的电子工业部规划教材《计算机操作系统》的第二版(修订版)。这次修订增加了第七、第八章,改用 UNIX 系统 V 版本取代了原来的 UNIX V.6,对原教材各章都做了大量更新、删旧工作。

本教材可作为计算机和计算机应用专业的教科书,也可作为从事计算机工作的科技人员学习操作系统的参考书。

高等学校教材

计算机操作系统

(第 二 版)

汤子瀛 杨成忠 哲凤屏

责任编辑·李纪澄

西安电子科技大学出版社出版

西安电子科技大学印刷厂印刷

陕西省新华书店发行 各地新华书店经售

开本 787×1092 1/16 印张 26 4/16 字数 619 千字

1984 年 11 月新 1 版 1992 年 5 月第 2 版 1992 年 5 月第 8 次印刷 印数 133 001—138 000

ISBN 7-5606-0192-8/TP·0066 (课)

定价: 6.75 元

第一版 出版说明

根据国务院关于高等学校教材工作分工的规定,我部承担了全国高等学校工科电子类专业课教材的编审、出版的组织工作。从一九七七年底到一九八二年初,由于各有关院校,特别是参与编审工作的广大教师的努力和有关出版社的紧密配合,共编审出版了教材 159 种。

为了使工科电子类专业教材能更好地适应社会主义现代化建设培养人才的需要,反映国内外电子科学技术水平,达到“打好基础、精选内容、逐步更新、利于教学”的要求,在总结第一轮教材编审出版工作经验的基础上,电子工业部于一九八二年先后成立了高等学校《无线电技术与信息系统》、《电磁场与微波技术》、《电子材料与固体器件》、《电子物理与器件》、《电子机械》、《计算机与自动控制》,中等专业学校《电子类专业》、《电子机械类专业》共八个教材编审委员会,作为教材工作方面的一个经常性的业务指导机构,并制定了一九八二~一九八五年教材编审出版规划,列入规划的教材、教学参考书、实验指

导书等 217 种选题。在努力提高教材质量,适当增加教材品种的思想指导下,这一批教材的编审工作由编审委员会直接组织进行。

这一批教材的书稿,主要是从通过教学实践、师生反映较好的讲义中评选择优和从第一轮较好的教材中修编产生出来的。广大编审者,各编审委员会和有关出版社都为保证和提高教材质量作出了努力。

这一批教材,分别由电子工业出版社、国防工业出版社、上海科学技术出版社、西北电讯工程学院出版社、湖南科学技术出版社、江苏科学技术出版社、黑龙江科学技术出版社和天津科学技术出版社承担出版工作。

限于水平和经验,这一批教材的编审出版工作肯定还会有许多缺点和不足之处,希望使用教材的单位、广大教师和同学积极提出批评建议,共同为提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

第二版 前 言

《操作系统原理》在 1981 年出版以后, 1984 年对它进行了第一次重大修改, 将 RTOS 实例换为 UNIX 系统, 改名为《计算机操作系统》, 并作为电子工业部规划教材出版。该教材从出版至今已有 7 年多的时间。这期间, 原教材中所介绍的许多技术和方法, 都有了很大的发展。有些原来只能在少数大型机上实现的技术, 现已可应用于小型机甚至微型机, 但也有一些当时曾广为流行的方法, 现在却已很少采用。换言之, 1984 年版的该教材内容, 已有部分显得陈旧, 需要更新。

在此期间, 微处理机得到迅速发展, 而利用多个结点处理机构成的多微处理机系统来取代传统的大、中型机, 已成为一种重要趋势。此外, 自 80 年代以来, 计算机网络, 特别是局域网已成为市场商品, 它以极其迅猛的速度渗透到诸如银行、宾馆、工矿企业及政府机关等的辅助管理系统和办公自动化系统中。为此, 我们在再版计算机操作系统教材时, 特增加了“多处理机操作系统”和“计算机网络操作系统和分布式操作系统”两章, 使学生能更好地适应操作系统的发展。

自 70 年代后期, UNIX 系统已成为世界上最著名的操作系统之一。在 80 年代, 它无论在功能上还是在性能上, 都有了显著的提高, 促进它进一步占领市场, 成为事实上的 16 位和 32 位计算机的多用户操作系统标准。其版本在不断更新, 由 UNIX V.6 经历 UNIX V.7、UNIX 系统 II 到 UNIX 系统 V。系统 V 一直延续至今, 当然其版本仍在不断更新。此

次修订本教材时, 将以 UNIX 系统 V 来取代原来的 UNIX V.6。

本书共分十章, 前六章仍介绍操作系统的基本原理, 每章的名称虽与原教材相同, 但实际上全部进行了重写, 各章内容都有很大更新。第七、八章是新增加的, 介绍主要在 80 年代形成和发展起来的多处理机操作系统和网络操作系统, 第九、十两章为实例部分, 分别介绍 UNIX 系统 V 的内部结构和 UNIX 的 shell。

本教材第一、二、七、八章由汤子瀛编写, 第三、四、五、六章由杨成忠编写, 第九、十章由哲凤屏编写。最后, 全书由汤子瀛统一编排定稿。

本教材在整个编写过程中, 得到浙江大学计算机系张德馨教授的指导与关心; 此外, 肖志俊、祝邦文、汤小丹等同志, 在整理、校对、绘图及誊写等工作中都付出了辛勤劳动; 还有西安电子科技大学出版社的大力支持与合作, 使本教材得以如期地与读者见面。在此, 谨向上述各位表示衷心感谢。

限于编者的水平, 在这次的修订版教材中, 仍难免存在错误或不当之处, 恳请读者批评指正。

编 者

1991 年 10 月 20 日

第一版 前 言

本教材系由计算机与自动控制专业教材编审委员会计算机专业教材编审小组评选审定,并推荐出版。

该教材由杭州电子工业学院汤子瀛、哲凤屏,成都电讯工程学院杨成忠编著。重庆大学童頔教授担任主审。编审者均依据计算机编审小组审定的编写大纲进行编写和审阅的。

自1981年2月《操作系统原理》出版以来,经三届学生使用,积累了一些教学经验,也发现了教材中存在的一些问题。况且,该书是1979年着手编写的,五年以来,国际国内在操作系统领域中又取得了长足的进步,UNIX系统已是世界上最著名的分时系统之一,已成为PDP-11, VAX-11和Interdata 8/32等小型机系列以及M68000、Z8000、Inter8086等16位微型计算机的主要操作系统,国内外不少专家认为UNIX系统将成为16位计算机的标准操作系统。鉴于上述两个原因,我们借修编此书的机会,对原教材作了较大修改。再版操作系统教材,仍以操作系统原理为主,把RTOS实例换为UNIX系统,以使学生对目前世界上最有影响的操作系统有较深入的了解,以适应四化建设的需要。原书名改为《计算机操作系统》。

本书共分十章。原理部分共六章,基本内

容与原版书一~六章相同。实例部分分为四章,即第七~十章。第七章较深入地阐述了存储器与处理机管理;第八章对块设备和字符设备的管理作了简要的介绍;第九章介绍UNIX系统中最有特色的文件系统;第十章是有关高级进程管理和控制等方面的内容。

本书是按100学时编写的,若按80学时讲授,建议免讲部分内容。

本书第一、二、五章由汤子瀛修改,第三、四、六章由杨成忠修改。实例部分的第八、十两章由汤子瀛编写,第七、九章由哲凤屏编写。这次参加本书审稿的有上海交通大学尤晋元同志,国防科技大学的李玉玺同志,重庆大学的梁光春、李朴、彭玉禄和苟荣秀等同志。清华大学的金兰教授、上海交通大学的张吉铎同志对本书的出版给予了很大的支持。

原《操作系统原理》教材出版后,得到了从事计算机科学和工程工作的同志们,特别是广大教师和学生的大力支持,提供了很多非常宝贵的意见,特在此表示衷心感谢。限于编者的水平,这次再版教材定然还存在着不少错误和不妥之处,仍恳请读者批评、指正。

编者 1984年3月

目 录

第1章 操作系统引论

1.1 操作系统的形成	1
1.1.1 人工操作方式	1
1.1.2 脱机输入输出技术	2
1.1.3 批处理技术	2
1.1.4 多道程序设计技术	2
1.2 操作系统的基本类型	3
1.2.1 批处理系统	4
1.2.2 分时系统	5
1.2.3 实时系统	7
1.3 操作系统的特征和功能	8
1.3.1 操作系统的特征	8
1.3.2 操作系统的功能	10
1.4 操作系统的发展	12
1.4.1 微机操作系统	12
1.4.2 多处理机操作系统	14
1.4.3 网络操作系统	15
1.4.4 分布式操作系统	16
1.5 操作系统的层次结构	18
1.5.1 模块接口法	18
1.5.2 有序分层法	20
1.6 小结	22
习题	22

第2章 进程管理

2.1 前趋图	24
2.1.1 程序的顺序执行与特征	24
2.1.2 前趋图的定义	25
2.1.3 程序的并发执行和特征	26
2.1.4 Bernstein 条件	28
2.2 进程的基本概念	29
2.2.1 进程的定义和特征	29
2.2.2 进程状态及其演变	30
2.2.3 进程控制块	31
2.3 进程控制	33
2.3.1 内核	33
2.3.2 进程的创建和撤消	34

2.3.3 进程的挂起与激活	36
2.3.4 进程的阻塞和唤醒	37
2.4 进程同步	38
2.4.1 临界区	38
2.4.2 硬件指令机制	39
2.4.3 信号量机制	40
2.4.4 信号量的应用	42
2.4.5 “信号量集”机制	44
2.4.6 管程	46
2.5 经典进程同步问题	47
2.5.1 生产者-消费者问题	47
2.5.2 读者-写者问题	49
2.5.3 哲学家进餐问题	51
2.6 进程通信	52
2.6.1 进程通信类型	52
2.6.2 直接通信方式	54
2.6.3 直接通信的实例 ——消息缓冲通信	55
2.7 进程调度	57
2.7.1 调度的基本概念	58
2.7.2 进程调度算法	59
2.7.3 进程调度的实现	63
2.8 死锁	65
2.8.1 产生死锁的原因和必要条件	65
2.8.2 预防死锁	68
2.8.3 避免死锁	69
2.8.4 检测死锁	72
2.8.5 死锁的解除	74
2.9 小结	75
习题	76

第3章 作业管理

3.1 引言	79
3.2 作业的管理和调度	79
3.2.1 作业的状态	79
3.2.2 作业调度	81
3.3 用户与操作系统之间的接口	89

3.4 命令接口	90	4.4 覆盖和交换	131
3.4.1 命令接口的类型	91	4.4.1 覆盖	131
3.4.2 命令接口的结构	94	4.4.2 交换技术	132
3.4.3 终端处理程序	95	4.5 分页存储管理	133
3.5 命令语言	96	4.5.1 纯分页系统	133
3.5.1 命令语言结构	96	4.5.2 请求式分页系统——单段式虚拟 存储器	136
3.5.2 命令的种类	97	4.5.3 分页系统的优缺点	147
3.5.3 键盘命令的打入和执行	98	4.5.4 分页系统小结	148
3.5.4 命令文件	98	4.6 分段存储管理	148
3.6 程序接口	99	4.6.1 分段地址空间	149
3.6.1 什么是系统调用	100	4.6.2 实现原理	150
3.6.2 初始化与终止	101	4.6.3 保护措施	152
3.6.3 系统调用简介	101	4.6.4 分段存储管理——多段式虚拟 存储器	152
3.6.4 系统调用执行过程	103	4.6.5 分段动态链接	155
3.7 程序库和链接编辑	104	4.6.6 分段的共享	160
3.7.1 编译/链接编辑/执行序列	104	4.6.7 分段管理的优缺点	162
3.7.2 编译程序和源语句库	105	4.7 段页式存储管理——另一种段式虚拟 存储器	163
3.7.3 目标模块	108	4.7.1 实现原理	163
3.7.4 装入目标模块	109	4.7.2 管理算法	166
3.8 作业控制	110	4.7.3 段页式系统的优缺点	167
3.8.1 脱机作业控制	111	4.7.4 段式系统小结	167
3.8.2 联机作业控制	112	习题	168
3.9 小结	112		
习题	113		
第4章 存储器管理		第5章 I/O 设备管理	
4.1 引言	114	5.1 引言	170
4.1.1 存储器管理的目的和功能	114	5.2 设备管理的任务和功能	170
4.1.2 存储分配的三种方式	115	5.3 处理输入输出请求的步骤	171
4.1.3 重定位	116	5.3.1 I/O 请求	171
4.1.4 虚拟存储器概念	118	5.3.2 对输入输出控制系统的调用	172
4.1.5 小结	119	5.3.3 输入输出控制系统的功能	172
4.2 单用户系统的存储器管理	120	5.3.4 设备驱动程序	172
4.2.1 存储器的分配	120	5.4 输入输出控制系统(I/OCS)	173
4.2.2 存储保护	120	5.4.1 I/O 服务的请求	173
4.3 多用户系统存储器管理 ——分区式分配	122	5.4.2 逻辑设备和物理设备	173
4.3.1 固定式分区	122	5.4.3 数据缓冲区	175
4.3.2 可变式分区	123	5.4.4 I/O 挂起队列	176
4.3.3 可重定位分区分配	126	5.4.5 输入输出控制系统示例	176
4.3.4 多重分区分配	126	5.5 设备驱动程序	178
4.3.5 动态分区的操作和数据结构	127	5.5.1 服务请求的校验	178
4.3.6 分区的存储保护	129	5.5.2 设备的状态	179
4.3.7 小结	131	5.5.3 启动 I/O 请求	180

5.5.4 中断处理	181	6.9.2 符号文件系统	234
5.5.5 I/O 请求的完成	181	6.9.3 基本文件系统	234
5.5.6 设备驱动程序示例	181	6.9.4 存取控制验证	234
5.5.7 I/O 请求的处理步骤	183	6.9.5 逻辑文件系统	235
5.6 挂起队列的调度策略	184	6.9.6 物理文件系统	235
5.7 设备分配程序	187	6.9.7 设备策略模块	236
5.7.1 I/O 系统结构	187	6.9.8 I/O 调度和控制系统	236
5.7.2 设备管理中的数据结构	188	6.9.9 分配策略模块	236
5.7.3 设备分配机制	191	6.9.10 主要文件命令的算法	237
5.7.4 设备分配程序	193	6.10 小结	240
5.8 小结	195	习题	240
习题	195		
第 6 章 文件管理		第 7 章 多处理机操作系统	
6.1 引言	197	7.1 多处理机系统概述	242
6.2 文件、文件系统	198	7.1.1 多处理机系统的结构	242
6.2.1 文件	198	7.1.2 多处理机系统的分类	244
6.2.2 逻辑文件和物理文件	198	7.1.3 多处理机系统的优点	245
6.2.3 文件类型	199	7.2 多处理机系统的硬件结构	246
6.2.4 文件系统	199	7.2.1 分时和公用总线	246
6.3 文件组织和存取方法	200	7.2.2 交叉开关矩阵	248
6.3.1 顺序存取方法	202	7.2.3 多端口存储器	249
6.3.2 直接存取方法	203	7.3 多处理机操作系统的特征与分类	251
6.3.3 索引顺序存取方法	204	7.3.1 多处理机操作系统的特征	251
6.4 文件的物理组织	206	7.3.2 多处理机操作系统的功能	251
6.4.1 文件物理组织的不同方式	206	7.3.3 多处理机操作系统的类型	253
6.4.2 文件类型与文件存储器、 存取方法的关系	210	7.4 进程同步	254
6.4.3 多级索引	210	7.4.1 集中式同步与分布式同步	254
6.5 文件存储器存储空间的管理	212	7.4.2 集中式同步机构	255
6.5.1 空白文件目录	212	7.4.3 利用事件计数实现同步	257
6.5.2 空白物理块链	213	7.4.4 利用中心进程实现同步	259
6.5.3 位示图	215	7.5 进程通信	259
6.6 文件目录	216	7.5.1 信箱通信的原理	259
6.6.1 单级文件目录	216	7.5.2 信箱通信的实现方式	261
6.6.2 多级目录	218	7.5.3 异常情况处理	262
6.6.3 便于共享的目录组织	221	7.6 任务分配	263
6.6.4 符号文件目录的查询技术	223	7.6.1 设计任务分配算法时应考虑的问题	263
6.7 文件的存取控制	226	7.6.2 最早调度算法	265
6.8 对文件的各种操作	228	7.6.3 子集调度算法	266
6.8.1 文件的使用	229	7.6.4 搜索树	268
6.8.2 文件控制块	231	7.7 小结	270
6.9 文件系统的一般模型	232	习题	271
6.9.1 用户接口及初始化模块	233		
		第 8 章 网络操作系统与分布式操作系统	
		8.1 计算机网络概述	272

8.1.1 计算机网络的类型	272	9.1.6 进程上下文的保存和切换	317
8.1.2 交换网	273	9.1.7 进程调度	318
8.1.3 局域网	275	9.2 存储器管理	320
8.2 开放系统互连参考模型	277	9.2.1 对换	320
8.2.1 OSI/RM 的组成和分层体系 结构的概念	277	9.2.2 请求调页管理的数据结构	324
8.2.2 OSI/RM 中的低三层	280	9.2.3 换页进程	327
8.2.3 OSI/RM 的中间层	282	9.2.4 请求调页	329
8.2.4 OSI/RM 的高两层	284	9.3 设备管理	330
8.3 网络操作系统的类型	284	9.3.1 块设备的数据缓冲池	330
8.3.1 按网络控制方式分类	285	9.3.2 核心与驱动程序接口	334
8.3.2 按支持 NOS 的单机操作系统 类型分类	287	9.3.3 磁盘设备驱动程序	336
8.4 标识符系统	289	9.4 文件的内部结构	338
8.4.1 标识符系统的目标	289	9.4.1 索引结点的管理	339
8.4.2 标识符的类型	290	9.4.2 正规文件的物理结构	342
8.4.3 标识符管理	291	9.4.3 空闲磁盘块的管理	345
8.5 进程通信	292	9.4.4 目录文件	347
8.5.1 数据传输	293	9.4.5 文件表的管理	349
8.5.2 流量控制	294	9.5 与文件有关的系统调用	351
8.5.3 路径控制	296	9.5.1 系统调用 open	351
8.6 分布式进程同步	297	9.5.2 系统调用 creat	352
8.6.1 事件排序	298	9.5.3 系统调用 close	354
8.6.2 Lamport 算法	298	9.5.4 系统调用 link	354
8.6.3 Ricart and Agrawala 算法	299	9.5.5 系统调用 unlink	356
8.6.4 令牌传送法	300	9.5.6 系统调用 read 和 write	357
8.7 共享资源的管理	301	9.5.7 其它系统调用	358
8.7.1 硬盘共享	301	9.6 进程控制	359
8.7.2 打印机共享	302	9.6.1 系统调用 fork	359
8.7.3 文件和数据的共享	303	9.6.2 系统调用 exec...	361
8.8 网络提供的服务	304	9.6.3 系统调用 exit	362
8.8.1 电子邮件系统	304	9.6.4 系统调用 wait	363
8.8.2 文件传送、存取和管理	306	9.7 进程通信	364
8.8.3 作业传送和操纵	307	9.7.1 sleep 和 wakeup 通信机制	364
8.9 小结	307	9.7.2 软中断	366
习题	308	9.7.3 管道通信机制	368
第 9 章 UNIX 系统 V 的内部结构		9.7.4 消息机构	369
9.1 进程结构和进程调度	310	9.7.5 共享存储器机构	373
9.1.1 UNIX 的基本特性	310	9.7.6 信号量集机制	375
9.1.2 UNIX 系统核心的体系结构	311	9.8 小结	376
9.1.3 进程的数据结构	312	习题	377
9.1.4 进程状态及其转换	314	第 10 章 SHELL	
9.1.5 进程上下文	315	10.1 概述	379
		10.2 shell 命令语言	379
		10.2.1 简单命令	379

10.2.2 重定向与管道命令	383	10.4.2 shell 变量及其赋值	393
10.2.3 后台命令	384	10.4.3 命令表与命令行	394
10.2.4 通信命令	384	10.4.4 流程控制命令	395
10.3 正文编辑程序	385	10.4.5 命令替换与参数替换	397
10.3.1 行编辑程序 ed	386	10.4.6 shell 过程的运行	399
10.3.2 全屏幕编辑程序 vi	388	10.5 shell 命令解释程序	401
10.3.3 另一个正文编辑程序 ex	390	10.6 小结	404
10.3.4 源程序的编译、链接和运行	390	习题	404
10.4 shell 过程	392	参考文献	407
10.4.1 shell 过程的引入	392		

第1章

操作系统引论

计算机系统由硬件和软件两部分组成。通常把未配置软件的计算机称为裸机。直接使用裸机不仅不方便,而且将严重降低工作效率和机器的利用率。操作系统(Operating System)是为了填补人与机器之间的鸿沟,即建立用户与计算机之间的接口,而为裸机配置的一种系统软件。配置操作系统的目标有三:(1)提供一个计算机用户与计算机硬件系统之间的接口,使计算机系统更易于使用;(2)有效地控制和管理计算机系统内的各种硬件和软件资源,使之得到更有效的利用;(3)合理地组织计算机系统的工作流程,以改善系统性能(如响应时间、系统吞吐量等)。

操作系统是裸机上的第一层软件,它是对硬件系统功能的首次扩充。它在计算机系统中占据特殊重要的地位,所有其它软件,如编辑程序、汇编程序、编译程序、数据库管理系统等系统软件,以及大量的应用软件都是建立在操作系统基础上的,并得到它的支持和取得它的服务。从用户角度看,当计算机配置了操作系统后,用户不再直接使用计算机系统硬件,而是利用操作系统所提供的命令和服务去操纵计算机,因此把操作系统看作是用户与计算机之间的接口。事实上,操作系统已成为现代计算机系统(包含大、中、小及微型机)中必不可少的最重要的系统软件。

1.1 操作系统的形成

操作系统是在人们不断地改善计算机系统性能和提高资源利用率的过程中,逐步地形成和发展起来的。

1.1.1 人工操作方式

早期计算机的工作,基本上采用人工操作方式:由操作员将纸带(或卡片)装入纸带输入机(或卡片输入机),以把程序和数据输入计算机,当程序运行完毕,由用户取走纸带和计算结果后,才让下一个用户上机操作。这种人工操作方式具有以下两个特点:

(1) 用户独占全机。一台计算机为一个用户独占,系统中的全部资源由他一人支配,因此用户可以较方便地使用各种资源,不会出现因资源已被其它用户占用而等待的现象,但资源的利用率却非常低。

(2) CPU 等待人工操作。用户仅在上机时才能将纸带或卡片装入相应的输入设备,显然,此时 CPU 空闲;当计算完成,进行卸带取卡操作时, CPU 又空闲。可见, CPU 的利用率极不充分,这在运行短程序时尤为突出。

人工操作方式严重地损害了资源的利用率,此即所谓人机矛盾。对早期的计算机来说,

此矛盾尚不突出,因为计算机本身拥有的资源并不多,且计算速度低,计算所需时间相对较长。但随着计算机运算速度的提高、规模的扩大,人机矛盾就变得严重起来,甚至可使资源利用率降为百分之几,甚至更低。

随着 CPU 速度的大幅度提高, CPU 和 I/O 设备之间速度不匹配的矛盾也日益突出,这就导致了一种关键性硬件——通道的出现。它可使 CPU 的运算和 I/O 操作并行执行;采用缓冲技术使速度不匹配的矛盾得到进一步的缓和;而脱机输入输出技术的引入,获得了较为令人满意的结果。

1.1.2 脱机输入输出技术 (Off-Line I/O Technic)

1. 脱机输入技术 为解决低速输入设备与 CPU 速度不匹配的问题,可将用户程序和数据,在一台外围计算机的控制下,预先从低速输入设备输入到磁带上,当 CPU 需要这些程序和数据时,再直接从磁带机高速输入到内存,从而大大加快了程序的输入过程,减少了 CPU 等待输入的时间。

2. 脱机输出技术 当程序运行完毕或告一段落, CPU 需要输出时,无须直接把计算结果送至低速输出设备,而是高速地把结果送到磁带上,然后在另一台外围机的控制下,把磁带上的计算结果,由相应的输出设备输出,这就大大加快了程序的输出过程。

在采用脱机输入输出方式时,由于程序和数据的输入输出都是在外围计算机的控制下完成的,或者说它们是脱离主机进行的,故称之为脱机输入输出操作;反之,由主机控制输入输出的方式称之为联机输入输出。

1.1.3 批处理技术 (Batch Processing Technic)

批处理技术是指计算机系统对一批作业自动进行处理的一种技术。所谓作业是指用户程序及其所需的数据和命令的集合。

在脱机输入输出方式中,事先已把一批作业记录在一盘磁带上,这意味着作业的处理是成批的,且处理顺序已经排定。我们在系统中再配置监督程序,在它的控制下,先把磁带上的第一个作业传送到内存,并把运行的控制权交给它,当第一个作业处理完后又把控制权交还给监督程序,由监督程序再把第二个作业输入内存。按这种方式对磁带上的作业自动地、一个接一个地处理,直至把磁带上的所有作业全部处理完毕。这便形成了最初的批处理技术。可见,批处理技术是在解决人机矛盾和 CPU 与 I/O 设备速度不匹配的矛盾的过程中,亦即在提高资源利用率的过程中形成的。

1.1.4 多道程序设计技术 (Multiprogramming Technic)

一、多道程序

在采用批处理技术时,内存中仅存放一道程序,每当该程序发出 I/O 请求后, CPU 便处于等待 I/O 完成状态,致使 CPU 空闲。为改善 CPU 的利用率而引入了多道程序设计技术,

即同时把几个作业放入内存,并允许它们交替执行,共享系统中的各种硬、软件资源。当一道程序因 I/O 请求而暂停执行时,CPU 便立即转去执行另一道程序。这样不仅使 CPU 能得到充分利用,同时还可改善 I/O 设备和内存的利用率。允许多道程序运行的系统称为多道程序系统。

二、多道程序系统需解决的问题

虽然多道程序技术能提高系统的吞吐量和有效地改善资源利用率,但实现多道程序还需妥善地解决下述一系列问题:

(1) 在一个连续的内存空间中,同时驻留了多道程序。应为每道程序分配好内存空间,使它们各得其所,不会相互重叠而丢失信息;又应防止某道程序由于人为的因素或出现异常情况而破坏其它程序。

(2) 在单处理机系统中,处理机只有一个,且它为各道程序所共享,这必将引起各道程序对处理机的争夺。系统应协调这些程序对处理机的使用,使各道程序最终都能获得处理机而运行;对于紧急的任务,还应能使之优先获得处理机。

(3) 通常,一个系统中的 I/O 设备数量少于多道程序所要求的数量,这同样会引起各道程序对 I/O 设备的争用,系统应为各道程序分配 I/O 设备;此外,I/O 设备品种繁多,驱动程序又直接与硬件相关,又需要各种类型的缓冲,系统应对它们进行有效管理,使之既能提高 I/O 设备的利用率,又能极大地方便用户对设备的使用。

(4) 在一个系统中运行的程序,可能具有不同的类型。如有的程序属计算型,即该程序需要经过大量的计算后,方才要求输入输出;而有的程序属 I/O 型,其计算量小,但要求 I/O 操作的量却较大。不同程序所要求的资源也有多有少,紧急程度也各不相同。系统应组织好这些程序的运行,以保证系统的吞吐量最大、资源利用率最高,而又不延误紧急程序的运行。

(5) 通常,系统把大量有意义的信息以文件的形式存放在各种存储介质上。为提高存储空间的利用率,加速对信息的检索速度,系统应对它们进行组织和管理;并且为了方便用户,系统应提供存储和检索文件信息的手段。

为解决上述问题,在多道程序系统中,必须设置一组有机结合的软件;此外,还应提供方便用户使用计算机的软件。这样便形成了操作系统。至此,我们可以把操作系统定义为:“操作系统是控制和管理计算机硬件和软件资源、合理地组织计算机工作流程,以及方便用户的程序的集合。”

1.2 操作系统的基本类型

迄今为止,操作系统的诞生已有 30 年。在此期间,计算机系统得到蓬勃发展,相应地也就出现了各式各样的操作系统。对操作系统可采取多种方法进行分类,最常用的方法是按功能特征的不同而把操作系统分为三种基本类型:(1)批处理(操作)系统;(2)分时(操作)系统;(3)实时(操作)系统。随着计算机体系结构的发展,又出现了许多新型操作系统:(1)微机操作系统;(2)多处理机操作系统;(3)网络操作系统;(4)分布式操作系统。本节介绍按功能特征分类的几种操作系统。

1.2.1 批处理系统 (Batch Processing System)

一、批处理系统的类型

1. 单道批处理系统 在早期计算机系统中所配置的一种操作系统类型。其特征是：

(1) 无须作业调度和进程调度。只按作业在外存中排定的顺序，依次将它们调入内存，将处理机分配给调入的作业使之运行；(2) 内存中仅有一道作业在运行。仅当前一道作业运行完成，或出现不能处理的异常情况时，才重新调入其后继作业进入内存运行，在内存中始终只保持一道作业；(3) 作业完成的顺序与作业进入内存的顺序直接相关，即先进入内存的先完成。

2. 多道批处理系统 在60年代中期以后常配置的一种操作系统类型。其特征是：

(1) 须经过作业调度和进程调度。从作业提交给系统，并在外存上形成后备作业队列开始，到它获得处理机运行，须经过两级调度：①作业调度。系统按一定的算法将后备队列中的若干个作业送入内存，并给它们分配除处理机以外的必要资源，再将它们排成一个队列；②进程调度。系统又按一定算法从该队列中取出一个作业，使之获得处理机执行。

(2) 内存中可同时驻留多道作业。宏观上，有几道作业在同时运行，在单处理机时，每个时刻只有一道作业在执行。

(3) 作业完成的先后顺序与其进入内存的顺序并无严格的对应关系，即后进入内存的作业有可能先完成。

3. 远程批处理系统 这是配置在联机系统和计算机网络中的、能接收从远程系统送来的批量型作业，对它进行处理后，将结果送至指定系统的一种批处理系统。该系统是在通常的批处理系统基础上，再加上远程作业录入程序(RJE)后形成的。RJE程序具有接收远地作业，并将它排入本地批处理作业队列中的功能。以后便像本地作业一样地处理。远地批处理的功能十分有用，它是构成分布式处理系统的基本条件之一。

二、批处理系统的优缺点

批处理系统虽诞生于60年代初期，但至今仍作为一种主要的操作系统类型，并已将它配置于网络系统中。批处理系统的主要优点是：

(1) 系统吞吐量大。所谓吞吐量是指系统在单位时间内所完成的总工作量。批处理系统能以较多的时间对作业进行处理，操作系统的开销较小，故可获得较大的系统吞吐量；

(2) 资源利用率高。在批处理系统中采取了一系列措施来提高资源利用率，使系统资源得到有效的利用。

批处理系统也存在着某些严重缺点：

(1) 平均周转时间长。所谓周转时间是指从作业进入系统开始，到作业完成所经历的时间。由于在批处理系统中，一个作业一旦运行便将运行到完成。这必然使许多短作业的周转时间显著增长；

(2) 不能提供交互作用能力。用户将作业提交给系统后，直到该作业运行结束，都无法与自己的作业进行交互作用，因此，用户必须提供一份作业说明书，详细地说明作业应如

何运行,以及对可能出现的某种情况应采取的措施等。这将给程序的修改和调试带来不便。

1.2.2 分时系统 (Timesharing System)

分时系统既是操作系统的一种类型,又是对配置了分时操作系统的计算机系统的一种称呼。通常,在一台主机上连接了多个键盘显示终端,用户可以通过各自的终端,以交互作用方式使用计算机,共享主机上所配置的各种硬、软件资源。

一、分时系统的实现方法

为使终端用户不仅在编辑时能和系统进行会话,而且在作业处理过程中的各个阶段,也能与自己的作业交互作用,这就要求系统能对用户键入的命令及时响应。显然,在分时系统中,作业的运行方式不能像批处理系统那样,一个作业长期占有 CPU 运行,其它作业则因不能获得 CPU 而处于长期等待的状态,致使其用户请求不能得到及时响应。实现分时系统有下述几种方法:

1. 简单分时系统 在简单的分时系统中,内存中只有一道程序作为现行作业,其它作业仍在外存上。为使系统能及时响应用户请求,规定每个作业在运行一个时间片 (time slice) 的时间 (例如 100 ms) 后便暂停运行,由系统将它调至外存 (调出),再从外存上选一作业装入内存 (调入),作为下一个时间片的现行作业投入运行。若在不太长的时间内,例如 3 秒钟内能使所有的作业都轮流运行一个时间片,亦即在指定时间内每个用户作业都一定能运行,这就能使终端用户与自己的作业交互作用,从而保证每个用户请求都能获得及时响应。在简单分时系统中,由于内存中只存放一道程序,故系统性能欠佳。

2. 具有“前台”和“后台”的分时系统 为了改善系统性能,引入了所谓“前台”和“后台”的概念。这里,把内存划分为“前台”和“后台”两部分。“前台”存放按时间片调进/调出的作业流,其工作方式同前;“后台”部分存放批处理作业。仅当“前台”正在调进/调出或无调进/调出作业流时,才运行“后台”的批处理作业,并给它分配更长的时间片。图 1-1 示出了具有“前台”和“后台”的分时系统的工作情况。

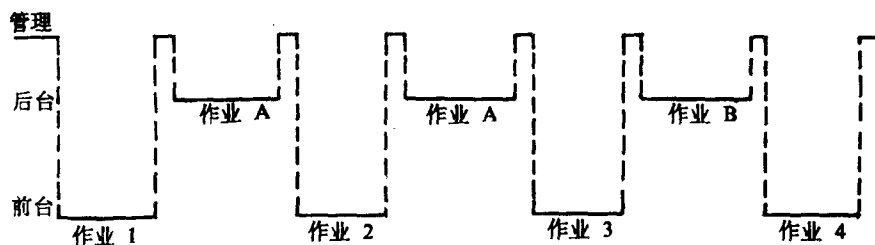


图 1-1 具有“前台”和“后台”的分时系统

3. 基于多道程序设计的分时系统 为了进一步改善系统性能,在分时系统中引入多道程序设计技术。在内存中可同时装入多道程序,每道程序无固定位置,对小作业可多装入几道程序,对一些较大作业则少装入几道程序。系统把所有具备运行条件的作业排成一个队列,使它们依次地轮流获得一个时间片来运行。当系统中除了终端型用户作业外,还有

批处理作业时，应赋予终端型作业以较高的优先权，并将它们排成一个高优先权队列；而将批处理作业另外排成一个队列。平时轮转运行高优先权队列的作业，以保证终端用户的请求能获得及时响应，仅当该队列空时，才运行批处理作业队列中的作业。

二、分时系统的特征

综上所述，我们可以归纳出分时系统有以下四个基本特征：

1. **多路性** 系统将若干个用户终端通过多路卡连接到一台主机上。宏观上，多个用户同时工作，共享系统资源；微观上，各终端作业是轮流运行一个时间片。多路性提高了资源利用率，节省了开支，促进了计算机的广泛应用。
2. **独立性** 每个用户各占一台终端，彼此独立操作，互不干扰。从用户角度说，每一用户并不感觉到有其它用户的存在，就像整个系统被他所独占。
3. **及时性** 终端用户请求能在很短时间间隔内获得响应。该响应时间间隔的大小由人们所能接受的等待时间来确定，通常把响应时间规定为 2~3 秒钟。
4. **交互性** 用户能与系统进行较广泛的人机对话，即用户从键盘输入命令，请求系统服务和控制程序的运行。系统能及时响应该命令，并在终端上显示响应结果。交互性有力地支持了工程设计和方案论证，显著地提高了研究、检查和调整程序的效率。由于交互作用是分时系统的重要特征，因而分时系统也被称为交互作用系统。

三、响应时间的讨论

在批处理系统中，主要考虑如何提高机器的利用率；而在分时系统中，必须注意响应时间。

1. 影响响应时间的若干因素

(1) 系统开销。在单道分时系统中，系统开销主要是进行调进/调出，即对换；在多道分时系统中，开销主要是进程调度和对换。对换所需的时间，主要由对外存的访问时间和信息的传输时间两部分组成。内、外存之间信息的对换速度愈高，CPU 便能以更多的时间去处理终端用户作业，从而减少响应时间，为此，应选用存取速度高的磁盘作为外存。

(2) 用户数目。如果系统中有 n 个同时性用户，时间片为 q ，则每个用户轮转一次所需时间为 nq ，可把 nq 近似地看做响应时间。当 q 一定时，响应时间与用户数目成比例，这就限制了同时性用户的数目。

(3) 时间片。当 n 一定时，响应时间比例于 q ，因而适当地减少时间片长度，便可改善对用户的响应时间。但随着时间片的减少，使原来只需一个中等时间片即能完成的作业或命令，就可能需要若干个时间片才能完成，这样反而增加了响应时间且降低了系统效率。因此，在考虑时间片大小时，应针对 70~80% 的命令和作业，使之能在一个时间片内完成。

(4) 对换信息量。在单道批处理环境下，对换所需时间将随着对换信息量的增加而增加，CPU 用于处理终端用户程序的时间将因对换时间的增加而减少，从而延长了作业在内存的驻留时间。

2. **响应时间的改善** 减少用户数目及减少时间片的大小，无疑可改善响应时间，但这会影响系统性能。这里介绍两种用减少对换信息量来改善响应时间的方法。

(1) 采用重入码文件以减少对换信息量。利用重入码编制成的文件可供多个终端用户