

高等院校计算机系列教材

计算机操作系统

主 编 王志刚 胡玉平
副主编 张如健 罗 心 徐雨明



WUHAN UNIVERSITY PRESS
武汉大学出版社

高等院校计算机系列教材

计算机操作系统

主 编 王志刚 胡玉平
副主编 张如健 罗 心 徐雨明



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

计算机操作系统/王志刚,胡玉平主编. —武汉:武汉大学出版社,
2005.8

(高等院校计算机系列教材)

ISBN 7-307-04585-0

I. 计… II. ①王… ②胡… 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(2005)第 084960 号

责任编辑:杨 华 黄金文 责任校对:黄添生 版式设计:支笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

印刷:湖北省孝感日报社印刷厂

开本:787×980 1/16 印张:22.375 字数:354 千字

版次:2005 年 8 月第 1 版 2005 年 8 月第 1 次印刷

ISBN 7-307-04585-0/TP·175 定价:33.50 元

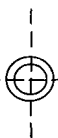
版权所有,不得翻印;凡购我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。



内 容 简 介

本书全面介绍了计算机操作系统的基本概念、原理和实现方法。全书共分十一章,第一章讲述了操作系统的发展概况;第二章至第十章分别阐述了操作系统的基本原理、概念和实现方法,包括中断技术,进程和线程的管理、进程的同步和通信,存储器管理,虚拟存储器,处理机调度,死锁问题,设备管理和文件系统;第十一章介绍了 Linux 操作系统;在第二章到第十章后面还介绍了 Windows 2000 操作系统,并较详细地分析了这两个系统的基本结构、主要的功能模块及其相互之间的关系。

本书吸收了国内外近几年出版的同类教材的优点,内容丰富。它既可以作为计算机专业和相关专业的本科生教材,也可作为从事计算机工作人员的参考书。



前言

操作系统是计算机系统中最基本的系统软件之一,它在计算机系统中具有中心地位,也是计算机教学中最重要的环节之一。操作系统被认为是信息系统的生命线,信息的处理和安全只能依靠技术手段来解决,而信息技术的核心是操作系统。操作系统的质量直接影响整个计算机系统的性能和用户对计算机的使用。

书中详细介绍了操作系统的基本原理和概念,吸收了国内外近几年出版的同类教材的优点,内容丰富,为读者学习、使用和分析操作系统提供了基本的原理和方法。这些知识对于计算机专业人员应用好计算机是不可缺少的。

全书共分十一章。第一章阐述了现代操作系统的发展概况;第二章至第十章分别阐述了操作系统的基本原理、概念和实现方法,包括中断技术,进程和线程的管理、进程的同步和通信,存储器管理,虚拟存储器,处理机调度,死锁问题,设备管理和文件系统。

在阐述基本原理和概念的基础上,介绍了当代操作系统的新概念、新技术和新方法。为了使读者对操作系统建立一个整体概念,对所学的知识能融会贯通,第十一章介绍了 Linux 操作系统,在第二章到第十章结合原理介绍了当今流行的 Windows 2000 操作系统,并较详细地分析了这两个系统的基本结构、主要的功能模块及其相互之间的关系。这些内容是深入了解和使用这些系统的重要知识。

本书第一章由王志刚、熊江编写,第二章由李燕清编写,第三章、第四章由雷清、张晓梅编写,第五章由胡玉平编写,第六章由戴祖雄编写,第七章由张如健编写,第八章由陈曦、罗心编写,第九章由徐雨明编写,第十章由周昱、彭大钧编写,第十一章由伍雁鹏编写,Windows 2000 由丁湘陵、贺慧琳、刘亚琴、汤小康编写。全书由王志刚、胡玉平统稿。

由于编者水平有限,书中可能会有不妥之处乃至缺点和错误,恳切希望读者赐教。

在本教材的编写过程中,得到了武汉大学出版社的大力帮助,在此表示衷心的感谢。

作者

2005 年 6 月

目 录

第一章 绪 论	1
1.1 操作系统概述	1
1.2 操作系统的发展过程	4
1.2.1 人工操作阶段	4
1.2.2 单道批处理阶段	4
1.2.3 执行系统阶段	6
1.2.4 多道程序系统阶段	6
1.3 操作系统的功能	7
1.3.1 用户接口	7
1.3.2 处理机管理	9
1.3.3 存储管理	9
1.3.4 设备管理	10
1.3.5 文件管理	10
1.4 操作系统的结构	11
1.4.1 系统结构	11
1.4.2 两种机器状态	13
1.4.3 两种系统界面	14
1.5 操作系统的特征	16
1.5.1 操作系统的基本特征	16
1.5.2 现代操作系统的某些新特征	18
1.6 操作系统的分类	20
1.6.1 多道批处理系统	20
1.6.2 分时系统	21
1.6.3 实时系统	23
1.6.4 网络操作系统	24
1.6.5 分布式操作系统	26
1.7 操作系统的启动和工作过程	27
1.8 Windows 2000 的模型	28



第二章 中断技术	35
2.1 中断在操作系统中的地位	35
2.2 中断的概念、作用和类型	36
2.2.1 中断的作用	36
2.2.2 中断的有关概念	37
2.2.3 中断的类型	38
2.2.4 中断嵌套、中断优先级和中断屏蔽	39
2.3 中断响应过程	39
2.4 中断处理过程	41
2.4.1 中断处理过程	41
2.4.2 中断处理例程简介	41
2.5 向量中断	44
第三章 进程和线程的描述与控制	46
3.1 进程的引入	46
3.1.1 程序的顺序执行	46
3.1.2 程序的并发执行及特点	47
3.2 进程的概念	51
3.2.1 进程的定义和特点	52
3.2.2 进程的基本状态	53
3.2.3 进程控制块	55
3.3 进程控制	56
3.3.1 进程控制的有关概念	56
3.3.2 进程创建	58
3.3.3 进程终止	59
3.3.4 进程等待	60
3.3.5 进程唤醒	61
3.3.6 进程挂起	61
3.4 线程	65
3.4.1 线程的引入	65
3.4.2 线程的控制	66
3.4.3 线程与进程的比较	67
3.4.4 用户级线程和内核级线程	68
第四章 进程的同步与通信	72
4.1 互斥与同步的基本概念	72
4.2 解互斥问题的算法	75

4.3	同步与互斥的基本工具——信号量和 P,V 操作	78
4.4	经典的互斥、同步问题	81
4.4.1	生产者-消费者问题	81
4.4.2	读者-写者问题	83
4.4.3	哲学家进餐问题	85
4.5	管程机制	87
4.5.1	管程的引入	87
4.5.2	管程概念	88
4.5.3	利用管程解生产者-消费者问题	90
4.5.4	利用管程解哲学家进餐问题	91
4.6	进程通信	93
4.6.1	消息缓冲通信	93
4.6.2	信箱通信	95
4.6.3	共享文件通信	96
4.7	Windows 2000 进程互斥与同步	97
4.8	Windows 2000 进程间通信	98
第五章	处理机调度	103
5.1	处理机调度类型	103
5.1.1	作业调度	104
5.1.2	中级调度	105
5.1.3	进程调度	106
5.2	调度算法性能的衡量	108
5.2.1	确定调度算法应考虑的因素	108
5.2.2	调度算法响应性能的衡量	109
5.3	调度算法	110
5.3.1	先来先服务调度算法	111
5.3.2	最短作业(短进程)优先调度算法	112
5.3.3	最高响应比优先算法	113
5.3.4	优先级调度算法	114
5.3.5	时间片轮转调度算法	117
5.3.6	多级反馈队列调度算法	118
5.3.7	均衡调度算法	120
5.3.8	实时调度算法	121
5.3.9	几种常见调度算法的比较	122
5.4	多处理机调度	124
5.4.1	负载分配调度	124



5.4.2	专用处理机分配调度	125
5.4.3	群调度	126
5.4.4	调度类和多模式调度器	126
5.5	线程调度	127
5.5.1	线程的状态	127
5.5.2	线程控制	128
5.5.3	线程调度的特征	129
5.5.4	线程优先级	129
5.5.5	线程时间配额	130
5.5.6	提高前台线程优先级的问题	132
5.5.7	调度数据结构	132
5.5.8	调度策略	133
5.5.9	线程优先级提升	135
5.5.10	对称多处理机系统上的线程调度	137
5.5.11	空闲线程	138
第六章 死 锁		142
6.1	死锁的基本概念	142
6.1.1	什么叫死锁	142
6.1.2	死锁产生的原因	143
6.1.3	产生死锁的必要条件	145
6.1.4	死锁表示方法	146
6.1.5	死锁的判定	147
6.1.6	处理死锁的基本方法	148
6.2	死锁的预防	149
6.3	死锁的避免	152
6.3.1	数据结构	152
6.3.2	系统的安全状态	152
6.3.3	死锁避免算法	154
6.4	死锁的检测和消除	157
6.4.1	死锁检测	157
6.4.2	死锁消除	158
6.5	处理死锁的综合措施	159
第七章 存储器管理		163
7.1	存储器管理的功能	164
7.2	存储器的地址变换	165

7.3 存储器的分区存储管理	167
7.3.1 固定式分区存储管理	167
7.3.2 动态分区存储管理	168
7.3.3 碎片问题及拼接技术	172
7.4 存储器的分页存储管理	174
7.4.1 分页存储管理的基本原理	175
7.4.2 存储空间分配和回收	177
7.4.3 地址变换机构	177
7.4.4 多级页表	180
7.5 存储器的分段管理	181
7.5.1 分段管理的原理	181
7.5.2 分段存储管理的实现	182
7.5.3 段的共享和保护	184
7.6 段页式管理	185
7.6.1 段页式管理的基本思想	185
7.6.2 段页式管理的实现方法	186
第八章 虚拟存储器	192
8.1 虚拟存储器概述	192
8.1.1 虚拟存储器的基本原理	192
8.1.2 虚拟存储器的理论基础	193
8.2 请求分页存储管理	194
8.2.1 页表结构	194
8.2.2 缺页中断处理	195
8.2.3 地址变换	196
8.3 页面置换算法	198
8.3.1 最佳置换算法	198
8.3.2 先进先出(FIFO)置换算法	199
8.3.3 最近最久未使用(LRU)置换算法	200
8.3.4 时钟置换算法	203
8.3.5 页面缓冲置换算法	205
8.4 页面分配算法和页面置换范围	206
8.4.1 进程正常运行所需的最少块数	206
8.4.2 页面分配算法	206
8.4.3 页面的分配和置换范围	207
8.5 请求分页系统性能分析	208
8.5.1 缺页率对有效访问时间的影响	208



8.5.2 抖动现象	210
8.5.3 页面大小的选择	213
8.6 请求分段存储管理	217
8.6.1 请求分段存储管理的实现原理	217
8.6.2 段的共享和保护	221
第九章 设备管理	226
9.1 设备管理概述	226
9.1.1 设备分类	226
9.1.2 设备管理的任务和功能	227
9.1.3 设备控制器与 I/O 通道	228
9.1.4 I/O 系统结构	230
9.2 输入/输出控制方式	232
9.2.1 程序直接控制方式	232
9.2.2 中断控制方式	232
9.2.3 DMA 控制方式	233
9.2.4 通道控制方式	235
9.3 缓冲技术	235
9.3.1 缓冲的引入	235
9.3.2 单缓冲	236
9.3.3 双缓冲	237
9.3.4 循环缓冲	237
9.3.5 缓冲池	238
9.4 设备分配	238
9.4.1 设备分配中的数据结构	239
9.4.2 设备分配策略	240
9.4.3 设备分配程序	241
9.4.4 Spooling 系统	242
9.5 驱动调度	243
9.5.1 移臂调度	244
9.5.2 旋转调度	247
9.6 软件的层次结构	248
9.6.1 中断处理程序	248
9.6.2 设备驱动程序	249
9.6.3 与设备无关的 I/O 软件	250
9.6.4 用户空间的软件	251
9.7 Windows 2000 I/O 系统结构和模型	253

9.7.1 I/O 管理器	254
9.7.2 即插即用管理器	255
9.7.3 电源管理器	257
9.8 Windows 2000 磁盘管理	258
9.8.1 磁盘存储类型	258
9.8.2 驱动程序 Ntldr	259
9.8.3 多重分区管理	260
9.8.4 高速缓存	261
第十章 文件管理	265
10.1 文件系统的概念	265
10.1.1 文件和文件系统	265
10.1.2 文件分类	267
10.2 文件结构与存储设备	268
10.2.1 文件的逻辑结构	268
10.2.2 文件的物理结构	268
10.2.3 文件的存取方法	269
10.2.4 文件的存储设备	270
10.3 文件存储空间的分配与管理	274
10.3.1 文件存储空间的分配	274
10.3.2 空闲存储空间的管理	277
10.4 文件目录管理	279
10.4.1 文件目录	279
10.4.2 单级目录结构	280
10.4.3 二级目录结构	281
10.4.4 多级目录结构	282
10.5 文件共享及文件管理的安全性	283
10.5.1 文件共享	284
10.5.2 文件保护	287
10.5.3 文件的转储和恢复	290
10.6 文件的使用	291
第十一章 Linux 操作系统	293
11.1 内存管理	294
11.1.1 虚拟内存	294
11.1.2 内存映射	295
11.1.3 页的分配和回收	298
11.1.4 按需换页	299



11.1.5	页的交换和释放	300
11.1.6	缓 存	301
11.1.7	页 缓 存	302
11.1.8	缓存页面的交换和释放	302
11.1.9	交 换 缓 存	304
11.2	进 程	305
11.2.1	Linux 进程	306
11.2.2	标 识 符	307
11.2.3	调 度	308
11.2.4	文 件	310
11.2.5	虚 拟 内 存	311
11.2.6	创建进程	313
11.2.7	时间和计时器	314
11.2.8	执行程序	315
11.3	进程通信机制	317
11.3.1	信 号	318
11.3.2	管 道	319
11.3.3	系统 V IPC 机制	320
11.4	设备驱动程序	325
11.4.1	Linux 设备驱动程序的特点	325
11.4.2	轮询与中断	326
11.4.3	DMA(直接内存存取)	328
11.4.4	存 储 器	329
11.4.5	设备驱动程序接口	329
11.4.6	硬 盘	331
11.5	文 件 系 统	332
11.5.1	第二扩充文件系统(EXT2)	334
11.5.2	虚拟文件系统(VFS)	338
11.5.3	缓冲区缓存	340
11.5.4	/proc 文件系统	341
11.5.5	特殊设备文件	342
	参 考 文 献	344



第一章 绪论

【学习目标】

了解操作系统的形成过程,建立起操作系统的整体概念;

熟悉操作系统的基本类型和服务方式;

掌握操作系统的定义、特征和功能。

【学习重点、难点】

操作系统的各种观点;

操作系统的基本类型和特征;

操作系统的服务和功能。

1.1 操作系统概述

操作系统,简称 OS (Operating System),是计算机系统中最基本和最重要的系统软件,是其他软件的支撑软件。它管理计算机系统资源,并通过这种管理为用户使用计算机提供公共的和基本的服务,从而成为用户与计算机之间的接口。

首先,操作系统是软件而不是硬件。特别是其内核,是一个确定的可执行程序,由几百个程序模块组成。然而,第一,操作系统的某些功能(如地址变换、内存保护等)是直接由硬件实现的;第二,操作系统的某些程序,如基本输入/输出系统(BIOS)里的许多程序,是固化在 ROM 内的,已经成了固件。所以,从这两点来说,操作系统又不是纯粹的软件,而是一个以软为主、软硬结合的整体。

其次,操作系统是系统软件,它是由系统程序员研制和编写的,并把它记录在外存(光盘或磁盘)上,随计算机硬件一起销售和运行。对于任何一个实用的计算机系统来说,可以没有其他软件,但绝不可以没有操作系统。

操作系统不仅与应用程序不同,也与其他系统程序不同。一个完善的机器系统通常都配有众多的系统软件,如编辑程序,编译、汇编程序,子程序库,各种命令的解释、执行程序,数据库管理系统等。所有这些程序,虽然与操作系统一样都属于系统软件,但它们都受操作系统的管理和控制,并得到操作系统的支持和服务,它们之间的关系如图 1-1 所示。

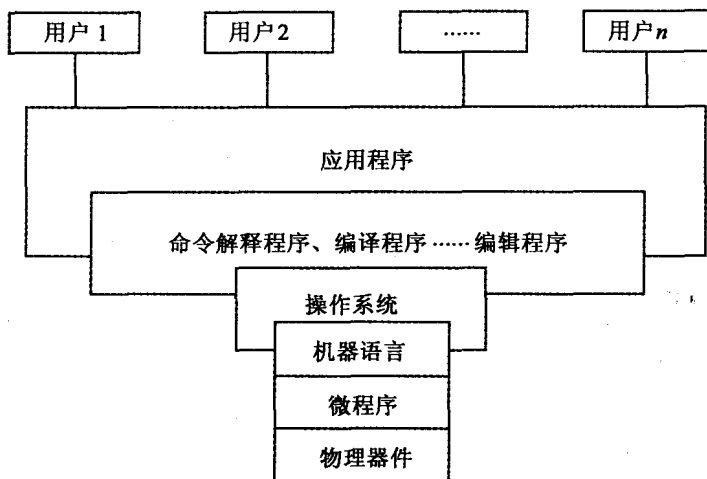


图 1-1 计算机系统的抽象层次结构

上述解释说明了如何看待操作系统的观点和设置操作系统的目的。看待操作系统有两个观点,即资源管理观点和服务用户观点。设置操作系统的两个主要目的是:高效地利用计算机系统资源和为用户使用计算机创造良好的工作环境。资源管理观点认为操作系统是管理计算机系统资源的程序。

计算机的系统资源包括硬件资源和软件资源。硬件资源包括 CPU、内存储器、外存储器、各种外部设备;软件资源主要指的是各种文件。所有系统程序以及用户要求长期保存的程序和数据都是以文件的形式存放在系统内的。

现代计算机系统通常都是多道程序系统,即同时有多个程序在主机内运行的系统。因此,所谓资源、管理,主要指的是在多道程序之间合理地分配和回收各种资源,使资源得到充分有效的使用,使程序得以有条不紊地运行。

从资源管理观点出发,可以把整个操作系统分成以下几部分:存储管理、CPU 调度(即处理机管理)、设备管理和文件系统。存储管理主要是指在多道程序之间分配和回收内存空间,其中包括作为内存的扩展和延伸的辅助空间;CPU 调度负责在多道程序之间分配 CPU,使每道程序都能得以运行;设备管理统一管理各种已登录系统的外部设备并负责外部设备与主机之间的信息交换;文件系统统一管理各种以文件形式保存在外存上的信息,负责文件的建立、读、写和删除等。

操作系统的服务用户观点是指,操作系统作为软件,它是一个为用户服务的大型复杂程序。用户在使用计算机的过程中需要的公共和基本服务主要有以下几个方面:

(1) 程序执行。任何一个用户程序都要在 CPU 上运行,而在此之前必须被先装入内存。



(2) 输入/输出操作。任何一个在计算机上运行的程序都必须先通过外部设备输入主机,程序在运行前和运行中都有可能输入其所加工的数据,任何程序都有运行结果,这些结果都要通过外部设备以适当的形式输出。

(3) 信息保存。用户如果认为他的程序或数据是重要的,需要长期保存以便将来再次运行或进一步完善,则要以文件的形式将它们存储在外存上。系统要为用户提供支持和服务,既要使用户能方便存取和使用自己的文件,又要防止文件的泄密、被盗用或丢失。

(4) 错误检测和处理。用户程序在运行过程中有可能出现各种错误,如算术运算溢出、程序中出现非法指令、地址越界等,操作系统应有能力对这些错误进行处理。此外,机器本身也有可能出现各种软、硬件故障,操作系统应及时地检测并修复它们,以保证用户程序的顺利运行。

上述服务功能是任何一个用户都需要的,是公共的和基本的。如果实现这些功能的程序都由每个用户去编写,这不仅增加了用户的工作量,而且这种重复性劳动往往容易产生各种各样的错误,使计算机系统的可靠性和有效性大大下降,甚至整个机器无法使用。因此,将这些功能集中起来,由系统程序员编写实现这些功能的程序是明智的、有效的、可靠的和易于实现的。这些程序的集合就是操作系统。除此之外,如果程序的运行需要服务,操作系统还为其他系统程序运行提供同样的服务。

操作系统的服务观点不仅使我们从一个侧面对操作系统的功能有了进一步的了解,而且也给出了其工作流程的一个大致的轮廓,因为操作系统的工作流程也就是沿着用户程序的执行提供各种服务的过程。此流程可粗略表示如下:程序和数据输入→作业收容→作业调度→作业运行→结果输出,如图 1-2 所示。

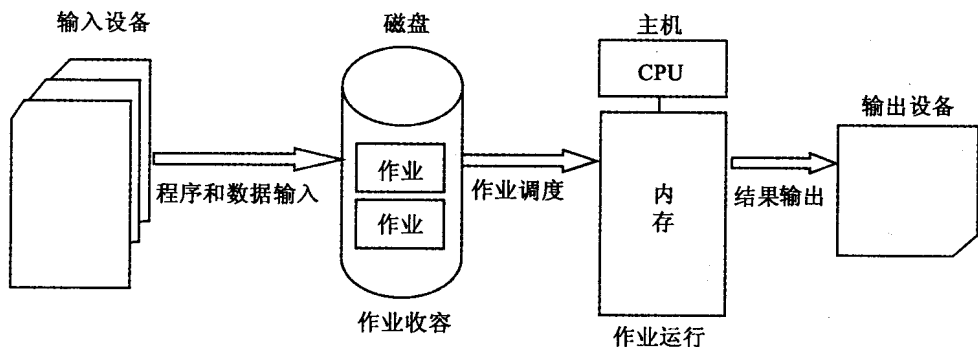


图 1-2 操作系统的主要工作流程

在输入阶段,操作系统提供输入服务,把程序和数据从输入设备上调入辅存,这时程序变成了计算机收容的作业(jobs);然后,操作系统按一定的调度算法,把这些作业调到主机(CPU + 内存)运行,这里涉及为作业分配外设和内存等资源;作业调



入主机后,在多道环境下运行,操作系统要为它们合理地分配 CPU,使它们都能得以运行;当一作业运行终止有结果需要输出时,操作系统还要提供输出服务,把用户所需要的结果在不同的设备(如显示终端、打印机、绘图仪等)上输出。操作系统就是按照这样的流程周而复始地为用户服务的。

操作系统的管理观点和服务观点是一致的:通过管理而达到服务的目的,通过管理为用户提供良好的工作环境。良好的标准有两条:方便和安全。一个系统是否方便好用,这可以由用户的直觉来决定。至于“安全可靠”主要指的是多用户、多作业共享同一计算机时互相不发生干扰以及防止“黑客”对计算机系统软、硬件资源的破坏和窃取。

1.2 操作系统的发展过程

操作系统同其他任何事物一样,也有一个产生与发展的过程。为了进一步加深对操作系统的理解,下面先对操作系统的形成过程进行一个简单回顾。

1.2.1 人工操作阶段

在 20 世纪 50 年代以前的第一代计算机中是没有配置操作系统的,计算机只是由控制台控制的一个庞大的物理机器。当时人们使用计算机的过程可大致描述为:首先由程序员将其编好的程序从纸带或卡片上装入内存,然后再通过控制台上的按钮或开关启动程序执行,最后当程序运行完毕时,取下纸带和运算结果,开始下一个用户程序。依次重复上述过程。这种人工操作方式存在以下三个主要问题:

(1) 资源独占。当一个用户开始操作后,计算机中的全部资源都归该用户所有,直到程序运行结束或出现错误时才会把资源转让给下一个用户。

(2) 串行工作。手工操作、计算机的运行以及 CPU 与 I/O 设备之间都是按串行方式工作的。

(3) 人工干预。计算机是在人的直接联机干预下进行工作的。

可见,人工操作方式严重降低了计算机资源的利用率,形成了所谓的人工操作方式与机器利用率的矛盾,简称人机矛盾。对早期的计算机而言,这种矛盾尚不突出,原因是那时的计算机本身所拥有的资源数量较少,计算速度也较慢,一般程序所需的计算时间要比其他操作时间(如装卸带、卡,输入/输出等)占用的时间长。但是,随着 CPU 速度的提高、存储容量的增大以及外部设备的增多,人机矛盾就变得越来越尖锐,使得这种方式到了非改不可的地步。

1.2.2 单道批处理阶段

为了解决上述问题,人们自然首先想到的是如何摆脱从一个用户程序过渡到另一个用户程序时的手工干预,使其能自动进行。这就产生了由计算机对一批用户程