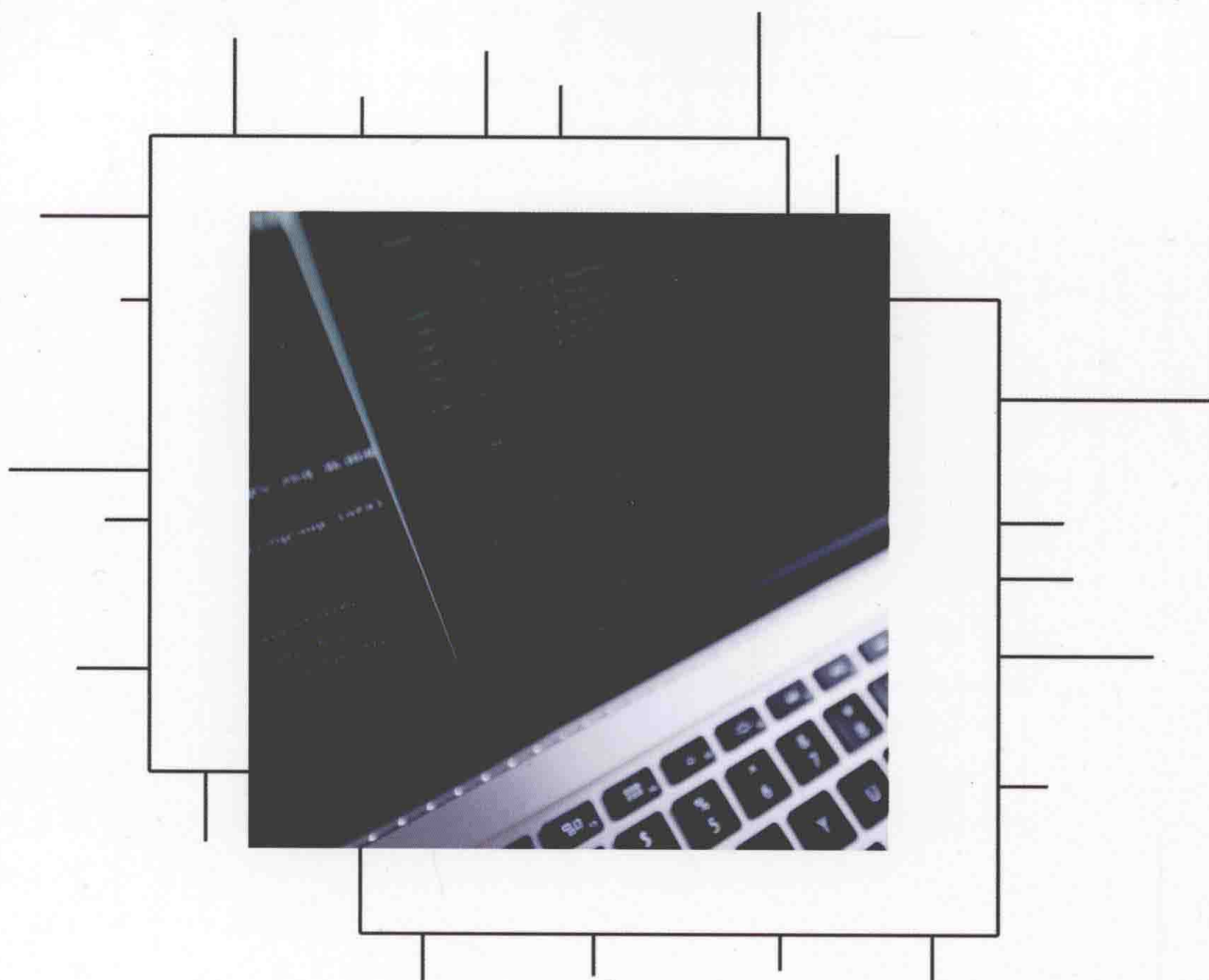




国家新闻出版改革发展项目库入库项目
高等院校计算机类规划教材



操作系统原理

周延森 编



北京邮电大学出版社
www.buptpress.com



国家新闻出版改革发展项目库入库项目
高等院校计算机类规划教材

操作系统原理

周延森 编



北京邮电大学出版社
www.buptpress.com

内 容 简 介

本书系统、深入地介绍了操作系统的基本原理,以操作系统的五大功能模块作为主线,分章节阐述了操作系统的理论基础。全书共分7章,内容包括绪论、用户接口、进程管理、处理机调度、存储管理、文件管理、设备管理。此外,每章均配有精选习题,题型丰富,有助于读者领会和掌握相关知识。

本书力求做到理论结合实际、突出应用,以帮助读者学习和掌握操作系统的理论知识。本书可作为计算机相关专业的本科生教材,也可作为研究生教材,还可供从事计算机及通信相关工作的科技人员参考。

图书在版编目(CIP)数据

操作系统原理 / 周延森编. -- 北京 : 北京邮电大学出版社, 2022.5

ISBN 978-7-5635-6637-2

I. ①操… II. ①周… III. ①操作系统 IV. ①TP316

中国版本图书馆 CIP 数据核字(2022)第 067790 号

策划编辑:姚 顺 刘纳新 责任编辑:王晓丹 谢亚茹 封面设计:七星博纳

出版发行:北京邮电大学出版社

社 址:北京市海淀区西土城路 10 号

邮政编码:100876

发 行 部:电话:010-62282185 传真:010-62283578

E-mail: publish@bupt.edu.cn

经 销:各地新华书店

印 刷:保定市中画美凯印刷有限公司

开 本:787 mm×1 092 mm 1/16

印 张:16.25

字 数:424 千字

版 次:2022 年 5 月第 1 版

印 次:2022 年 5 月第 1 次印刷

ISBN 978-7-5635-6637-2

定价:48.00 元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

操作系统是计算机系统的重要组成部分，操作系统课程则是计算机及相关专业的重要基础课程，同时也是计算机行业的工程或技术人员必须深入了解的基础知识。

考虑到本科生对基础内容的学习，本书结合作者多年教学的经验，主要讨论了操作系统的共性内容。本书具有以下特点：

(1) 主要关注操作系统基本概念、基本技术、基本方法的阐述，着眼于操作系统学科知识体系的系统性和实用性；

(2) 根据国内计算机专业研究生招生考试大纲的要求编排，基本覆盖了考试大纲的内容，并在每章习题部分安排了一些考研真题。

全书共分7章。第1章为绪论，主要介绍操作系统的形成与地位、发展阶段、定义与特征、基本类型，以及基本功能。第2章介绍了操作系统提供给用户的界面，即用户接口，主要有用户输入/输出方式、系统调用、图形化用户接口等。第3章围绕进程管理展开论述，从程序运行的方式开始，引入了进程的概念并讨论了进程的特征、状态变化及模型，重点阐述进程的同步与互斥问题、进程的通信、死锁，并对线程与进程进行了比较，最后讨论了管程。第4章为处理机调度，深入阐述了作业的基本概念、处理机调度的层次以及调度算法，提出了评价调度算法的定量和定性指标，并对实时调度算法进行了分析。第5章为存储管理，主要讨论存储管理的基本功能、分区存储管理、页式存储管理、段式存储管理以及段页式存储管理，其中包括最新的存储管理技术，如多级页表、快表等。第6章为文件管理，讨论文件及文件系统的基本概念、文件逻辑结构、文件物理结构、目录管理、文件存储空间管理、文件的共享和保护等。第7章为设备管理，在讨论了I/O系统、4种I/O数据传输控制方式之后，讲解了缓冲技术的引入及单缓冲、双缓冲和缓冲池的概念，对I/O软件的层次、I/O过程和SPOOLing技术进行了阐述，并详细介绍了磁盘调度算法。

本书参考了很多国内外同行关于操作系统的最新研究内容以及大量学术著作，有些已经在参考文献中列出，但篇幅所限，未能一一列出，在此一并表示感谢！

由于作者水平有限，书中难免有不妥之处，殷切希望广大读者批评指正。

第 1 章 绪论	1
1.1 操作系统的形成与地位	1
1.1.1 操作系统的形成	1
1.1.2 操作系统的地位	2
1.2 操作系统的发展阶段	3
1.2.1 手工操作系统阶段	3
1.2.2 早期批处理系统阶段	4
1.2.3 多道程序系统阶段	5
1.2.4 新一代操作系统阶段	6
1.3 操作系统的定义与特征	6
1.3.1 操作系统的定义	6
1.3.2 操作系统的特征	7
1.4 操作系统的基本类型	8
1.4.1 多道批处理操作系统	8
1.4.2 分时操作系统	9
1.4.3 实时操作系统	10
1.4.4 网络操作系统	11
1.4.5 分布式操作系统	12
1.4.6 多处理机操作系统	13
1.4.7 嵌入式操作系统	13
1.5 操作系统的基本功能	14
1.5.1 处理机管理	14
1.5.2 存储器管理	15
1.5.3 文件管理	17
1.5.4 设备管理	17
1.5.5 用户接口	18
1.6 研究、分析操作系统的几种观点	19
习题	21
第 2 章 用户接口	23
2.1 操作系统提供给用户的接口	23

2.2 用户的输入/输出方式	24
2.2.1 联机输入/输出方式	24
2.2.2 脱机输入/输出方式	25
2.2.3 直接耦合方式	25
2.2.4 SPOOLing 系统	26
2.3 系统调用	27
2.3.1 系统调用的基本概念	27
2.3.2 系统调用的主要种类	28
2.3.3 系统调用的功能	28
2.3.4 系统调用与一般过程(函数)调用的主要区别	29
2.3.5 系统调用的运行过程与使用方法	29
2.4 图形化用户接口	31
习题	32
第3章 进程管理	33
3.1 进程	33
3.1.1 进程的引入和定义	34
3.1.2 进程上下文	37
3.1.3 进程状态	38
3.1.4 PCB 内容	42
3.2 进程控制	45
3.2.1 进程创建原语	46
3.2.2 进程终止原语	47
3.2.3 进程阻塞与唤醒原语	48
3.2.4 进程挂起与激活原语	49
3.3 进程的同步与互斥	50
3.3.1 进程之间的制约关系	50
3.3.2 信号量机制	51
3.3.3 进程的同步与互斥实现	54
3.3.4 进程的同步与互斥实例分析	57
3.4 进程的通信	65
3.4.1 消息缓冲通信	65
3.4.2 信箱通信	67
3.4.3 管道通信	68
3.5 死锁	71
3.5.1 死锁的基本概念	71
3.5.2 死锁的预防与避免	73
3.5.3 死锁的检测与解除	77
3.6 线程	80
3.6.1 线程的概念	80

3.6.2	线程实现机制	82
3.6.3	线程与进程的比较	83
3.7	管程	84
3.7.1	管程概念	84
3.7.2	管程语法与应用	85
3.7.3	管程的主要特征	86
	习题	87
第4章	处理机调度	91
4.1	作业	91
4.1.1	作业和作业步	91
4.1.2	作业控制块	92
4.1.3	作业状态及相互转换	93
4.2	处理机调度层次	93
4.2.1	高级调度	94
4.2.2	中级调度	94
4.2.3	低级调度	95
4.3	进程调度目标和调度方式	96
4.3.1	进程调度目标	96
4.3.2	进程调度方式	97
4.4	调度队列模型和选择调度算法的准则	98
4.4.1	调度队列模型	98
4.4.2	选择调度算法的若干准则	99
4.5	调度算法	101
4.5.1	先来先服务调度算法	102
4.5.2	短进程优先调度算法	102
4.5.3	高优先级优先调度算法	102
4.5.4	最高响应比优先调度算法	103
4.5.5	基于时间片的轮转调度算法	103
4.5.6	多级反馈队列调度算法	104
4.6	实时调度	105
4.6.1	实时调度的基本条件	106
4.6.2	实时调度算法的分类	107
4.6.3	常用的实时调度算法	108
	习题	110
第5章	存储管理	112
5.1	存储管理的功能	112
5.1.1	内存空间的分配和回收	112
5.1.2	地址变换	113

5.1.3	内存空间的共享和保护	114
5.1.4	内存空间的扩充	115
5.2	存储器的层次结构	115
5.2.1	多级存储器结构	115
5.2.2	内存与寄存器	116
5.2.3	高速缓存和磁盘缓存	116
5.3	程序的装入和链接	117
5.3.1	装入与重定位	118
5.3.2	链接	120
5.4	分区存储管理	121
5.4.1	单一连续分区	122
5.4.2	固定分区	122
5.4.3	可变分区	124
5.4.4	动态重定位分区	126
5.5	覆盖技术与交换技术	127
5.5.1	覆盖技术	127
5.5.2	交换技术	128
5.6	页式存储管理	129
5.6.1	页式存储管理的基本原理	129
5.6.2	页式存储管理的地址变换机构	131
5.6.3	页式存储管理中逻辑地址到物理地址的计算	132
5.6.4	页表的硬件实现	132
5.6.5	页表的组织	134
5.7	页面置换算法	135
5.7.1	最佳置换算法和先进先出置换算法	135
5.7.2	最近最久未使用置换算法	138
5.7.3	Clock 置换算法	140
5.8	段式存储管理	143
5.8.1	段式存储管理的基本原理	144
5.8.2	段式存储管理系统地址变换过程	145
5.8.3	段式存储管理和页式存储管理的区别	146
5.8.4	段的共享与保护	146
5.9	段页式存储管理	150
5.9.1	段页式存储管理的基本原理	150
5.9.2	段页式存储管理的地址变换	150
5.9.3	段页式存储管理的特点	153
5.10	局部性原理和抖动问题	153
5.11	内存三大管理技术对比	154
	习题	155

第 6 章 文件管理	159
6.1 文件与文件系统	159
6.1.1 数据项和记录	159
6.1.2 文件分类	160
6.1.3 文件系统	161
6.1.4 文件操作	163
6.2 文件逻辑结构	164
6.2.1 文件逻辑结构类型	165
6.2.2 顺序文件	167
6.2.3 索引文件	168
6.2.4 索引顺序文件	169
6.2.5 堆文件	170
6.2.6 直接文件和哈希文件	170
6.3 文件物理结构	171
6.3.1 连续存储结构	171
6.3.2 链接存储结构	173
6.3.3 索引存储结构	175
6.3.4 存取设备、物理结构和存取方法之间的关系	178
6.4 目录管理	179
6.4.1 文件控制块、文件目录和目录文件	179
6.4.2 索引节点	180
6.4.3 目录结构	181
6.4.4 目录查询技术	185
6.5 文件存储空间管理	187
6.5.1 空闲表法和空闲链表法	187
6.5.2 位示图法	188
6.5.3 成组链接法	189
6.6 文件共享	190
6.6.1 基于索引节点的共享方式	190
6.6.2 利用符号链实现文件共享	191
6.6.3 绕道法	193
6.6.4 基本文件目录表法	193
6.7 文件保护	194
6.7.1 存取控制矩阵	195
6.7.2 存取控制表	195
6.7.3 用户权限表	196
6.7.4 口令	196
6.7.5 密码	196
习题	197

第7章 设备管理	199
7.1 I/O 系统	199
7.1.1 I/O 设备	199
7.1.2 设备控制器	201
7.1.3 I/O 通道	203
7.2 I/O 数据传输控制方式	204
7.2.1 程序直接控制方式	204
7.2.2 中断控制方式	205
7.2.3 直接存储器访问控制方式	207
7.2.4 通道控制方式	210
7.3 缓冲管理	214
7.3.1 缓冲的引入	214
7.3.2 缓冲的种类	216
7.4 I/O 软件	222
7.4.1 I/O 软件的设计目标和原则	223
7.4.2 中断处理程序	225
7.4.3 设备驱动程序	227
7.4.4 设备独立性	229
7.4.5 用户层的 I/O 软件	231
7.5 设备管理	231
7.5.1 设备分配中的数据结构	231
7.5.2 设备分配时应考虑的因素	233
7.5.3 独占设备的分配程序	235
7.5.4 SPOOLing 技术	236
7.6 中断技术	238
7.7 磁盘存储器的管理	239
7.7.1 磁盘性能简述	239
7.7.2 磁盘调度	242
习题	246
参考文献	249

1.1 操作系统的形成与地位

当今计算机系统与智能手机的使用都离不开操作系统(Operating System, OS)。可以说,每位计算机用户都是通过某种操作系统去使用计算机或智能手机的,都需要基本掌握某种操作系统的操作方法以及系统调用。由此可见,操作系统的地位是十分重要的。

1.1.1 操作系统的形成

从计算机系统组成的角度看,一个完整的计算机系统是由硬件系统和软件系统两大部分组成的,如图 1-1 所示。硬件系统包括构成计算机系统的各种物理设备,比如控制器、运算器、存储器、外部设备等。软件系统是指计算机系统中使用的各种程序组成的系统。没有任何软件支持的计算机称为“裸机”。让用户直接面对裸机工作是十分困难的,同时用户编程时也不想涉足硬件的具体细节。因此,隐蔽对硬件的复杂操作,建立起一个服务体系,为用户提供良好的操作环境和服务功能就成为计算机系统的主要任务之一。在计算机的硬件之上覆盖一层层的管理软件,由内层向外层提供某种服务,每经过一层覆盖,系统功能便会增强一个级别。到了用户层,“裸机”就扩展成了一台操作界面简单、功能强大的计算机。

软件系统按其所面向的对象和着眼点的不同而分为两大类:面向计算机本身功能进行组织管理、维护,从而简化用户在各个环节上工作的软件,称为系统软件;而面向用户、为用户解决各种具体实际问题的软件,称为应用软件。应用软件需要在系统软件创造的适当环境下运行。操作系统就是系统软件中的一种,它是最基本的系统软件,是其他软件在计算机上运行的基础,而且可以说它是系统软件的核心。操作系统以外的其他系统软件统称为系统实用软件,例如编辑软件和编译软件等。

操作系统是最基本的系统软件,也是对硬件系统功能的第一次扩充。操作系统直接控制和管理所有的系统硬件,也为其他系统软件和应用软件提供基本的支持环境。当代的计算机都离不开操作系统。

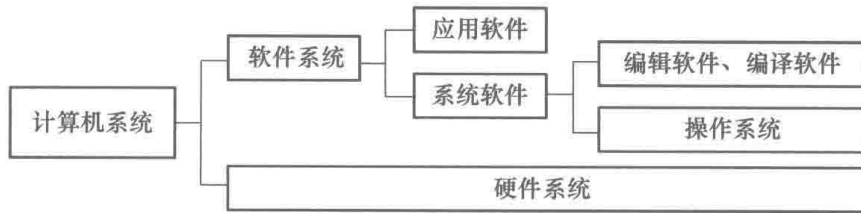


图 1-1 计算机系统组成

1.1.2 操作系统的地位

图 1-2 展示了计算机系统的层次结构。根据与用户距离的不同,从里向外(或从裸机到用户)分别是:硬件、操作系统、系统实用软件、应用软件。操作系统是最靠近硬件的一个层次,它控制和管理着内层的硬件系统,也控制和管理着外层的系统实用软件和应用软件,为其他软件提供了良好的开发与运行环境;并与各种系统实用软件协作,从而使各种应用软件得以开发和正常、高效率地运行。而从用户的角度讲,操作系统则是用户与计算机之间的接口。上述这些层次既相互独立,又紧密相连、互相依赖,形成完整的计算机系统,完成各种信息处理任务。

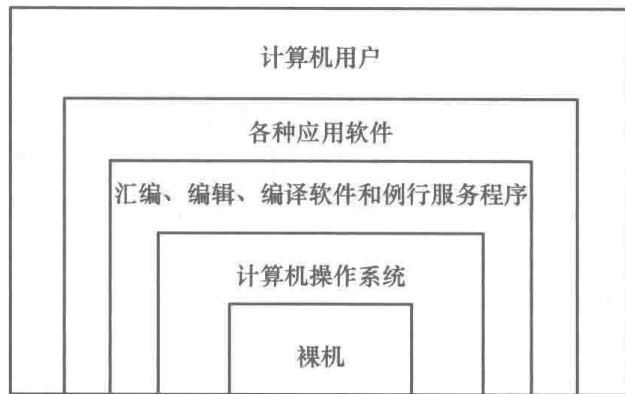


图 1-2 计算机系统的层次结构

从图 1-3 可知,操作系统是其他软件与硬件的接口和桥梁:其他应用程序的各种命令需要被操作系统翻译并转到硬件进一步处理,硬件处理后的结果同样需要通过操作系统转交给各种应用程序。

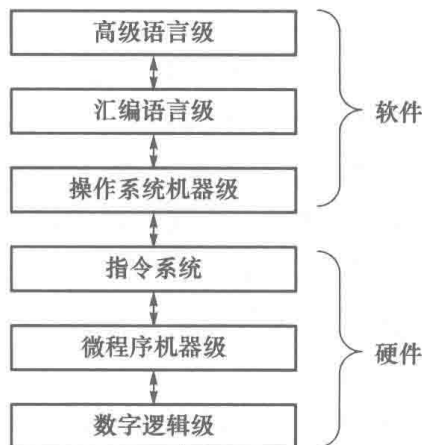


图 1-3 操作系统所处地位

从图 1-4 可知,应用用户无须了解操作系统的实现细节,即这些细节对应用用户来说是透明的;应用开发人员在编写各种应用程序时需要调用操作系统的应用程序接口(Application Programming Interface, API),从而完成各种系统调用功能;操作系统开发人员必须了解操作系统的工作原理,并需要了解各种硬件的实现细节。

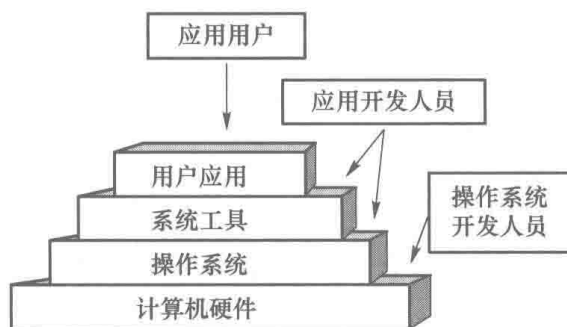


图 1-4 操作系统与使用人员的关系

1.2 操作系统的发展阶段

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,逐步形成和发展起来的。随着机器硬件功能和性能的不不断提高,机器的处理能力越来越强,提供的资源越来越多。如果仍然采用单个用户独占一台计算机的使用方式,势必造成相当大的浪费。为了更加有效地使用机器的资源,希望有多个用户同时利用机器来完成各自的工作,即各个用户同时使用不同的资源[例如外部设备、中央处理器(Central Processing Unit, CPU)等],既相对独立,又彼此协调。正是在这种实际要求面前,操作系统才得以问世。下面介绍操作系统形成过程中的几个重要阶段。

1.2.1 手工操作系统阶段

20 世纪 50 年代末,也就是第一台计算机问世若干年后,并没有出现操作系统。那时的计算机运行速度慢,外部设备少,程序的装入、调试及控制程序的运行等工作都由手工完成。计算机工作时,由操作人员把程序纸带(或卡片)装入输入机,输入机把程序和数据输入计算机后,操作人员启动程序,运行完后取走结果,卸下纸带(或卡片)。然后,下一个用户才能上机。

这种手工操作方式的缺点如下。

(1) 用户独占计算机的全部系统资源。一台计算机被一个用户独占,系统中的全部资源由他一人支配。尽管用户可以较方便地使用各种资源,不会出现因资源已被其他用户占用而等待的现象,但资源利用率非常低。

(2) CPU 的利用率低,大量的 CPU 时间空闲,作业运行过程需人工干预

CPU 需要等待人工操作,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;在计算完成后,进行卸带取卡操作时,CPU 又空闲。可见,CPU 的利用极不充分,这在运行短程序时尤为突出。

可见,手工操作方式严重地降低了资源的利用率,此即人-机矛盾。随着 CPU 速度的提高,CPU 和输入/输出(Input/Output, I/O)设备间不匹配的矛盾日益严重。为缓和此矛盾,

必须摆脱手工干预,实现作业的自动过渡,因此出现了批处理系统。

1.2.2 早期批处理系统阶段

在计算机发展的早期阶段,用户上机时需自己建立和运行作业,并作结尾处理。为了缩短作业的建立时间,人们研制了监督程序,它是一小段常驻内存(primary storage)的核心代码。当若干用户作业合成一个作业运行序列时,监督程序自动地依次运行。早期的批处理可分为两种方式:联机批处理和脱机批处理。

1. 早期批处理方式

(1) 联机批处理

联机输入/输出是指程序和数据的输入/输出都是由主机控制的,即慢速的 I/O 设备是直接和主机相连的。联机批处理下,作业的运行过程大致为:

- ① 用户提交作业;
- ② 作业被做成穿孔纸带或卡片;
- ③ 操作人员有选择地将若干作业合成一批,通过输入设备(输入机或读卡机)把它们存入磁带;
- ④ 监督程序读入一个作业(若系统资源能满足该作业要求);
- ⑤ 从磁带调入汇编程序或编译程序,将用户作业源程序翻译成目标代码;
- ⑥ 连接装配程序,把编译后的目标代码及所需的子程序装配成一个可运行的程序;
- ⑦ 启动运行;
- ⑧ 运行完毕,由善后处理程序输出计算结果;
- ⑨ 再读入一个作业,重复步骤④~⑧;
- ⑩ 等完成一批作业后,返回步骤③,处理下一批作业。

联机批处理系统实现了作业的自动过渡,同手工操作系统相比,计算机的使用效率提高了。但在这种批处理系统中,作业的输入/输出是联机的,也就是说作业从输入机到磁带,再由磁带调入内存,以至结果的输出打印都是由 CPU 直接控制的。在这种联机操作方式下,虽然解决了作业自动转接的问题,减少了作业建立和人工操作的时间,但是随着 CPU 速度的不断提高,CPU 和 I/O 设备之间的速度差距就形成了一对矛盾,即在运行结果的输出过程中,CPU 仍处于停止阻塞状态。所以,在联机批处理系统中,慢速的 I/O 设备和快速主机之间仍处于串行工作状态,CPU 时间仍有很大的浪费。

(2) 脱机批处理

脱机输入/输出是指程序和数据的输入/输出都是在外围机(卫星机)的控制下完成的,或者说它们是脱离主机进行的。脱机批处理系统由主机和外围机组成。在一台外围机的控制下,预先将用户程序和数据从低速设备输入磁带或磁盘,当 CPU 需要这些数据时,再直接从磁带或磁盘高速地调入内存,大大地加速了输入过程。类似地,当 CPU 需要输出时,可立即将输出数据送到磁带或磁盘上,再在外围机的控制下,把磁带或磁盘上的处理结果通过相应的输出设备输出,大大加速了数据的输出过程。脱机输入/输出过程如图 1-5 所示。

2. 批处理技术

在早期的脱机 I/O 方式中,事先把一批作业输入磁带,这意味着作业的处理是成批进行的;为使这一批作业能自动连续地进行处理,在系统中还配置了监督程序。在监督程序的控制下,系统先把磁带上第一个作业装入内存,并将运行的控制权交给该作业,当该作业被处理完

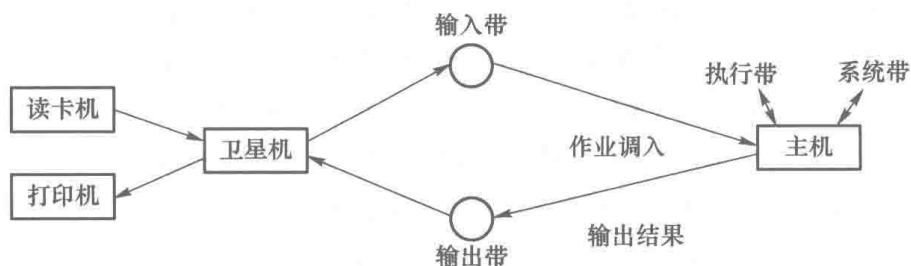


图 1-5 脱机输入/输出过程

后,系统又把控制权还给监督程序,由监督程序将第二个作业装入内存。这样,作业一个个地自动进行处理,直到磁带上的作业全部完成,这就是早期的批处理系统。可见,批处理系统是在解决人机矛盾和 CPU 与 I/O 设备速度不匹配矛盾的过程中发展起来的,或者说,批处理技术旨在提高系统吞吐量(system throughput,指系统在单位时间内所完成的作业数目)和资源的利用率。该系统的主要特征如下。

(1) 自动性(减少了手工操作):实现了作业的自动过渡,改善了 CPU 与 I/O 设备的使用情况,提高了计算系统的处理能力。

(2) 顺序性:磁带上的各道作业是按顺序进入内存的,在正常情况下,各道作业完成的顺序与它们进入内存的顺序应当完全相同,即先调入内存的作业先完成。

(3) 单道性:在某一时刻,内存中仅有一道程序在运行,仅当该程序完成或发生异常情况时,才将后继程序调入内存继续运行。

早期批处理技术的主要缺点如下。

(1) 平均周转时间长。所谓周转时间,是指从作业进入系统到作业完成所经历的时间。在批处理系统中,一个作业运行完成才能开始下一个作业,这必然使许多短时作业因为等待时间增加,周转时间显著增长。

(2) 不能提供交互功能,用户使用机器不方便。在批处理运行过程中,无法实现人机交互,必须等正在运行的批作业运行完毕或因出现错误使运行终止,用户才可以与机器交互。

(3) CPU 利用率较低(单道、串行)。在进行批处理过程中,监督程序、系统程序和用户程序之间存在着一种调用关系,任何一个环节出了问题,整个系统都会停顿;用户程序也可能会损坏监督程序和系统程序,这时,只有操作人员进行干预才能使程序恢复正常使用。后来,通道和中断技术的使用,使操作系统进入运行系统阶段。运行系统是常驻内存的监督程序,不过其功能扩大了,它不仅负责作业运行的自动调度,而且还要提供输入/输出的控制功能。但是,这时计算机系统运行的特征仍是单道顺序地处理作业,即用户作业仍然是一道一道地按顺序处理。这时,可能会出现以下两种情况:

① 对于以计算为主的作业,输入/输出量少,外部设备空闲(计算型作业);

② 对于以输入/输出为主的作业,又会造成主机空闲(I/O 型作业)。

总的来说,计算机资源的使用效率仍然不高。因此,操作系统进入了多道程序系统阶段,即多道程序合理搭配交替运行,充分利用资源,提高效率。

1.2.3 多道程序系统阶段

在早期批处理系统中,内存中只存放一道程序,称为单道运行。这种系统的管理很简单,不存在高度管理的问题。但是这种单任务系统对 CPU 的利用率极低,原因是:CPU 与外界交

换数据时,CPU 的速度很快,而外部设备的速度很慢,导致 CPU 大部分时间在等待外部设备的 I/O 操作。为了提高 CPU 的利用率,引入了多道程序技术。多道程序技术是指,将一个以上的程序放入内存,并允许它们交替运行,共享系统中的各种资源。当正在运行的程序因 I/O 操作而暂定运行时,CPU 立即转去运行另一道程序;当第二道程序又因 I/O 操作而暂定运行时,CPU 又转去运行第三道程序。显然,多道程序设计技术提高了 CPU 的利用率,同时也显著改善了内存和 I/O 设备的利用率,使系统吞吐量获得大幅度提高。允许多道程序运行的系统称为多道程序系统。现代计算机操作系统的设计一般都基于多道程序技术。

多道程序系统具有以下特征。

① 多道性:计算机内存中同时存放多个相互独立的程序。

② 无序性:多个程序完成的先后顺序与它们进入内存的顺序之间并无严格的对应关系,即先进入内存的程序可能较后甚至最后完成,而后进入内存的程序又可能先完成。

③ 调度性:作业从提交给系统开始直至完成,需要经过两次调度——作业调度和进程(process)调度。

④ 宏观上并行、微观上串行:在单核 CPU 系统中,进入系统的几道程序都处于运行状态,即它们先后开始了各自的运行,但都未运行完毕;实际上,各道程序轮流地使用 CPU 进行交替运行。

优点:①提高了 CPU、内存、I/O 设备的利用率;②系统的吞吐量显著提高。

缺点:平均周转时间长。

在批处理系统中采用多道程序设计技术,就形成了多道批处理系统。要处理的许多作业存放于外部存储器(secondary storage)中,形成作业队列,等待运行。当需要调入作业时,操作系统中的作业调度程序根据该批作业对资源的要求并按照一定的调度原则,调几个作业进入内存,让它们交替运行。当一个作业完成后,再调入一个或几个作业。使用这种处理方式时,内存中总是同时存在几道程序,系统资源得到比较充分的利用。

多道程序及运行程序(常驻内存、功能扩大了了的监控程序)的出现,标志着操作系统的初步形成。

1.2.4 新一代操作系统阶段

随着计算机技术的飞速发展,出现了智能计算和网络计算,并随之出现了微机操作系统、网络操作系统(network operating system)、智能手机操作系统和分布式操作系统(distributed operating system)等新型操作系统。在这个阶段,计算机系统和操作系统的发展都异常迅速,操作系统技术逐渐成熟,新的技术不断引入,使新一代操作系统具有了开放环境、高效数据处理、友好人机界面、强功能开发支持、网络互联与通信,以及多媒体处理等功能。

1.3 操作系统的定义与特征

1.3.1 操作系统的定义

一般来说,给出操作系统的准确定义是很困难的,许多关于操作系统的论著中也有着不同的提法。一个操作系统包括哪些部分、不包括哪些部分,也没有统一的规定。现在,操作系统



操作系统的概念、
特征与功能

已经成为一种软件产品,对于其使用范围,各生产厂商也有不同的界定。

操作系统是计算机系统最重要的系统软件,是其他软件的支撑。它管理着计算机的系统资源,并通过这种管理为用户提供公共和基本的服务,从而成为用户与计算机之间的接口。下面从不同的角度来阐述操作系统。

(1) 科普观点:操作系统是计算机系统的管理指挥机构和控制中心。

(2) 功能观点:操作系统是计算机的资源管理系统,负责对计算机的全部软、硬件资源进行分配、控制、调度和回收。通过系统注册表对软件资源实施有效管理,通过设备管理器对硬件资源实施有效管理。

(3) 用户观点:操作系统是用户使用计算机的一个界面。

(4) 管理员观点:操作系统是计算机工作流程得以自动高效运行的组织者,系统软、硬件资源合理协调的管理者。

(5) 软件观点:操作系统是由程序和数据集组成的大型系统软件。

综合上述观点,将操作系统定义为:操作系统是计算机系统中的一个系统软件,它是这样一些程序模块的集合——以尽量有效、合理的方式组织和管理计算机的软硬件资源,合理地组织计算机的工作流程,向用户提供控制程序运行的各种服务功能,使用户能够灵活、方便和有效的使用计算机,使整个计算机系统能高效地运行,是计算机与用户之间的接口。

1.3.2 操作系统的特征

不同的操作系统具有不同的特征,但它们都具有以下 4 个特征。

1. 并发性

并发(program concurrence)是指两个或多个活动在同一时间间隔内发生。在多道程序环境下,并发是指在一段时间内可有多道程序同时运行。假设计算机系统中同时有两个或两个以上程序存在,它们都已经开始运行而且都还没有结束运行,宏观上它们在同时运行,但微观上,同一个系统硬件(例如单 CPU)还是被几个程序轮流地使用。只不过 CPU 在运行完某个程序的某一条指令后,并不一定就接着运行该程序的下一条指令,而是很可能转而去运行另一个程序的一条指令。

在单 CPU 系统中,每一时刻仅能运行一道程序,因此并发性是宏观上的,而微观上这些程序在 CPU 上是交替运行的。在多 CPU 系统中,在运行程序的数量不超过 CPU(核)数的情况下,多个活动不仅在宏观上是并行的,而且在微观上也是并行的。在分布式系统中,多台计算机并存,使程序的并发性特征得到更充分的体现。

并行和并发是 2 个既相似又有区别的概念。并行是指两个或多个事件在同一时刻发生,而并发是指两个或多个事件在同一时间间隔内发生。

程序的并发运行,能够提高系统资源的利用率。

2. 共享性

资源共享是指系统中的软硬件资源不再为某个程序所独占,而是供多个用户程序共同使用。操作系统要在多个并发程序间通过调度来分配 CPU 时间,并且保证系统数据共享的正确性以及数据的完整性。

根据资源的属性不同,共享可分为互斥共享和同时共享两种方式。互斥共享是指系统中的资源虽能提供给多个进程使用,但在一段时间内却只允许一个进程访问该资源,这种资源又称为临界资源。同时共享是指在一段时间内,允许多个进程同时对资源进行访问,当然,这里