

高等院校精品课程系列教材

操作系统教程

谢旭升 朱明华 张练兴 李宏伟◎编著



Operating System

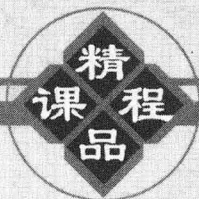


机械工业出版社
China Machine Press

高等院校精品课程系列教材

操作系统教程

谢旭升 朱明华 张练兴 李宏伟 编著



机械工业出版社
China Machine Press

操作系统是计算机系统的灵魂，是计算机系统必不可少的核心软件，是计算机专业工作者必须掌握的知识，因而操作系统课程成为计算机相关专业的必修课。本书系统地阐述了操作系统的基本概念、原理、技术、方法及功能，尽可能全面地展示操作系统的精髓和特性，力求做到概念清晰、结构合理、取舍得当、由浅入深、通俗易懂。

全书共分8章。第1章介绍操作系统的基本知识，第2章阐述处理器管理，第3章阐述作业管理，第4章阐述存储管理，第5章阐述设备管理，第6章阐述文件管理，第7章介绍Windows操作系统，第8章介绍Linux操作系统。

本书可作为高等院校计算机及相关专业的教材，也可作为计算机专业技术人员的参考书。

封底无防伪标均为盗版

版权所有，侵权必究

本书法律顾问 北京市展达律师事务所

图书在版编目（CIP）数据

操作系统教程 / 谢旭升等编著. —北京：机械工业出版社，2012.1
(高等院校精品课程系列教材)

ISBN 978-7-111-36896-0

I. 操… II. 谢… 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 (2011) 第 272313 号

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑：余洁

北京瑞德印刷有限公司印刷

2012 年 3 月第 1 版第 1 次印刷

185mm × 260mm · 20.25 印张

标准书号：ISBN 978-7-111-36896-0

定价：36.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88378991；88361066

购书热线：(010) 68326294；88379649；68995259

投稿热线：(010) 88379604

读者信箱：hzsj@hzbook.com

前言

操作系统是计算机系统中不可缺少的系统软件，在计算机专业的课程体系中占有重要的地位，是计算机专业及相关专业的一门必修课程，也是计算机专业工作者必须掌握的知识。因此，一本适用的教材对操作系统课程的学习显得特别重要。

本书是作者在多年教学工作的基础上，参阅了有关文献而编写的。考虑到本课程的课时数有限，我们对其内容进行了精选，着重于操作系统基本概念、基本原理、基本技术、基本方法的阐述，力求做到概念清晰、表述准确、结构合理、取舍得当、由浅入深、循序渐进、通俗易懂、便于自学，以期达到较好的教学效果。

全书共分8章。第1章引论，介绍了操作系统的基本概念、操作系统的形成与发展、操作系统的功能和分类、操作系统的特征和作用；第2章处理器管理，介绍了并发执行的特征以及进程与线程的概念，对进程控制、进程调度及算法、进程互斥与同步、进程通信、进程死锁等问题进行了分析和讨论；第3章作业管理，介绍了作业的概念、作业控制方式、作业管理和用户接口，着重讨论了批处理作业调度算法；第4章存储管理，介绍了存储管理的基本知识，讨论了存储管理的基本功能及各种存储管理技术和虚拟存储管理技术；第5章设备管理，介绍了I/O系统结构、I/O控制方式和I/O缓冲技术，对设备的分配、磁盘存储的管理、驱动调度及算法、虚拟设备技术进行了讨论；第6章文件管理，介绍了文件及文件系统的概念，对文件目录、文件组织、文件存储、文件操作、文件保护与保密等问题进行了分析和讨论；第7章Windows操作系统，主要介绍了Windows系统的实现技术；第8章Linux操作系统，主要介绍了Linux系统的实现技术。

本书由谢旭升、朱明华、张练兴、李宏伟编写。由于编者水平有限，书中难免会有错误和不当之处，恳请读者批评指正。

编者

2011年10月

教学建议

教 学 内 容	学习要点及教学要求	课 时 安 排	
		全部讲授	部分选讲
第1章 引 论	• 掌握操作系统的定义 • 了解操作系统的形成和发展过程 • 了解操作系统的分类及其特点 • 掌握操作系统的功能 • 了解操作系统的特征和作用 • 掌握多道程序设计技术	4	2
第2章 处理器管理	• 了解程序的顺序执行和并发执行特征 • 掌握进程的概念及其状态转换过程 • 了解进程控制过程 • 掌握进程调度的功能和进程调度算法 • 了解线程的概念和主要特征 • 了解临界区概念、信号量和PV原语 • 掌握PV操作实现进程互斥案例 • 掌握PV操作实现进程同步案例 • 了解管程的概念及其简单应用 • 了解进程通信类型 • 掌握死锁的定义、产生的原因和必要条件	20	16
第3章 作业管理	• 了解作业、作业步和作业调度的基本概念 • 了解作业管理的主要功能 • 掌握批处理作业的调度过程 • 掌握批处理作业的典型调度算法和应用 • 了解批处理作业的控制过程 • 了解交互式作业的管理过程 • 了解操作系统与用户的三种接口机制	8	6
第4章 存储管理	• 了解存储管理的主要功能 • 掌握程序装入与链接、逻辑地址与物理地址、内存保护等概念 • 了解交换与覆盖技术 • 掌握连续分配管理方式（单一连续分配和分区分配） • 掌握非连续分配管理方式（分页管理方式、分段管理方式、段页式管理方式） • 掌握虚拟存储器的基本概念 • 掌握请求分页管理方式，常用的页面置换算法 • 理解抖动现象、工作集的概念 • 了解请求分段管理方式 • 了解请求段页式管理方式	12	10

(续)

教 学 内 容	学习要点及教学要求	课 时 安 排	
		全部讲授	部分选讲
第 5 章 设备管理	- 了解设备管理的主要功能 - 了解 I/O 控制方式 - 了解缓冲技术实现的基本思想, 了解高速缓存与缓冲区技术 - 了解设备的分配和回收 - 掌握磁盘组织与管理, 理解掌握磁盘调度算法 - 掌握虚拟设备的实现思想, 理解假脱机技术 (SPOOLing 技术)	8	6
第 6 章 文件管理	- 了解文件的定义、分类和层次结构 - 了解文件系统的功能 - 掌握文件的逻辑结构和物理结构 - 掌握一级、二级、多级目录结构及功能 - 掌握空闲块表、空闲块链、位示图和成组链接方法等辅存空间管理技术 - 了解文件的操作和使用 - 掌握目录链接、基于索引结点的链接和符号链接三种文件共享技术 - 掌握文件的保护和保密技术	10	8
第 7 章 Windows 操作系统	- 了解 Windows 系统组成和设计目标 - 掌握 Windows 系统的处理器管理机制 - 了解 Windows 系统的内存管理基础知识 - 了解 Windows 系统常用的文件系统 - 了解 Windows 系统的 I/O 系统	4	2
第 8 章 Linux 操作系统	- 了解 Linux 系统的发展历史以及基本特点 - 掌握 Linux 系统的进程管理机制 - 掌握 Linux 系统的文件系统结构 - 掌握 Linux 内存管理的实现机制 - 了解 Linux 系统的 I/O 系统基础知识	4	2
合 计		70	52

目 录

前言

教学建议

第 1 章 引论	1
1.1 操作系统概念	1
1.2 操作系统的形成与发展	2
1.2.1 手工操作阶段	3
1.2.2 早期批处理系统	3
1.2.3 执行系统	4
1.2.4 多道批处理系统	5
1.2.5 分时系统	7
1.2.6 实时系统	8
1.2.7 通用操作系统	8
1.2.8 微机操作系统	9
1.2.9 网络操作系统	9
1.2.10 分布式操作系统	10
1.2.11 嵌入式操作系统	10
1.3 操作系统的分类	12
1.4 操作系统的功能	12
1.4.1 处理器管理	12
1.4.2 作业管理	12
1.4.3 存储管理	13
1.4.4 设备管理	13
1.4.5 文件管理	13
1.5 操作系统的特征和作用	13
1.5.1 操作系统的特征	13
1.5.2 操作系统的作用	14
1.6 本章小结	15

习题一	15
-----------	----

第 2 章 处理器管理	17
2.1 程序的顺序执行和并发执行	17
2.1.1 程序的顺序执行	17
2.1.2 程序的并发执行	18
2.2 进程的概念	19
2.2.1 进程的定义	19
2.2.2 进程的基本状态和转换	21
2.2.3 进程控制块	22
2.2.4 进程队列	23
2.3 进程控制	25
2.3.1 进程创建	25
2.3.2 进程撤销	26
2.3.3 进程阻塞与唤醒	26
2.4 进程调度	28
2.4.1 进程调度的功能	28
2.4.2 进程调度的时机	28
2.4.3 进程调度的算法	29
2.4.4 进程调度算法的选择	31
2.5 线程	32
2.5.1 线程的引入	32
2.5.2 线程的定义	32
2.5.3 线程的状态	33
2.5.4 线程的调度	33
2.5.5 线程的特征	33
2.5.6 线程的分类	34
2.5.7 线程与进程结构	35

2.6 进程互斥	36	3.1.3 作业控制方式	70
2.6.1 并发进程	36	3.2 批处理作业的管理	71
2.6.2 与时间有关的错误	36	3.2.1 批处理作业输入	72
2.6.3 临界区	38	3.2.2 批处理作业的调度	73
2.6.4 进程的互斥	39	3.2.3 作业调度算法	77
2.7 进程同步	43	3.2.4 算法综合应用	83
2.7.1 同步的概念	43	3.2.5 批处理作业的控制	92
2.7.2 用PV操作实现进程的同步 ..	44	3.3 交互式作业的管理	94
2.7.3 生产者-消费者问题	45	3.3.1 交互式作业的控制	94
2.7.4 读者-写者问题	49	3.3.2 终端作业的管理	95
2.7.5 时间同步问题	50	3.4 操作系统与用户的接口	96
2.8 管程	51	3.4.1 命令接口(作业级接口)	96
2.8.1 管程的概念	51	3.4.2 程序接口	105
2.8.2 管程的语法描述	52	3.4.3 图形接口	106
2.8.3 条件变量	53	3.5 本章小结	110
2.8.4 利用管程解决生产者-消费者 问题	53	习题三	110
2.8.5 利用管程解决哲学家进餐 问题	54		
2.9 进程通信	55	第4章 存储管理	113
2.9.1 进程通信的类型	55	4.1 存储管理概述	113
2.9.2 直接通信	56	4.1.1 存储器的存储结构	113
2.9.3 间接通信	57	4.1.2 存储管理的功能	114
2.10 死锁	58	4.2 程序的装入与链接	115
2.10.1 死锁的原因	58	4.2.1 物理地址和逻辑地址	115
2.10.2 死锁的必要条件	59	4.2.2 程序的装入	116
2.10.3 解决死锁的方法	60	4.2.3 程序的链接	118
2.11 本章小结	66	4.3 连续存储管理	119
习题二	66	4.3.1 单一连续存储管理	119
		4.3.2 固定分区存储管理	120
第3章 作业管理	70	4.3.3 可变分区存储管理	123
3.1 作业管理概述	70	4.3.4 覆盖与交换技术	130
3.1.1 作业	70	4.4 页式存储管理	132
3.1.2 作业步	70	4.4.1 基本原理	132
		4.4.2 存储空间的分配与去配	132
		4.4.3 页表与地址转换	133

4.4.4 快表	134	5.5.1 磁盘结构	175
4.4.5 页的共享与保护	135	5.5.2 磁盘空间的管理	178
4.5 段式存储管理	135	5.5.3 驱动调度	178
4.5.1 基本原理	136	5.5.4 提高磁盘 I/O 速度的一些 方法	183
4.5.2 空间的分配与去配	136	5.6 设备处理	185
4.5.3 地址转换与存储保护	137	5.6.1 设备驱动程序的处理过程 ...	185
4.5.4 段的共享	138	5.6.2 设备的中断处理	186
4.5.5 分页和分段存储管理的主要 区别	139	5.7 虚拟设备	187
4.6 段页式存储管理	139	5.7.1 脱机外围设备操作	187
4.7 虚拟存储管理	140	5.7.2 联机同时外围设备操作	188
4.7.1 虚拟存储器	141	5.7.3 SPOOLing 应用实例	189
4.7.2 请求分页式存储管理	142	5.8 本章小结	190
4.7.3 请求分段式存储管理	153	习题五	191
4.7.4 请求段页式存储管理	154		
4.8 本章小结	154	第 6 章 文件管理	193
习题四	155	6.1 文件管理概述	193
第 5 章 设备管理	159	6.1.1 文件和文件系统	193
5.1 设备管理概述	159	6.1.2 文件的分类	194
5.2 I/O 系统	160	6.1.3 文件系统的功能	195
5.2.1 I/O 系统结构	160	6.1.4 文件系统的层次结构	196
5.2.2 I/O 控制方式	165	6.2 文件的组织结构和存取方式	197
5.3 缓冲技术	169	6.2.1 文件的逻辑结构	197
5.3.1 缓冲的引入	169	6.2.2 文件的存储介质	199
5.3.2 单缓冲	170	6.2.3 文件的存取	199
5.3.3 双缓冲	171	6.2.4 文件的物理结构	200
5.3.4 多缓冲	171	6.2.5 记录的成组和分解	205
5.3.5 缓冲池	171	6.3 目录管理	206
5.4 独占设备的分配	172	6.3.1 文件目录	207
5.4.1 设备的逻辑号和物理号	173	6.3.2 文件目录结构	207
5.4.2 设备的独立性	173	6.4 辅存空间的管理	212
5.4.3 独占设备的分配	174	6.4.1 空闲块表法	212
5.5 磁盘管理	175	6.4.2 空闲块链法	213
		6.4.3 位示图法	214

6.4.4 成组链接法	216	7.5.2 文件系统	261
6.5 文件的使用	218	7.6 本章小结	264
6.5.1 主存打开文件表	218	习题七	264
6.5.2 文件基本操作	218	第8章 Linux 操作系统	266
6.5.3 文件的具体使用	220	8.1 Linux 系统概述	266
6.6 文件的共享	221	8.1.1 Linux 的产生及版本	266
6.6.1 绕道法	222	8.1.2 Linux 的特点	267
6.6.2 链接法	222	8.1.3 Linux 系统模型	267
6.7 文件的保护与保密	224	8.1.4 Linux 系统机制	268
6.7.1 存取控制	224	8.2 Linux 进程管理	271
6.7.2 容错技术	227	8.2.1 进程的数据结构	271
6.7.3 数据转储	230	8.2.2 进程和线程	273
6.8 本章小结	231	8.2.3 进程的调度	277
习题六	232	8.2.4 进程的通信机制	279
第7章 Windows 操作系统	234	8.3 Linux 内存管理	285
7.1 Windows 2000 系统概述	234	8.3.1 请求分页存储管理	285
7.1.1 设计目标	235	8.3.2 多级页表	287
7.1.2 系统模型	235	8.3.3 内存页的缺页中断	288
7.1.3 系统机制	239	8.3.4 内存空间的分配与回收	288
7.2 Windows 2000 进程管理	244	8.3.5 页面交换机制	290
7.2.1 进程	244	8.4 Linux 输入和输出系统	291
7.2.2 进程互斥和同步	245	8.4.1 设备管理概述	292
7.2.3 线程	247	8.4.2 I/O 控制方式	295
7.3 Windows 2000 内存管理	248	8.4.3 设备驱动程序	296
7.3.1 概述	248	8.4.4 高速缓冲机制	300
7.3.2 内存对象	251	8.5 Linux 文件系统	300
7.3.3 动态链接库	253	8.5.1 常见文件系统格式	301
7.3.4 页面文件	255	8.5.2 虚拟文件系统	301
7.4 Windows 2000 输入和输出系统	256	8.5.3 EXT2 文件系统	304
7.4.1 I/O 系统的组成	257	8.5.4 日志文件系统	306
7.4.2 驱动程序	258	8.6 本章小结	307
7.5 Windows 2000 文件系统	259	习题八	308
7.5.1 磁盘管理	259	参考文献	310

第 1 章

引 论

1.1 操作系统概念

任何一个计算机系统都是由两部分组成：计算机硬件和计算机软件。计算机硬件通常是由中央处理器（包括运算器和控制器）、存储器、输入设备和输出设备等部件组成，它构成了系统本身和用户作业赖以活动的物质基础和环境。由这些硬件部件组成的机器称为“裸机”。

然而用户最不喜欢裸机这种工作环境，因为裸机上没有任何一种可以协助他们解决问题的手段，只提供最低级的机器语言。为了对硬件的性能加以扩充和完善，方便用户使用计算机，于是在裸机外添加了能实现各种功能的软件程序。例如，为了方便用户描述自己的算题任务而提供了程序设计语言及相应的翻译程序（汇编程序或编译程序）；为了方便、有效地解决各类问题，提供了各种服务性程序和实用程序，如系统程序库、编辑程序、链接装配程序等；为了维护系统正常工作，有查错程序、诊断程序和引导程序；还有用户应用程序、数据库管理系统、数据通信系统等。软件又称为软设备，它是程序和数据的集合。这些程序是为了方便用户和充分发挥计算机效能而编制的。在这些软件中有一个很重要的软件系统，即操作系统，它管理系统中所有的软、硬件资源，并组织控制整个计算机的工作流程。

软件一般可以分为以下几类：

- 1) 系统软件：操作系统、编译程序、程序设计语言、链接装配程序以及与计算机密切相关的程序。
- 2) 应用软件：应用程序、软件包（如数理统计软件包、运筹计算软件包等）。
- 3) 工具软件：各种诊断程序、检查程序、引导程序。

硬件是计算机系统的物质基础，没有硬件就不能执行指令和实施最原始、最简单的操作，软件也就失去了效用；而若只有硬件，没有配置相应的软件，计算机就不能发挥它潜在的能力，这些硬件也就没有了活力。因此，硬件和软件二者是互相依赖、互相促进的。可以这样说：没有软件的硬件是一具僵尸；而没有硬件的软件则是一个幽灵。只有软件和硬件有机地结合在一起的系统，才能称得上是一个完整的计算机系统。操作系统将系统中的各种软、硬件资源有机地组合成一个整体，使计算机真正体现了系统的完整性和可利用性。

计算机系统的组成用图 1-1 来描述。从图中可看出，硬件和软件以及软件各部分之间是层次结构的关系。硬件在最里层，它的外面是操作系统，经过操作系统提供的资源管理功能和方

使用户的各种服务手段把硬件改造成功能更强、使用更为方便的机器，通常称为“虚拟机”。而其他系统软件和应用软件则运行在操作系统之上，需要操作系统支撑。

因此，引入操作系统的目的可从两方面来考察：

1) 从系统管理人员的观点来看，引入操作系统是为了合理地组织计算机工作流程，管理和分配计算机系统硬件及软件资源，使之能为多个用户所共享。因此，操作系统是计算机资源的管理者。

2) 从用户的观点来看，引入操作系统是为了给用户提供一个良好的界面，以使用户无需了解许多有关硬件和系统软件的细节，就能方便灵活地使用计算机。

综上所述，我们可以把操作系统定义为：操作系统是计算机系统中的一个系统软件，它统一管理计算机的软、硬件资源和控制程序的执行。

操作系统的主要作用可归结为：

- 1) 方便用户使用：操作系统通过提供用户与计算机之间的友好界面来方便用户使用。
- 2) 扩展机器功能：操作系统通过扩充硬件功能和提供新的服务来扩展机器功能。
- 3) 管理系统资源：操作系统有效地管理系统中的所有硬件和软件资源，使之得到充分利用。
- 4) 提高系统效率：操作系统合理组织计算机的工作流程，以改进系统性能和提高系统效率。
- 5) 构筑开放环境：操作系统遵循国际标准来设计和构造一个开放环境。其含义主要是指：遵循有关国际工业标准和开放系统标准，支持体系结构的可伸缩性和可扩展性；支持应用程序在不同平台上的可移植性和互操作性。

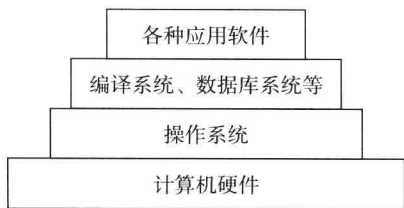


图 1-1 计算机系统的组成

1.2 操作系统的形成与发展

为了更好地理解操作系统的基本概念、功能和特点，让我们首先回顾一下操作系统形成和发展的历史过程。

操作系统由于客观的需要而产生，它伴随着计算机技术本身及其应用的日益发展而逐渐发展和不断完善。它的功能由弱到强，在计算机系统中的地位不断提高。至今，它已成为计算机系统的核心，但凡计算机系统都配置了操作系统。

世界上第一台电子计算机于 1946 年问世，此后，电子计算机在其运算速度、存储容量方面急剧上升，而价格、体积、热辐射和功耗却不断下降。人们通常按照元件工艺的演变把计算机的发展过程分为以下四个阶段：

- 1946 年至 20 世纪 50 年代末：第一代（电子管时代）；
- 20 世纪 50 年代末至 60 年代中期：第二代（晶体管时代）；
- 20 世纪 60 年代中期至 70 年代中期：第三代（集成电路时代）；
- 20 世纪 70 年代中期至今：第四代（大规模集成电路时代）。

现在电子计算机正向着巨型、微型、并行、分布、网络化、智能化几个方向发展。在此基础上为适应客观需要，操作系统也经历了如下的发展过程：手工操作阶段（无操作系统）、早期批处理系统、执行系统、多道批处理系统、分时系统、实时系统、通用操作系统、微机操作

系统、网络操作系统、分布式操作系统和嵌入式操作系统。

1.2.1 手工操作阶段

在第一代计算机时期,构成计算机的主要元件是电子管,其运算速度慢(只有几千次/秒)。计算机由主机(运控部件、主存)、输入设备(如读卡机)、输出设备(如穿孔机)和控制台组成。当时没有操作系统,甚至没有任何软件。

人们在利用这样的计算机解题只能采用手工操作方式,而且用户只能轮流地使用计算机。每个用户的使用过程大致如下:先把程序卡片装上读卡机,然后启动读卡机把程序和数据送入计算机,接着通过控制台开关启动程序运行。计算完毕,用户通过穿孔机输出结果。在这个过程中需要人工装卡片、人工控制程序运行、人工卸卡片,这些都是人工操作,即所谓“人工干预”。这种由一道程序独占机器且有人工操作的情况,在计算机运行速度较慢时是允许的,因为此时的计算时间相对较长,手工操作所占比例还不大。

20世纪50年代后期,计算机的运行速度有了很大的提高,从每秒几千次、几万次发展到每秒几十万次、上百万次。由于计算机运行速度几十倍、上百倍地提高,手工操作的慢速度和计算机的高速度之间产生了矛盾,即所谓人一机矛盾。随着计算机速度的提高,人一机矛盾已到了不能容忍的地步。为了解决这一矛盾,必须得摆脱人的手工干预,实现作业的自动过渡,这样就出现了批处理系统。

1.2.2 早期批处理系统

在计算机发展的早期,用户上机时需要自己建立和运行作业,并做结束处理,没有任何用于管理的软件,所有的运行管理和具体操作都由用户自己承担。为了缩短作业的建立时间,减少错误操作,尽可能地提高CPU的利用率,我们可以采取以下两个措施:首先配备专门的计算机操作员,程序员不再直接操作机器,以减少操作机器的错误;另一个重要措施是进行批处理,操作员把用户提交的作业进行分类,把一批中的作业编成一个作业执行序列,每一批作业将有专门编制的监督程序自动依次处理。

1. 联机批处理系统

用户上机前,需向机房的操作员提交程序、数据和一个作业说明书,后者提供了用户标识、用户想使用的编译程序以及所需的系统资源等基本信息。这些资料必须变成穿孔信息,(例如穿孔卡片的形式),操作员把各用户提交的一批作业装到输入设备上(若输入设备是读卡机,则该批作业是一叠卡片),然后在监督程序的控制下传送到磁带上。之后,监督程序自动输入第一个作业的说明记录,若系统资源能满足其要求,则将该作业的程序、数据调入主存,并从磁带上调入所需要的编译程序。编译程序将用户源程序翻译成目标代码,然后由链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执行的程序,接着启动执行。计算完成后输出该作业的计算结果。一个作业处理完毕后,监督程序又可以自动地对下一个作业进行处理。重复上述过程,直到该批作业全部处理完毕。

这种联机处理方式解决了作业自动转换,从而减少了作业建立和人工操作时间。但是在作业的输入和执行结果的输出过程中,CPU仍处于停止等待状态,CPU时间仍有很大的浪费,于是低速的输入/输出设备与高速的CPU之间形成了一对矛盾。如果把输入/输出工作直接交给价格便宜的专用机去做,就能充分发挥主机的效率,为此出现了脱机批处理系统。

2. 脱机批处理系统

脱机批处理系统由主机和卫星机组成，如图 1-2 所示。卫星机又称外围计算机，它不与主机直接连接，只与外部设备打交道。卫星机负责把输入机上的作业逐个传输到输入磁带上，当主机需要输入作业时，就把输入带与主机连上。主机从输入带上调入作业并运行，计算完成后，输出结果到输出磁带上，再由卫星机负责把输出带上的信息进行输出。在这样的系统中，主机和卫星机可以并行操作，二者分工明确，可以充分发挥主机的高速计算能力，因此与联机批处理系统相比，脱机批处理系统大大提高了系统的处理能力。

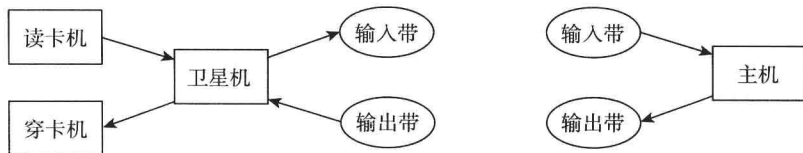


图 1-2 脱机批处理系统

批处理系统出现于 20 世纪 50 年代末，它是在解决人 - 机矛盾以及中央处理器的高速度和 I/O 设备的低速度之间矛盾的过程中发展起来的。它的出现促使了软件的发展，更重要的是它产生了具有管理作用的监督程序，该程序完成作业的自动过渡并且负责装入和运行各种语言翻译程序（如汇编程序、编译程序）以及实用程序（如链接装配程序）。在此期间也出现了程序库和程序覆盖等新的程序设计技术。解题操作过程变成了装入、汇编（或编译）、链接装配和执行四个步骤，从而使上机操作初步自动化。

1.2.3 执行系统

批处理系统实现了作业的自动过渡，它的出现改善了 CPU 和外设的使用情况，从而使整个计算机系统的处理能力得以提高。同时，它也存在一些缺点，如磁带需人工拆卸，这样既麻烦又容易出错，另一个重要的问题是系统保护问题。在进行批处理的过程中，所涉及的监督程序、系统程序和用户程序之间是一种互相调用的关系。对于用户程序没有任何检查，若目标程序执行了一条非法停机指令时，机器就会错误地停止运行。此时，只有当操作员进行干预，在控制台上按下启动按钮后，程序才会重新启动运行。另一种情况是，如果一个程序进入死循环，系统就会踏步不前，更严重的是无法防止用户程序破坏监督程序和系统程序。

20 世纪 60 年代初期，硬件获得了两方面的进展，一是通道的引入，二是中断技术的出现，这两项重大成果促使操作系统进入执行系统阶段。

通道是一种专用处理部件，它能控制一台或多台外设工作，负责外部设备和主存之间的信息传输。它一旦启动就能独立于 CPU 运行，这样可使 CPU 和通道并行操作，而且 CPU 和各种外部设备也能并行操作。所谓中断是指当主机接到外部信号（如设备完成信号）时，马上停止原来的工作，转去处理这一事件，处理完毕，主机又回到原来的断点继续工作。

借助于通道和中断技术，输入/输出工作可在主机控制下完成。这时，原有的监督程序的功能扩大了，它不仅负责调度作业自动地运行，而且还要提供输入/输出控制功能（用户不能直接使用启动外设的指令，输入/输出请求必须通过系统去执行）。这个发展了的监督程序常驻主存，称为执行系统。

执行系统实现的是联机操作，和早期批处理系统不同的是：输入/输出工作是由在主机控制下的通道完成的，主机和通道、主机和外设之间都可以并行操作。在执行系统中，用户程序

的输入/输出工作是委托给系统实现的,由系统检查其命令的合法性,以避免由于不合法的输入/输出命令造成对系统的威胁,因此提高了系统的安全性。另外,由于引入了一些新的中断,如算术溢出和非法操作码中断等,克服了错误停机的弊病,而时钟中断可以解决用户程序中出现的死循环现象。

1.2.4 多道批处理系统

早期批处理系统因为每次只调用一个用户作业程序进入主存并运行,故又称为单道批处理系统。其主要特征为:

1) 自动性。在顺利的情况下,在磁带上的一批作业能自动地逐个作业依次运行,而无需人工干预。

2) 顺序性。磁带上的各道作业是顺序地进入主存,各道作业完成的顺序与它们进入主存的顺序在正常情况下应当完全相同,亦即先调入主存的作业先完成。

3) 单道性。主存中仅有一道程序在运行,即监督程序每次从磁带上只调入一道程序进入主存运行,仅当该程序完成或发生异常情况时,才调入其后继程序进入主存运行。

图 1-3 说明单道程序运行时的情况。图中说明用户程序首先在 CPU 上进行计算,当它需要进行 I/O 传输时,向监督程序提出请求,由监督程序提供服务,并帮助启动相应的外部设备进行传输工作,这时 CPU 空闲等待。当外部设备传输结束时发出中断信号,由监督程序中负责中断处理的程序进行处理,然后把控制权交给用户程序,让其继续计算。

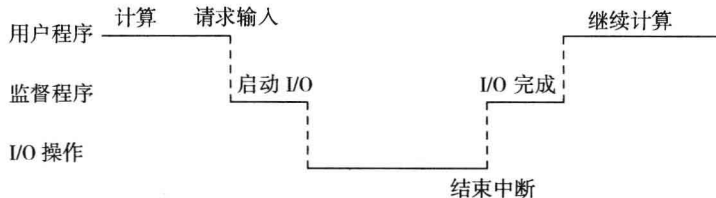


图 1-3 单道程序工作示例

从图 1-3 可以看出,当外部设备进行传输工作时,CPU 处于空闲等待状态;反之,当 CPU 工作时,I/O 设备又无事可做。这说明,计算机系统各部件的效能没有得到充分的发挥,其原因在于主存中只有一道程序。在计算机价格十分昂贵的 20 世纪 60 年代,提高设备的利用率是首要目标。为此,人们设想能否在系统中同时存放几道程序,这就引入了多道程序设计概念。

1. 多道程序设计

多道程序运行情况由图 1-4 来说明。图中,用户程序 A 首先在 CPU 上运行,当它需要从输入设备输入新的数据而转入等待时,系统帮助它启动输入设备进行输入工作,并让用户程序 B 开始计算。程序 B 经过一段计算后需要从输出设备输出一批数据,系统接收请求并帮助启动输出设备工作。如果此时程序 A 的输入尚未结束,也无其他用户程序需要计算,CPU 就处于空闲等待状态,直到程序 A 在输入结束后重新运行。若当程序 B 的输出工作结束时,程序 A 仍在运行,则程序 B 继续等待,直到程序 A 计算结束再次请求 I/O 操作时,程序 B 才能占用 CPU。

多道程序设计是一种软件技术,该技术使同时进入计算机主存的几个相互独立的程序在管理程序控制之下相互交替地运行。当某道程序因某种原因不能继续运行下去时(如等待外部

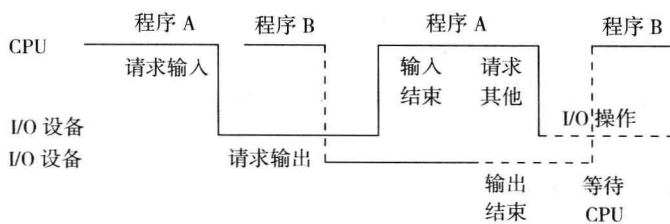


图 1-4 多道程序工作示例

设备传输数据)，管理程序便将另一道程序投入运行。这样可以使 CPU 及各外部设备尽量处于忙碌状态，从而大大提高计算机的使用效率。

引入多道程序设计，具有以下好处：

- 可提高 CPU 的利用率；
- 可提高主存和 I/O 设备利用率；
- 可增加系统吞吐量。

在单处理器系统中，多道程序运行的特点是：

- 1) 多道：计算机主存中同时存储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，主存中的多道程序轮流地或分时地占用 CPU，即多道程序交替执行。

2. 多道批处理系统

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。在多道批处理方式下，交到机房的许多作业由操作员负责将其由输入设备转存到辅存设备（比如磁盘）上，形成一个等待运行的作业队列。当需要调入作业时，管理程序中有一个名为“作业调度”的程序负责对磁盘上的一批作业进行选择，将其中满足资源条件且符合调度原则（比如按先来后到的顺序进行选择）的几个作业调入主存，让它们交替运行。当某个作业完成计算任务时，输出其结果，收回该作业占用的全部资源，然后根据主存和其他资源的情况再调入一个或几个作业。这种处理方式的特点是：在主存中总是同时存储几道程序，系统资源的利用率是比较高的。

要求计算机解决的问题是多种多样的，具有不同的特点。例如，科学计算问题需要使用较多的 CPU 时间，因为它计算量较大；而数据处理问题输入/输出量较大，需较多地使用输入/输出设备。若在调入作业时能注意到不同作业的特点，并能合理搭配，如将计算量大的作业和输入/输出量大的作业搭配，系统资源的利用率会进一步提高。

多道批处理系统是一种有效但又十分复杂的系统，为使系统中的多道程序能协调地运行，必须解决以下一些问题：

- 1) 并行运行的程序要共享计算机系统的硬件和软件资源，既有对资源的竞争，又必须相互同步。因此同步与互斥机制成为系统设计中的重要问题。
- 2) 多道程序的增加使之出现了主存不够用的问题，提高主存的使用效率也成为关键。因此出现了诸如覆盖技术、对换技术和虚拟存储技术等主存管理技术。
- 3) 多道程序存在于主存，为了保证系统程序存储区和各用户程序存储区的安全可靠，提

出了主存保护的要求。

多道批处理系统的出现标志着操作系统进入日趋成熟的阶段，并先后出现了作业调度管理、处理器管理、存储器管理、外部设备管理、文件系统管理等功能。

1.2.5 分时系统

在批处理方式下，用户以脱机操作方式使用计算机。用户上机前必须提交程序、数据和作业说明书，之后，用户就完全脱离了自己的作业，由系统对包含了该作业的一批作业进行自动处理，等该批作业全部处理结束后，用户才能拿到计算结果，并根据结果分析下一步处理方案，若有错，还得重复上述过程。这种工作方式给用户使用计算机带来了很大的麻烦，虽然计算机的效率较高。人们又十分留恋手工操作阶段那样的联机工作方式——用户独占机器，并且直接控制程序的运行，因此可以随时了解程序运行情况。但是，联机工作方式下，用户独占计算机造成计算机效率极低。那么，能否有一个两全之策，既保证计算机的效率，又方便用户的使用呢？

20 世纪 60 年代中期，计算机技术和软件技术的发展加速了这种可能变成现实。由于主机速度的不断提高和采用了分时技术，一台计算机可同时为多个终端用户服务，而每个用户可在自己终端设备上联机使用计算机，好像自己独占计算机一样。

所谓分时技术，就是把处理器的时间分成很短的时间片（如几百毫秒），这些时间片轮流地分配给各联机作业使用。如果某个作业在分配给它的时间片用完之时计算还未完成，该作业就暂时中断，等待下一轮继续计算，此时处理器让给另一个作业使用。这样，每个用户的各种要求都能得到快速响应，给每个用户的印象是好像他独占一台计算机。

采用这种分时技术的系统称为分时系统。在该系统中，一个计算机和许多终端设备连接。每个用户可以通过终端向系统发出各种控制命令，请求完成某项工作，而系统则分析从终端设备发来的命令，完成用户提出的要求，输出一些必要的信息，如给出提示信息、报告运行情况、输出计算结果等。用户根据系统提供的运行结果，向系统提出下一步请求。重复上述交互会话过程，直到用户完成预计的全部工作为止。

多道批处理系统和分时系统的出现标志着操作系统的初步形成。

1. 分时系统的类型

(1) 单道分时系统

在单道分时系统中，主存只驻留一道程序（作业），其余作业都保存在辅存上。每当主存中的作业运行一个时间片后，便被调至辅存（称为调出），再从辅存上选一个作业装入主存（称为调入）并运行一个时间片，依此方法使所有的作业都能在规定的时间内轮流运行一个时间片，这样便能使用户都能与自己的作业交互。由于单道分时系统只有一道作业驻留在主存，在多个作业的轮流运行过程中，每个作业往往可能频繁地调进调出多次，开销大，故使系统性能较差。

(2) 具有“前台”和“后台”的分时系统

在单道批处理系统中，作业调进调出时 CPU 空闲；主存中的作业在执行 I/O 操作时 CPU 也空闲。为了充分利用 CPU，我们引入了“前台”和“后台”的概念。在具有前、后台的系统中，主存被固定地划分为“前台区”和“后台区”两部分，“前台区”存放按时间片调进和调出的作业流，“后台区”存放批处理作业。仅当前台调进或调出、前台已无作业可运行时，