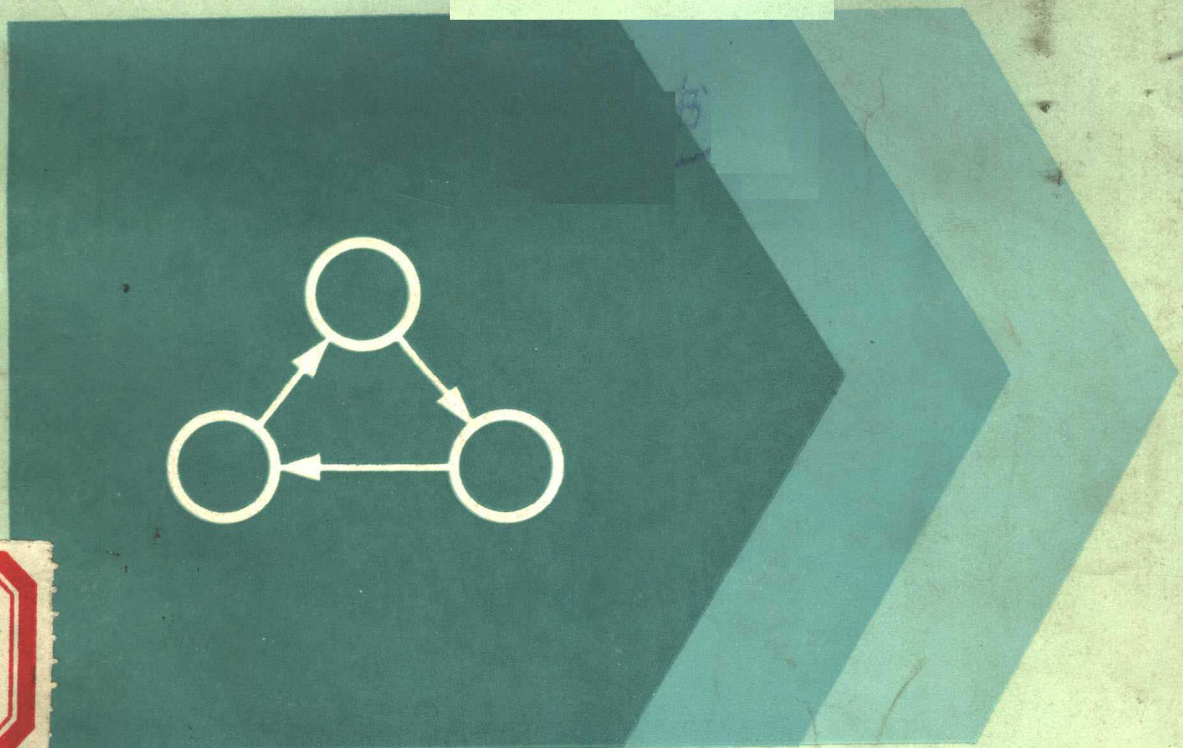
CAO ZUO XI TONG JI CHU JIAO CHENG

操作系统基础教程

徐甲同 编



西北电讯工程学院出版社

操作系统基础教程

徐甲同　编

西北电讯工程学院出版社

1987

内容简介

本书着重介绍计算机系统的一个重要的系统软件——操作系统的基本原理。全书共分九章：第一章为操作系统概述；第二章介绍操作系统的运行环境；第三章至第七章分别介绍操作系统的各种资源管理功能，即作业管理、进程管理、存储管理、设备管理和文件系统；第八章介绍操作系统的结构；第九章介绍一个实例——CP/M操作系统。

本书可作为大学专科计算机软件专业或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。

操作系统基础教程

徐甲同　编

责任编辑　王绍菊

西北电讯工程学院出版社出版发行

西北电讯工程学院印刷厂印刷

开本 787×1092 1/16　印张 16　字数 387 千字

1987 年 4 月第 1 版　1987 年 4 月第 1 次印刷　印数 1—5 000

ISBN7-5606-0015-8/TP·0007

统一书号：15322·87

定价：2.70 元

前　言

自1978年以来，作者一直为大专院校的本科生、专科生、研究生以及各种计算机的用户讲授操作系统课程。在总结多年教学经验的基础上，结合使用、分析、设计操作系统的实践，重新编写了本教材。

操作系统是用户和计算机系统之间的接口，是计算机软件和硬件之间的接口，它涉及的内容较为广泛，可以说是一门综合性的课程。因此，读者在阅读本书之前，应具备计算机原理、程序设计语言和数据结构等方面的基础知识。由于本书在叙述上力求由浅入深，语言通俗易懂，如果读者有了上述必要的基础知识，那么即使不通过讲授也能靠自学掌握操作系统的基本原理和基本技术。

本教材共分九章，第一章是操作系统概述，使读者在学习之前，对操作系统先建立一个总的概念，有一个总体的认识。第二章讲述操作系统的运行环境，介绍了必要的预备知识，为进一步学习操作系统铺平道路。第三章至第七章介绍操作系统资源管理的五大功能：作业管理、进程管理、存储管理、设备管理和文件管理系统，这几章是本书的核心内容。每一部分着重介绍了基本功能、实现原理和设计方法。把作业管理放在其它管理的前面讲述，一方面原因是把作业管理看成是处理机管理的一部分，另一方面的原因是它处于操作系统的最外层，与用户关系密切。第八章讲述操作系统结构，目的是为操作系统的设计打下基础。第九章介绍一个操作系统的实例——CP/M，选择CP/M的理由是它结构简单，程序短小，易于分析，易于理解。其它操作系统的内容分散在各章节中。由于本书的内容只涉及操作系统的基本概念、基本原理和基本方法，故本书取名为《操作系统基础教程》。

本教材按80学时编写，如按60学时讲授，可略去目录中带*号的章节。

在本教材的编写过程中，我系陈家正副教授在百忙中审阅了全部书稿，并提出了大量指导性的建议，在此谨致衷心的感谢。

由于编写时间紧迫，错误和不当之处在所难免，敬请读者批评指正。

作　者

1986年10月

目　录

第一章　操作系统概述

在进入操作系统这门课程学习之前，人们首先就会提出：操作系统“是什么”，“能干什么”以及“怎样干”等一系列问题。本章的目的是全面地介绍操作系统这一领域中的基本问题，使读者对操作系统有一个全面、概括的了解，以便有目的地、深入地学习它的主要内容。

§1.1　什么是操作系统

当今社会的科学技术发展如此之迅速，计算机技术的应用如此之普遍，甚至中、小学生也在开始学习或使用计算机。许多读者对计算机也是熟悉的。如果你使用过计算机，那么你也必定接触过操作系统。因为操作系统是计算机系统中必须配置的一种系统软件，几乎所有的计算机都离不开操作系统，而且操作系统在计算机系统中所处的举足轻重的地位，已经突出地显示出来了。

一、操作系统的定义

关于操作系统的严格定义有多种说法。不同的说法反映了不同作者从不同角度去揭示操作系统的本质。

首先从计算机系统的构成说起。大家知道，一个计算机系统，它由两部分构成：系统硬件和系统软件。系统硬件是构成计算机系统所必须配置的那部分设备，它为形成和组织一个系统提供了控制机构，是提供给操作系统的物质基础。系统软件是指计算机系统必须配备的那部分软件，它通常是对各种领域都适用的一些软件。诸如各种程序设计语言的处理程序，各种操作系统，标准程序库以及系统维护软件等。

由此可见，操作系统本身是系统软件的一部分，其物质基础是系统硬件。系统硬件和系统软件都可看作是计算机系统资源，而操作系统正是管理系统资源的机构。

简单说来，所谓操作系统实际上是一批组织在一起的程序和数据的集合，它专门用于计算机系统资源的管理，并使计算机的操作和程序设计得以简化，能控制用户程序在系统中的运行。

上述定义指出了操作系统“是什么”——程序和数据的集合，以及操作系统“能干什么”——简化计算机操作和程序设计，并对计算机系统资源进行管理。

总之，操作系统的根本任务是把计算机系统内的所有硬、软件资源有机地组织并管理起来，尽可能地充分发挥它们各自的效率，使整个系统向用户提供他们所需的各种服务。它是用户和计算机系统的接口，用户通过操作系统使用计算机。

二、系统资源

属于计算机系统的资源主要有硬件资源和软件资源

1．硬件资源

硬件资源指的是计算机系统硬件总和，它包括中央处理机(CPU)、主存储器(MS)、输入输出设备、外存储设备以及系统控制台。对于每一种设备都有一个"性能/价格比"指标，用以衡量它在系统中的经济效果。自然，操作系统最关心的是那些高性能/价格比的部件，例如中央处理机和主存储器。操作系统的目的之一就在于提高处理机效率和主存的利用率。

2．软件资源

软件资源，也称信息资源，主要指程序和数据。一个有效的操作系统以两种方式使这些信息资源的使用得以简化。

(1) 把信息预先存入计算机系统中，以供用户访问；

(2) 不同用户可共享同一信息资源。

于是，当两个或更多用户要求使用同一程序或数据时，应允许这些用户共同使用。因此，操作系统应提供信息管理的功能以及对存储这些信息的设备进行管理的功能。

三、操作系统的性能

一个操作系统的性能在很大程度上决定了一个计算机系统工作的优劣。到目前为止，还没有一个统一的标准来衡量一个操作系统的优劣，那么操作系统的性能是什么呢？

操作系统的性能包括如下两个方面：

1．系统效率

很明显，效率是操作系统的一个重要性能指标。

体现系统效率方面的指标是系统处理能力(Throughput)、各种资源的使用效率以及响应时间等。

(1) 系统处理能力。所谓一个系统的处理能力，是指在一给定的时间间隔内（比如一天或一小时）系统所完成的总工作量，包括处理大、中、小各类作业的数量。一般地说，处理能力是硬件速度和软件性能的综合。处理能力也称吞吐量。

(2) 各种资源的使用效率。各种资源的使用效率是指系统中各个部件、各种设备的使用程度。它反映了系统内资源的利用情况。如果一个系统中仅仅是中央处理机非常忙碌，而其它各种外部设备却常常空闲，那么这一系统的效率显然是不高的。

(3) 响应时间(Response Time)。在批量处理情况下，响应时间表现为作业的周转时间(Turnaround Time)，即从用户把作业提交给系统到用户收到该作业的处理结果的这一段时间，通常又称为解题周期。周转时间中常常包含着非系统因素（例如作业交接和操作的拖延等）。

在分时系统和实时系统中，响应时间更能反映系统效率。分时系统的响应时间是指用户通过终端发出命令到系统进行应答所需的时间。一般要求系统在3～5 s内作出响应，使用户感到比较满意。

2．系统的RAS

设系统(包括硬件和软件)的平均故障时间(Mean Time Between Failures)为MTBF，平均故障修理时间(Mean Time To Repair)为MTTR，则系统的RAS按下式定义：

$$R = MTBF$$

$$A = \frac{MTBF}{MTBF + MTTR}$$

$$S = MTTR$$

其中 R(Reliability)，A(Availability)，S(Serviceability) 分别称为系统的可靠性，可利用性，可维护性。

显然，如果一个系统经常由于硬件故障或软件故障而使系统停止工作，将影响到该系统的可利用性；如果平均故障时间长且平均维修时间短，则系统的可利用性就高。

(1) 系统的可靠性。用户总是希望在一个稳定、可以信赖的环境下工作，所以要求操作系统最好是绝对可靠的，不发生任何错误。但实际上不存在这样的系统，主要原因是系统中包含了大量的软硬件，至今还没有一种设计和实施技术确保它们不发生故障。此外，人为的误操作等原因也可能引起工作不正常。但是，如果在以下几个方面加以努力，也可指望产生一个可靠的操作系统：

① 在设计过程中采取各种技术措施，尽可能地避免软、硬件故障；

② 在系统运行过程中，一旦出错便能及时检测出来，以减少对系统造成的损害；

③ 对检测出来的错误，迅速查出错误原因和故障位置，并采取相应措施以排除故障；

④ 对系统进行有效的故障恢复。

(2) 可维护性。系统投入运行后，维护人员要对其进行经常性的维护。例如软件、硬件工作情况的测试，故障的检测和排除，文件的定期复制等。为了提高系统的可维护性，要求系统结构清晰，提供完整的说明文件以及配置较强的维护工具等。

§1.2 操作系统的形成和发展

为了帮助读者更好地理解操作系统，现在让我们回顾一下操作系统形成和发展的历史。操作系统是随着计算机的发展而形成并发展起来的。要介绍操作系统的发展过程的确切情况往往很难，因为有些重要概念早在普遍被接受之前就开始引入了。例如分页和虚存的概念，早在 1959 年就在 Atlas 系统中首次应用，但由于当时的物资条件不具备或其它原因，曾被埋没一段时间，到了六十年代中期又重新被几个一般系统采用。最后在 1972 年它被宣布作为 IBM 标准生产线的一部分。顺便指出，本章中出现的一些名词或术语，这里都没有给出确切的定义，只是用来介绍操作系统的概貌，这些名词或术语，将从第二章起给出明确定义和详细说明。

操作系统的形成和发展可划分为如下几个阶段：

一、手工操作阶段

在五十年代末期以前的第一代计算机中，操作系统尚未出现，那时只是手工操作。每个程序员都必须亲自动手操作计算机：装入卡片叠或纸带，按电钮，查看存储单元等。这时的手工操作过程大致如下：

用户或程序员向机房主管人员提出上机时间申请，当预约时间到达后，便可进入机房。首先须清除前一用户所遗留的作业信息，然后装上卡片或纸带，启动输入机。这时他便建立了自己的作业。接着他通过控制台开关启动程序，开始运行。在运行过程中，程序如果需要操作员干预，他就采取相应措施，之后再次启动运行。最后当作业完成时，卸下磁带，取走卡片或纸带，取出打印结果。下一作业再重复以上过程。

这种方式有两个突出缺点：

(1) 当一个用户开始操作后，全部计算机资源都归他占有，一直到他下机时才把这些资源转让给下一用户。

(2) 操作是联机的，输入输出也是联机的，因此，从上机到下机的操作时间拉得很长。

这种操作方式在计算机速度较慢的情况下是允许的。当计算机速度大大提高以后，就暴露了其严重缺点。譬如说，一个作业在速度为每秒一万次的计算机上运行，需 1h；而作业建立和手工操作的时间需要 3 min。这种情况下操作时间和运行时间的比为 1:20，若机器速度提高到每秒 60 万次，则此时的作业运行时间仅需 1min，而操作的速度不会有多大的改进，仍假定为 3 min。此时操作时间和运行时间的比为 3：1。这就是说，操作时间远远超过了机器运行时间。由此可见，缩短手工操作时间就显得非常必要了。

二、批量处理阶段

人们自然首先想到的就是一个作业到另一个作业的过渡，要摆脱人的手工干预，使其自动进行。批量处理又可分为两个阶段：

1. 早期的批量处理

为了缩短作业的建立时间，人们研制了监督程序。把很多的用户作业集中到一起，成批地进行处理，构成了作业自动执行序列，仅当一个作业处理结束后，利用监督程序来启动下一个作业。

早期的批量处理过程如下：

操作员把用户提交的若干个作业集中成为一批，并将其各个卡片叠放在读卡机上，由监督程序把这一批作业输入到磁带上，当该批作业输入完成之后，监督程序就开始执行。它自动地把磁带上该批的第一个作业调入内存，并对该用户程序进行汇编或编译。然后由装配程序把汇编或编译结果装入内存，再启动执行之。计算机完成全部计算或处理后，进行结果的输出。第一个作业全部完成之后，监督程序自动地调入该批的第二个作业，并重复执行以上过程，一直到该批作业全部完成为止。在完成了上一批作业之后，监督程序又从读卡机上输入另一批作业，保存在磁带上，并按上述步骤重复处理。这样，监督程序不停地处理各个作业，从而实现了作业到作业的自动转换，缩短了作业建立时间和手工操作时间。

2. 脱机批量处理

在早期批量处理中，作业的输入输出都是联机的。也就是说，作业信息由卡片送到磁带，再由磁带调入内存，以及计算结果在打印机上输出，这些都是由 CPU 来处理的，这种联机输入输出的缺点是速度慢。为此，在批量处理中引进了脱机输入输出技术。除主机外，另设一台卫星机，该机只与外部设备打交道，不与主机直接连接。如图 1.1 所示。读卡机上的作业卡通过卫星机输入到磁带上，而主机只负责从磁带上把作业调入内存，并予以执行。作业完成后，主机负责把结果输出到磁带上，然后再由卫星机把磁带上的信息在打印机上输出。

这样一来，输入输出工作脱离了主机，这里的卫星机的工作可以和主机的执行同时进行。这比早期的批量处理系统提高了处理能力。

批量处理出现在第二代计算机中，约在五十年代末期。它的出现又促进了其它软件的发展。其中主要有：

(1) 输入输出标准程序和程序库。在手工操作阶段，所有的输入输出指令都是由程序员

直接写入他的程序中，而采用脱机输入输出后，系统就必须提供一套输入输出程序供用户调用。

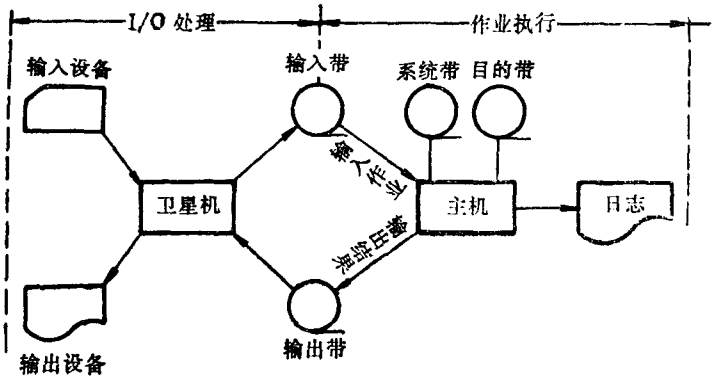


图 1.1　脱机批量处理

这时的系统程序越来越丰富，这就导致了程序库的建立。程序库中的程序称为库程序。库程序通常包括标准输入输出程序，汇编程序，FORTRAN 语言编译程序，装配程序，标准子程序以及善后处理程序等等。库程序通常放在磁带上，而只有监督程序放在内存中。

(2) 装配程序。由于一个用户程序无法单独运行，它必须调用一些库程序才能运行下去。因此，为了防止用户程序和系统程序之间、系统程序和系统程序之间的地址发生冲突，各程序应是可再定位的。换句话说，每个程序都必须按相对地址编写。这些系统程序可以按源程序的形式，也可以按其结果程序的形式存放在系统带上。当处理一个作业时，把用户作业的源程序翻译后送入执行带上去。所用的系统程序，若为结果形式，则立即送到执行带上去；若为源程序，则翻译后送到执行带上去。该作业在运行前，把执行带上的全部程序由装配程序进行装配，即把所有的结果程序装配成一个统一的可执行的绝对地址形式的目标程序，然后执行之。

(3) 覆盖技术。在目标程序的长度超过内存可用的容量时，程序员就得事先把程序和数据分成若干块，运行时逐块地调入内存并执行之。

(4) 运行日志和记帐。在批量处理中，为了记录系统的活动情况和用户程序的运行情况，往往要指定一台联机打印机，用于打印运行日志，包括运行过程中系统资源的利用情况以及使用该机的收费信息。

虽然批量处理系统克服了手工操作的一些缺点，并促进了软件的发展，但仍存在着一些根本问题没有解决。例如，监督程序、系统程序和用户程序之间是通过相互调用的办法进行转移的。因此，当目标程序企图执行一条非法指令时，整个系统就会停顿下来。若它陷入死循环，则整个系统也无法继续运行下去。更为严重的是它无法防止用户程序冲掉一部分监督程序，这样有可能把整个系统搞乱。如果出现这些情况，最后还得由操作员出面收拾。

三、执行系统阶段

在五十年代末期和六十年代初期，硬件获得了两方面的进展：一是通道的引进，二是中断的出现。

通道，即I/O数据通道，它是控制一台或多台外部设备的硬件机构。它一旦被启动，就独立于中央处理机而运行。这就能做到输入输出操作与主机并行工作。主机和通道的联络办法是由主机发出向通道的询问指令，询问通道工作完成否，若未完成，主机就循环询问，直至通道工作结束为止。

为使通道的数据传输能与主机并行并减少主机的等待时间，在程序设计中引进了缓冲技术。输入带上的信息提前送入输入缓冲区，要输出的信息也同样预先送到输出缓冲区。这样做，虽然也能减少询问等待时间，但并未解决根本问题。于是，在硬件中引进了中断技术。所谓中断，就是在输入输出结束时，或硬件发生某种故障时，由相应硬件向主机发出信号，主机立即停止正在执行的操作，转去处理中断请求。

为了获得主机操作与外部传输在时间上的重叠，就必须提供中断处理程序和输入输出控制服务(IOCS)，这样就把操作系统的概念从原来简单的监督程序扩大到执行系统。执行系统的程序可分三类：I/O系统、处理程序、管理程序。执行系统的程序这时相当庞大了，这些程序全部放在内存就会大大减少用户程序的可用空间，最好的办法是把所有的程序都要用到的中断处理程序和IOCS固定在内存，其它部分放在外存，常驻内存的部分称为执行程序。

执行程序对其它程序拥有控制权。用户程序的输入输出无例外地是通过委托执行程序来实现的。系统可对错误的I/O要求提供自动检查，受托程序完成之后，再用中断信号通知执行程序，这样便保证了系统的安全。此外，用户程序发生死循环也可以通过时钟中断进行检测处理。非法操作也通过非法操作中断得到及时处理。

随着执行系统的出现，在软件上又有了如下的一些进展：

(1) 系统程序模块化。执行系统只要有稍微的修改，就会影响到其它程序，但是修改和扩充又是不可避免的。为此，需要把整个系统模块化。所谓模块化，就是：第一，将系统中的某一或某些功能集中于某一程序段中，并且事先在不涉及其它模块的条件下进行调整和修改；第二，各模块之间通过固定的接口区进行模块间的通讯。模块化使得系统结构简单，并且易于修改和扩充。

(2) 作业控制语言的出现。执行系统的出现，促进了作业控制语言的发展。由于作业之间是自动过渡的，每个作业的运行过程也是自动的。因此，必须把控制作业运行的信息都反映在控制卡片上，然后提交给系统，由系统自动控制作业的运行。为了把控制卡与程序卡或数据卡区分开，控制卡的开头都带有特殊符号(通常为$或//)。

执行系统实现了通道和主机的并行操作，从而提高了计算机系统的处理能力。用户采用委托方式使用系统，从而保证了系统的安全。但执行系统也有两方面的缺点：第一，用户往往因不了解系统的使用规则而发生违例情况；第二，虽然主机和通道可并行操作，但不能完全消除处理机对于外部传输的等待。例如用户程序的输出量较大，主机往往要等待输出操作的结束。为了克服这些缺点，在此基础上又产生了多道程序和分时系统。

四、操作系统的形成

操作系统使用不久，人们就发现：若在机器主存内存放几道用户程序，每当一道程序等待外设传输而暂停时，可让另一道程序使用处理机，从而使CPU得到更充分的利用，这就出现了多道程序设计技术。不久，分时系统也相继出现。多道程序和分时系统的实现，标志了

操作系统的正式形成。

1．多道程序设计

所谓多道程序设计，是指同时把若干个作业存放在内存中，并且同时处于运行过程中。也就是说，这些作业都处于它们的开始点和结束点之间。但是，在某一给定时刻，真正在处理机上执行的只有一个作业(若只有一台处理机的话)，而其它作业，有的处于"就绪"状态，即它们具备了运行条件，但等待把处理机分配给它们；有的作业则可能因某种原因（如等待输入输出的完成）而处于"等待"(或称封锁或称阻塞)状态。这些作业的运行，完全由操作系统中的控制程序进行管理。

2．分时系统

分时系统，就是在一台计算机上，连接若干个终端，用户通过这些联机终端设备采用问答方式把他的程序和数据输入到计算机中，并控制程序的执行。系统把处理机时间轮流地分配给各个终端作业。每个作业只运行一段较短的时间。这样，各个用户的每次要求都能得到快速响应。

伴随多道程序和分时系统的出现，在软、硬件上又出现了如下进展：

(1) 存储管理。由于几个作业同时运行，必须有一种可靠的办法防止由于一个作业的错误而破坏了其它作业的正常运行。特别要防止侵犯系统程序。为此，硬件上提供了一种存储保护功能。其次，几个作业因共享内存，就要有一个存储分配的策略。第三，当实际内存容量不能满足需要时，就有一个内存扩充问题。这三点都属于存储管理的基本问题。

(2) 系统保护级。为了防止破坏系统程序，只有存储保护是不够的。于是引进了特权指令和系统保护级的概念。例如，规定用户程序在"算态"下运行，系统程序在"管态"下运行。特权指令(例如启动外设指令)只对"管态"开放，如在"算态"下执行特权指令，则发生故障中断。

(3) 作业控制语言。在多道和分时出现后，作业控制语言又有了进一步的发展，除了作业控制卡外，还出现了作业说明书和操作命令。

作业说明书由若干命令语句组成。这些语句反映了用户对作业进行控制的意图。系统解释这些语句，按语句的要求对作业实施控制。作业说明书在上机前写好，连同其程序、数据一起输入到计算机系统中去。

操作员命令是由操作员从终端键盘打入的命令。此外，对分时系统来说。还要提供一种会话语言。

(4) 文件系统。用户通过终端使用计算机，需要输入大量的程序和数据，为便于多次使用，用户希望能把这些信息保存起来，这就是产生文件系统的根本原因。

五、操作系统的进一步发展

到1968年为止，多道程序和分时系统都已完善，并且出现了实时操作系统、远程批量系统和计算机网。随着计算机网络和微计算机组成的多机系统的出现，近年来又发展了网络操作系统和分布式操作系统。

在1968年以前，操作系统的设计人员主要忙于为各类型号计算机配置相应的操作系统，主要精力都用于制定操作系统的设计方案，编制和调试操作系统，并且力求做到功能齐全。在操作系统的迅速而又有点混乱的发展中，有许多重要概念出现之后，过了一段时间又消失

了，后来又常常以不同形式再次出现。前面讲过的分时系统就是这种情况。

从1968年开始，操作系统的总结、提高和理论化的工作提到日程上来了。

从1968年开始，以后的主要工作有：

(1) 对过去的一切成果进行了统一和精选。消除了在操作系统发展过程中的混乱局面。这种倾向的一个例子是：把多道程序——批量处理的原理和分时技术结合起来，形成了操作系统，既能进行批量处理又能进行分时操作。

(2) 对操作系统的结构进行了研究和改革，并获得了几种较为成熟的结构设计方法。例如由Dijkstra在1968年提出的层次结构法代替了传统的模块结构法，使操作系统的可靠性得到了提高。1971年Brinch·Hansen发表了核扩散法，用核扩散法实现了操作系统，增加了操作系统的灵活性，并且也便于扩充。这些方法，工程实践证明了是可行的。

(3) 对进程通讯进行了研究。1968年初，Dijkstra提出了P、V操作作为进程间的通讯手段。1973年Brinch·Hansen提出了高级通讯语言——发送和接收原语。

(4) 关于对死锁问题和各种调度算法的研究，已取得了许多成果。对操作系统中的各种算法(例如处理机调度算法，存储分配算法等)进行了模拟、测量和分析。

(5) 对可靠性问题的研究做了大量的工作。例如对操作系统设计阶段的结构设计，实现阶段的程序正确性证明，以及对运行阶段的保护和容错等等的研究，都与可靠性问题密切相关。

(6) 对工具语言的研究。也就是研究编写操作系统程序的语言。最早使用汇编语言编制系统程序，后来陆续使用FORTRAN、ALGOL、PL/I，到了1970年初出现了顺序PASCAL、并行PASCAL语言。有了这些工具语言，对于缩短操作系统的研制周期，提高可靠性等方面都有很深远的意义。

操作系统理论研究的基本工具是数学。在性能测试和分析中使用了排队论、规划理论。在死锁问题的研究中使用了图论和集合论。在各种调度算法的研究中使用了排队论、图论和概率论。总之，离散数学、工程数学等许多基本数学理论都为操作系统的理论研究和进一步发展做出了巨大贡献。

综上所述，操作系统是在计算机广泛应用和计算机硬件不断发展中发展起来的。反过来，操作系统的发展又促进了计算机硬件的发展和计算机的广泛应用。

§1.3 操作系统的分类

操作系统有各式各样的分类方法，通常按其系统功能、运行环境以及服务对象来划分。尽管分类方法不同，迄今为止的各个操作系统均属于如下所列操作系统之一或它们的组合：批量处理系统；分时系统；实时系统；网络操作系统；分布式操作系统。下面分别加以简单说明。

一、批量处理系统

批量处理系统的突出特征是“批量”，它把提高系统的处理能力，即作业的吞吐量作为自己的主要目标，同时也兼顾作业的周转时间。在批量处理系统中，用户要使用计算机，事先必须准备好自己的作业，然后交给计算中心。计算中心的操作员并不立即进行输入，而是

等到一定时间或作业达到一定数量之后进行成批输入。计算结果也是成批进行输出。作业的执行采用“多道”形式，在作业执行过程中，用户不介入。实现批量处理的主要输入输出手段是 SPOOLing 系统。

在批量处理系统中，从作业的接收到作业的退出大体可分为四个阶段(如图 1.2 所示)。

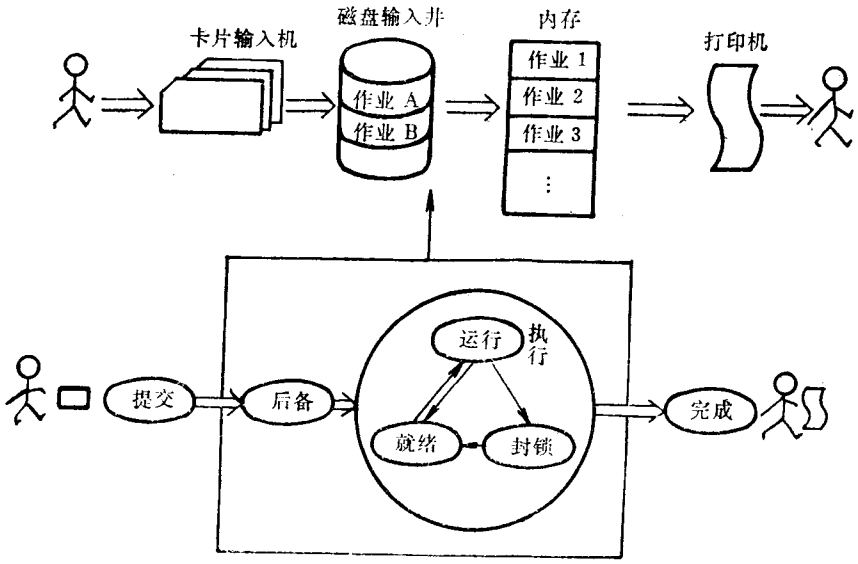


图 1.2　批量处理系统中作业状态的转换

(1) 用户首先准备好自己的作业，然后提交给计算机系统。

用户作业通常包括作业中的程序、数据以及说明作业如何运行的作业说明书。作业准备方式基本上有两种：一种是将上述三部分内容穿成卡片或纸带；另一种是利用文本编辑程序将它们构成文件存放在文件系统中。与此相应的作业提交形式也有两种：一种是将卡片或纸带提交给计算中心，由操作员将卡片或纸带信息输入到计算机系统中；另一种是用户使用终端或其它输入设备直接向系统提交作业。作业从用户手中送到计算中心所处的状态称为“提交”状态。

(2) 作业提交后，系统将它存放在磁盘上某一盘区中并等待运行。这些盘区通常称为输入井。其中可以同时存放许多等待运行的作业，这些作业称为后备作业。作业此时的状态称为后备状态或收容状态。

(3) 系统从后备作业中，挑选若干作业并将它们送入内存，再按一定方式使它们在处理机上运行。被选中的作业就处于“执行”状态。也就是说，这些作业处于执行过程的开始点和结束点之间。

在选择作业时要依据一定的算法和原则，要考虑到充分利用系统资源进行合理搭配，使处理机和各种输入输出设备的工作平衡。

由于有多道作业同时存放在内存中，所以这种批量处理系统称为多道批量处理系统。这是现代批量处理系统普遍使用的工作方式。

(4) 作业结束后，系统收回所需的各种资源并退出系统，此时作业所处的状态称为完成状态。

一个作业运行结束后，系统根据当时资源的使用情况以及各后备作业的特点，重新选择一个或一批作业送入内存，投入运行。

从上述多道批量处理系统来看，它实现了作业流程的自动化，提高了系统效率。但对用户来说，则感到越来越不方便了。这体现在以下三个方面：

（1）用户一旦把他的作业提交给系统后，他便失去了对作业运行的控制能力。虽然他可以通过作业说明书，把运行中可能出现的情况及解决办法告诉系统，但一般来说他很难预料到各种运行细节或出现的异常情况。一旦出现他所料想不到的情况，他也无法干预。

（2）在批量处理系统中，虽然也兼顾作业的周转时间，但这种周转时间通常是几小时或几天。然而，有时他的作业实际上只需几分钟内便可做完，这对用户来说显然是不合适的。

（3）对于小型作业，作业提交、结果输出以及再转交给用户的时间远远大于该作业的运行时间。因此一个小型作业的用户自然感到这种时间的浪费太可惜。

鉴于批量处理系统具有上述缺点，人们又提出了分时系统。它使用户通过终端使用计算机，直接控制作业的运行，这似乎又回到早期的手工操作阶段，但这是更高级的联机操作，它是对手工操作的否定之否定。

二、分时系统

1．分时和分时系统

在计算机硬件的发展中，自从出现了通道之后就有了分时的概念。在计算机系统中，按照通常习惯可以把分时定义为：两个或两个以上事件按时间划分轮流地使用计算机系统中的某一资源。在这样的定义下，CPU和通道分时地访问内存地址，以防止他们对内存访问的冲突。此外多台设备分时地使用通道，以及多道作业程序分时地使用处理机等等，这些都是分时。

在一个系统中，如果多个用户分时地使用同一计算机，那么这样的系统称为分时系统。在分时系统中，分时的时间单位叫做时间片。一个时间片通常是几十毫秒。在硬件上要利用中断机构和时钟。时钟使得 CPU 每运行一个时间片就产生一次时钟中断。中断后控制转向操作系统，操作系统轮流地处理各个用户作业，即把时间片分给各个终端用户。

一个分时计算机系统往往要连接几十甚至上百个终端设备，每个用户在他所占用的终端上控制其作业的运行。所以分时系统也称为多路存取系统(Multi-Access System)，图1.3给出了分时系统的概念图。

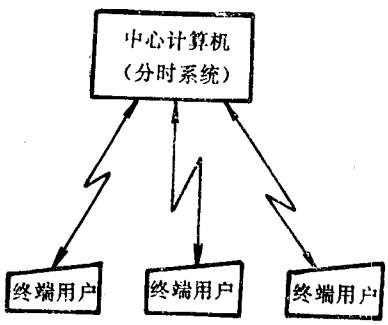


图 1.3　分时系统概念图

2．分时系统的特点

分时系统有如下四个特点：

（1）协调性。就整个系统来说，它要为各个终端用户服务，即要完成它们所请求的各种服务，又使各个用户都感到满意，这就要求分时系统能协调地工作。也就是说，若干用户可以同时操作，共同使用该系统。

（2）独立性。分时系统的一个突出特点是，每个终端用户都有一个共同感觉：我独占了整个系统的资源，该系统专为我一个用户服务。事实上，分时系统只在一个时间片内为一个用户服务，其余大部分时间为其它用户服务。由于时间片对人来说，感觉很短，