

5037

802524

08407:1

2

计算机软件基础

(下)

计算机应用软件人员水平考试
应试指导联合会教材编写组



087

08407:1

2

上海微电脑厂

前 言

对计算机应用人员实行统一的水平考试制度，在国外已有一些国家在实行。这种考试制度的实行将有利于在统一标准下显示计算机应用人员的实际能力和水平，保证各级各类培训的教育质量，而且有利于广大的科技人员学习和掌握计算机技术，促进计算机应用的普及和发展。

上海市政府1985年拟举行“计算机应用软件人员水平考试”。为了帮助有志者提高应试能力，为了使广大计算机应用人员进一步系统地掌握计算机知识，在上海计算技术研究所青年信息应用技术开发部、华东计算技术研究所青年电脑技术开发部、复旦大学计算机科学系研究生咨询中心、上海工业大学研究生学术交流协会、上海第二工业大学质量管理系(筹)、上海微电脑厂等单位的努力下，将组织举办计算机应用人员水平考试应试指导活动，拟开设有关计算机课程的学习班，由富有教学经验的教师讲课，并聘请著名计算机专家当顾问，进行水平考试的模拟考试等。为此我们根据上述需要有针对性地编写了一套教材和习题集，供大家复学、备考。由于时间仓促和编者水平有限，本书有欠妥之处，望读者谅解。

本书由计算机应用软件人员水平考试应试指导联合会教材编写组编写。

1984.11.24

第五章 操作系统

5.1 操作系统的概述

什么是操作系统呢？首先明确的是，操作系统是软件，它是裸机与用户之间的一个接口，这在以后的各节中可以清楚地看到这一点。

5.1.1 手工操作

在第一代计算机中，操作系统尚未问世，人们采用手工操作方式使用计算机，操作过程是：通过控制台按清除按钮清除前一用户在计算机内所留下的信息，然后通过外部设备（如磁带等）向计算机内输入信息，以此来建立作业；通过控制台启动开关来启动程序，在程序运行过程中需要操作员来干预计算机操作；当作业完成时，需要对外部设备进行作些处理（如卸下磁带等），接着打印计算结果。很显然，这种操作方式有如下的两个缺点。

(1) 一旦某用户开始操作，计算机的全部资源（处理机、内存、外部设备及一些简单的软件）都为该用户所独占，一直到他下机并把资源转给下一个用户为止；

(2) 操作是联机的，输入输出也是联机的，因此作业的运行时间拉得相当长。

这种操作方式在计算机速度较慢的情况下是允许的，但随着计算机的发展，计算机速度大大提高以后，就暴露出严重的弱点。例如，一个作业在速度为每秒1千次的机器上运行需要一小时，而作业的建立和整个操作过程只花三分钟，则操作时间与运行时间之比为1:20，若机器速度提高到每秒60万次，则该作业的运行时间缩短为六秒钟，但是操作速度不会有多大的改进，因为它主要由人的手动操作速度和思维速度决定，大致还是需要三分钟，这样，操作时间与运行时间之比为30:1，这就是说，操作时间远远超过了运行时间。因此，在五十年代末期，由于使用晶体管提高了计算机速度之后，缩短作业建立

时间与操作时间就成为急待解决的问题了。

5.1.2 成批处理

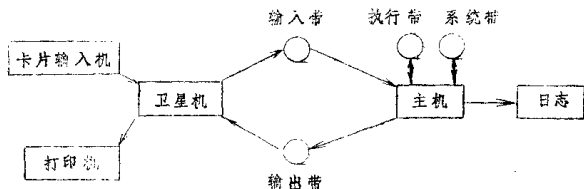
如何缩短作业的建立时间和操作时间呢？人们首先就想到，应该摆脱一个作业过渡到另一个作业中人的干预，使其为自动化，这样就出现了成批处理。

在早期的成批处理中，操作员把若干个作业合成一批，并把它们依次放在输入设备上，监督程序自动地把这批作业的第一个作业调入内存，并对该作业进行汇编或编译，然后由装配程序把编译后的结果程序及其所需的子程序（结果程序形式）一起装配成一个可执行的目标程序，然后启动并执行之，计算完成之后，由善后处理程序输出这个作业的计算结果。第一个作业完成后，监督程序又自动地把该批作业中的第二个作业调入内存，重复上面的过程，一直到这批作业全部处理完毕。综上所述，监督程序不间断地处理各作业，从而实现了作业间转换的自动化，从而缩短了建立时间和操作时间。

显然，在早期成批处理中，作业的输入输出是联机的，亦即作业从输入设备到内存，计算结果在输出设备上显示和打印，都是由中央处理机（CPU）直接控制完成的。当计算机速度提高后，由于CPU与输入输出设备（I/O）的速度差距明显的表现出来，这就使得高速的CPU在I/O时，要等待慢速度的联机I/O设备，从而不能发挥自己应有的效率。（应注意5.1.1中所述的是手工与计算机之间的时间差距，而此处的矛盾是慢速I/O设备与高速CPU之间的矛盾）。为了克服这一个缺点，在成批处理中引入了脱机输入输出技术。就是在主机之外另设一台小型卫星机，该机只与外设打交道，而不与主机有直接连结，如下图所示，卡片输入机上的作业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到内存，并予以执行；作业完成之后，主机只负责把结果记录到磁带上，而由卫星机负责把磁带上的信息在打印机上输出。这样，主机就摆脱了慢速的输入输出工作，而由卫星机去完成，而且卫星机与主机可以并行工作，因此这种系统比早期的成批处理系统大大提高了处理能力。这就是所谓的脱机成批处

理。

成批处理是在五十年代末期伴随第二代计算机而发展起来的。它的出现又促进了其它软件的发展。



脱机成批处理图

5.1.3 执行系统

在60年代初期，计算机硬件获得二个进展，一是通道的引入，二是通道中断主机的能力出现，这就导致了操作系统进入执行系统阶段。

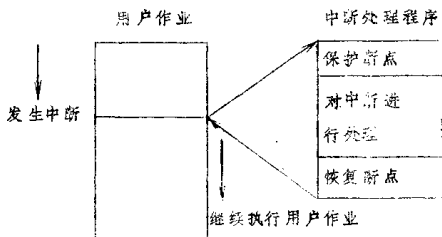
通道是一种硬件机构，它能控制一台或多台外部设备，它一旦被启动，就独立于CPU，这样就做到输入输出与计算过程重叠工作，主机（CPU）与通道同步的办法，是提供一些询问指令，由CPU发出这些指令，询问通道工作完成否，若未完成，主机就循环地询问，直至通道工作完成为止。

为了使通道传输与主机运行充分重叠，以减少主机询问等待时间，在程序设计中引进了缓冲技术。主机所需的输入信息由通道提前输入到缓冲区，同样，主机的输出信息先送到输出缓冲区，再由通道输出，从而减少主机的等待时间。

缓冲技术虽然减少循环询问等待时间，但是并不能彻底解决问题，于是硬件中引起了中断功能。所谓中断，就是在输入输出结束或发生某种故障时，相应的硬件向主机发信号，主机马上停止原来的工作去

处理中断要求。中断处理完成后，主机再回到原来的工作点继续工作。

中断的示意图如下所示：



上图中保护断点是指保护程序计数器、累加器、变址器的内容等，因为在执行中断处理程序过程中，会改变程序计数器，累加器和变址器等的内容，而在恢复执行被中断的程序之前，又要先恢复程序计数器的内容——中断点，及累加器、变址器的内容，这一切都是自动进行的。

为了获得主机计算与输入输出的充分重叠，就必须提供中断处理程序和输入输出控制程序（IOCS）。为了使可能多的为用户提供可用空间，把所有程序都受用到的中断处理程序和IOCS固定在内存，其余的程序则放在外存（辅存）上，永久留驻在内存的程序称执行程序（又称控制程序、监督程序、管理程序等）。

执行系统显然比成批处理进了一大步，它克服了成批处理的基本缺点，执行程序对其它程序拥有控制权，用户程序的输入输出是通过执行程序委托给IOCS实现的，受托要求完成后，通过中断信号通知执行程序，因而提高了系统的安全，另外由于硬件提供的时钟手段，故对程序死循环的情况可通过时钟中断信号而抑制。

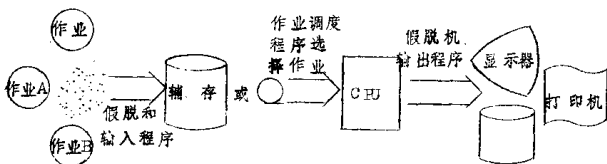
5·1·4 操作系统的形成

六十年代初期，执行系统使用不久人们就发现，在内存中同时存放几道用户程序是十分有利的。当某一道程序因等待外部传输而工作不下去时，系统马上可令主机执行另一道程序，从而使主机得到充分的利用，在这种思想的指导下，就出现了多道成批系统。多道成批系统的最大缺点就是不允许用户和机器进行交互性对话。为了克服这一缺点，不久就出现了分时系统。多道成批系统和分时系统的出现标志着操作系统最后形成。此后，由于实际需要，又出现了实时系统，这样就完成了通用操作系统。通用操作系统的出现，系统处理能力的提高，标志着操作系统已进入日至完善的阶段。

·多道成批系统

多道成批的基本思想，是主存中同时放入几道作业，一般说来，这些作业应是较为合理搭配的，例如可以把运算量大（输入输出量小）的科技计算作业和输入输出量大的数据处理作业搭配，这样的话，主机在处理科技计算作业的同时，输入输出设备正在处理数据处理作业，这样才能保证主机和外部设备高度并行地工作。

多道成批系统的工作流程如下所述：操作员把一批批作业随时提交给操作系统，假脱机输入程序及时把这些作业预输入到磁存中等待运行，以后当系统需要调入新的作业时，作业调度程序按一定的调度原则，选择相搭配的作业调入内存运行；内存中的各作业交替执行，当某个作业因某种原因（如等待输入输出完成）工作不下去时，系统就转去执行另一个作业，如此重复而已，直至某个作业工作完毕为止；当某作业完成时，系统把计算结果交给假脱机输出程序进行输出，并收回它所使用的全部资源。



·分时系统

手工操作是一种联机操作方式，其效率很低。成批处理否定并代替了手工操作，成批处理是一种脱机操作方式，多道成批系统发展了成批处理和执行系统的特点，进入了更高级的脱机操作阶段。

但在后来发现，仍有联机的必要，如几个系统分析员希望同时调试一个系统程序，且在调试时要直接干预程序。此外，用户也提出需要根据程序的运行情况动态地对其程序施以控制的要求，根据上述情况，分时系统就应运而生了。

分时系统有三个特点：多路调制性，交互性和独立性，基本思想是：一台计算机上可挂上几个控制台和上百台终端，每一用户都可以在一台终端或控制台上控制作业运行，这就使很多联机用户可以同时使用一台计算机（多路调制性），并与它进行对话（交互性），而用户又觉得自己在独占计算机（独立性）。

分时系统的工作流程是：每个用户可以通过终端向系统提交命令，请求完成某项工作；然后，系统分析从终端设备发来的消息，完成用户提出的要求，并把运行结果（成功或失败）再通过终端设备告诉相应的用户；以后，用户又根据系统提供的运行结果，向系统提出下一步请求。如此重复上述交互会话过程，直到用户完成预计的工作为止。由于终端上的用户工作很慢，所以系统能够在较短的时间（几秒钟）内同时响应所有用户的要求。实际上，系统是把处理机时间划分成很短的时间片，轮流地分配给各个联机作业使用。

分时系统的优点是：第一，使用户能在较短的时间内采用交互会话工作方式，及时地编写、调试、修改和运行自己的程序，因而加快了解题周期；第二，无论是本地用户或远地用户，只要与计算机连上一台终端设备，就可以随时随地使用计算机；第三，分时用户之间可以通过计算机的文件系统，彼此交流程序、信息和计算结果，便于多个用户协作完成共同关心的任务。

·实时系统

实时系统按其使用方式分成：① 实时控制 —— 如炼钢控制等，② 实时信息处理 —— 如飞机订座、情报检索等。

实时系统对响应时间的要求比分时系统更高，一旦向实时系统提出服务请求后，要求系统立即处理。实时系统没有分时系统那样强的交互式会话能力，但是它对系统可靠性和安全性要求很高。

多道成批系统、分时系统和实时系统的出现，对硬件和软件提出了一些要求，并促进了硬、软件的发展。主要有如下几个方面。

(1) 存贮管理。首先，当几个作业同时运行时，必须要有一个可靠的办法，用以防止由于一个作业的错误而影响其它作业的正常运行，特别要防止破坏系统程序，因此硬件一般都提供了存贮保护；其次，几个作业共享内存，就有一个存贮分配问题；再者，当各个作业对内存的要求总量或一个作业的内存要求量超过实际内存时，还有一个内存扩充问题。这三者都属于存贮管理所要研究的课题。

(2) 系统保护级。光靠存贮保护，还不能完全防止用户破坏系统程序，因此引进了特权指令和保护级的概念，所谓保护级，是指机器硬件可在多种不同的状态下工作，对于不同的状态，允许使用硬件的权力是不同的。例如，用户程序在算态下运行，系统程序在管态下运行，特权指令（如启动外部设备指令）只对管态开放。若在算态下执行特权指令，则发生故障中断。

(3) 文件系统。当用户通过终端使用计算机时，需要打入大量程序和数。若每次联机都要打入这么多信息，显然会产生重复性的浪费，因此要求在联机期间把打入的信息存贮起来，以备下次再用。在多道成批系统中，每个作业步都要接收大量的程序或数据作为输入，并产生计算结果和运行信息作为输出。这个输入输出都是以文件形式驻留在各种存贮介质上。因此，现代操作系统都配置文件系统，这是一种长期存贮手段，使得用户信息能长期存贮在系统中。

5.2 研究操作系统的几种观点

综上所述，可知操作系统具有以下特点：(1) 协调、控制并发执行的程序段；(2) 动态地管理各种软硬设备；(3) 其结构复杂、模块繁多，相互联系复杂，为此人们提出各种观点，试图给出一种系

统的方法，以便研究、分析和设计操作系统的功能、组成部分、工作部分以及体系结构。概言之，有如下三种观点：

· 资源管理 —— 研究和考察硬软设备的动态管理，目的是将一个操作系统划分为许多独立的模块。

· 分层结构机 —— 研究大型程序的编制方法，目的是如何安置操作系统的各模块，确定系统的体系结构。

· 进程 —— 研究并发程序的状态变化及它们之间的相互制约关系，目的是搞清操作系统各模块之间的动态连接关系。

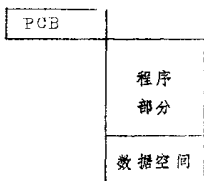
5.2.1 进程观点

操作系统的重要特点之一是并发性。所谓并发性，是指操作系统控制很多能并发执行的程序段，有些是系统程序段，有些是用户程序段。这些并发程序段彼此独立并发地执行自己的工作，有时彼此间却要以直接或间接方式发生着相互依赖、相互制约的关系。直接方式的制约关系通常是在彼此有逻辑关系的程序段之间发生的。如一程序段等待另一程序段的计算结果或信息。只当另一程序段送来相应的计算结果和信息时，这个程序段才能继续工作下去。间接方式的制约关系通常是由于并发程序段竞争相同资源而产生的，得到资源的程序段可以执行，反之就得暂时挂起，等到有可用后资源时再执行。

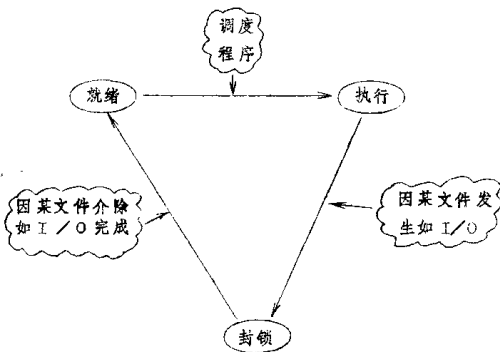
正因为这些并发执行的程序段之间存在着相互制约关系，故每个程序段不能与外界隔绝，不能随心所欲地在处理机上运行，它不仅受其它程序段的活动所制约，且要动态地依赖于系统资源的分配情况。因此使用程序这个概念，只能对它们进行简单、孤立、静止地研究分析，而不能深刻地揭示它们的活动联系及状态变化，故在操作系统中，引入“进程”概念。

进程有时也称作任务或活动。所谓进程，是并发程序段的执行。通常把进程表示成三部分：程序、数据空间、进程控制块(PCB)。程序部分表示了进程本身所要完成的功能；数据空间是程序进行加工的场所；进程控制块则包含了进程的有关信息，表示了进程在当前时刻的状态以及与其他进程和资源的关系。PCB所包含的信息有：进

程名字、当前状态、所需资源、已分配资源、调度信息、通讯信息、与其他进程的联系等等。每个进程可以下图为示：



引入进程之后，就可以研究它们的状态变化和相互制约关系了。下图是进程状态变化的一个简单模型。



进程可处于三种状态之一：

执行状态 —— 进程已分配到处理机，它的程序正在执行。

封锁状态 —— 进程等待某事件（如等 I/O 完成），此时，即

使分配给它处理机，它也不能执行。

就绪状态 —— 进程已具备执行条件，但因处理机数少于进程数，故暂时不执行，须等待分配处理机。

进程的各个状态可依据一定的原因和条件发生变化，如上图所示。

5.2.2 资源管理观点

操作系统要管理许多硬软设备，统称为资源，按其性质划分成四大类：处理机、存储器、外部设备、信息（程序和数据），这四部分资源构成了操作系统本身和用户作业赖以活动的环境，基于此，可把操作系统看成是一个资源管理程序所组成的。因此可把操作系统划分成处理机管理、存贮管理、设备管理和信息管理（即文件系统）加以讨论。

由于操作系统所管理的资源品种繁多，故需用统一的方式加以管理。因此研究资源管理的目的在于：为用户提供一种简单、有效使用资源的方法，充分发挥资源的利用率，防止分配中发生死锁现象，故从下面几个方面内容来研究每种资源管理：

(1) 记住资源的使用状态，记住资源是否被使用，以及被谁使用等情况。

(2) 确定资源的分配原则和调度原则。根据系统的设计目标，确定一组原则，用以决定“资源分配给谁？”、“何时分配？”、“分配多少？”以及“分配什么”等问题。

(3) 由用户的要求及(2)中的原则，执行资源分配。

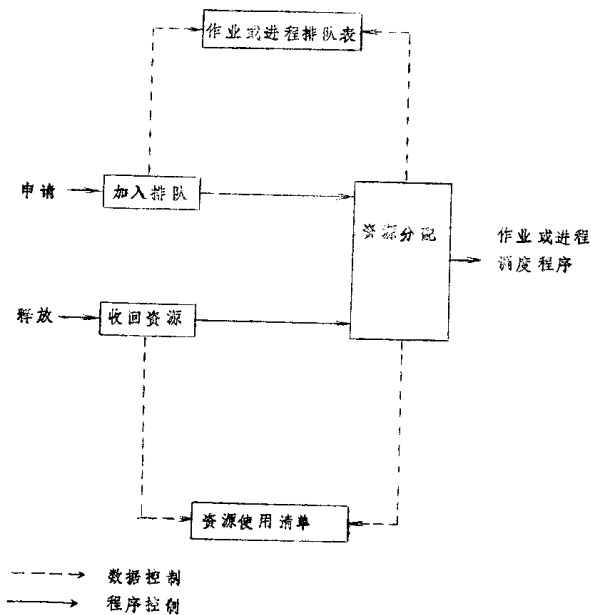
(4) 收回资源。当作业或进程不再需要某种资源时，系统应及时收回资源，以便重分配。

当代操作系统的一个重要特点是内存中可以有几道作业处于可运行状态，而每一作业可由多个彼此独立能并发执行的进程组成，这些作业和进程就构成了系统中分配资源的对象。一般来说，作业和进程是按照“申请 —— 使用 —— 释放”的步骤和资源发生联系的。

下图是一种较简单的资源管理模型，是由如下的数据结构和程序模块

组成的。

(1) 数据结构。该模型使用一张资源使用清单和一张作业或进程的排队表，资源使用情况记录了所有资源的使用情况，作业或进程的排队表记录了所有等待分配此种资源的作业或进程。



一种较简单的资源管理模型

(2) 程序模块。该模型是由“加入排队”、“收回资源”和“资源分配”三个模块组成的。

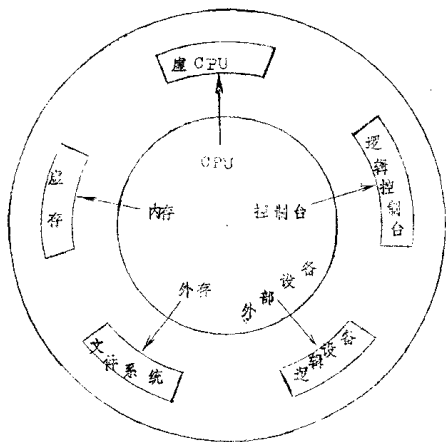
当正在执行的作业或进程执行“申请”资源命令时，控制转到相应资源的“加入排队”模块，该模块把当前作业或进程的名字和命令参数合在一起造登记项，之后把它插到排队表中等待分配资源。然后控制转给资源分配模块，由它查看排队表和资源使用清单，并根据资源分配原则确定能否满足相应作业或过程的资源请求。如能满足，则把相应作业或进程的状态置为“就绪”状态，否则置为“封锁”状态。

得到资源的作业或进程，其状态为“就绪”状态，当调度程序再次调度到它时，它就使用所分配到的资源继续工作，直至不再需要这资源并执行“释放”资源命令时为止。此时控制转给“收回资源”模块，它把所释放的资源标记为“未用”资源，然后再试图解除等待此资源的作业或进程。

5.2.3 分层虚拟机器系

操作系统是用户与计算机之间的介面，是协助用户解决问题有效使用计算机的有力工具。因此对用户来说，配备了操作系统的计算机是一台比原来计算机功能更强的“计算机”。因为这种计算机是概念上的计算机，而非真实的物理计算机，所以我们称它为虚拟计算机。

物理计算机通常由中央处理机(CPU)、内存、外存、外部设备和控制台组成。人们可通过操作系统分别对这五个部分加以扩充，使之成为功能完善，利于使用的计算机虚拟机。中央处理机经扩充后得到速度慢些但数量较多的虚CPU，内存经扩充后得到数量较多，容量较大的虚存，外存经扩充后得到一个按名字存取的大型存储的文件空间，外设经扩充后得到方便灵活，数量较多的逻辑设备，控制台经扩充后得到每一用户都有一个逻辑控制台。下图是经过扩充后的虚拟机。



虚 拟 机

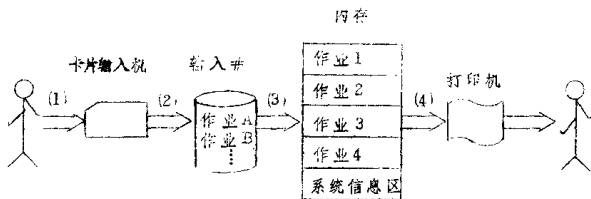
一般来说，扩充虚拟机的过程通常是分层进行的，即从最低层的物理机开始，然后相继地在每层虚拟机上编制程序去构造更高层次的虚拟机，重复这一过程，直至构造出所需要的系统为止。这种扩充虚拟机的方法称为分层虚拟机方法。当用这种方法用于上图所示的虚拟机时，可在每层中只构造扩充一种设备，也可以扩充几种或全部的设备，但每层只改造扩充设备的部分功能。在严格的分层结构中，各层之间要有简单准确的通讯方法，而且只允许高层调用低层，不允许低层调用高层。确定好各层虚拟机的功能后，就可以逐层地设计、编写和调试它们。由于各层的功能明确、结构清晰，所以这些工作相对来说都变得简单一些，甚至有可能严格地论证各层的功能及结构关系，从而保证整个系统的正确性。

5.3 处理机管理

上面已讲过，随着通道的引入，为使CPU和I/O通道并行工作，引了多道程序设计，即在一个时刻有几个作业同时在计算机内运行。在此情况下，各个作业都需要有CPU供它运行，但实际上，在计算机系统中，处理机的数目总是少于同时运行的用户作业数，且往往只有一台CPU，这样就产生了CPU的分配问题。最简单的办法是把CPU的时间轮流分配给各个作业，但在实际应用上往往根据不同的设计目标采用不同的CPU分配。如在分时系统中，就要求系统响应时间较快，使用户觉得自己独占计算机；又如实时系统中，系统响应时间就更短了，这就说明要更合理的调配CPU，这些CPU的管理，就属于处理机管理的范畴。

那么，处理机管理是如何管理CPU呢？先让我们看一下多道程序设计情况下作业的计算流程，然后讨论与此流程的每个阶段相对应必须有哪些处理机管理机构，以及这些机构的基本工作原理。

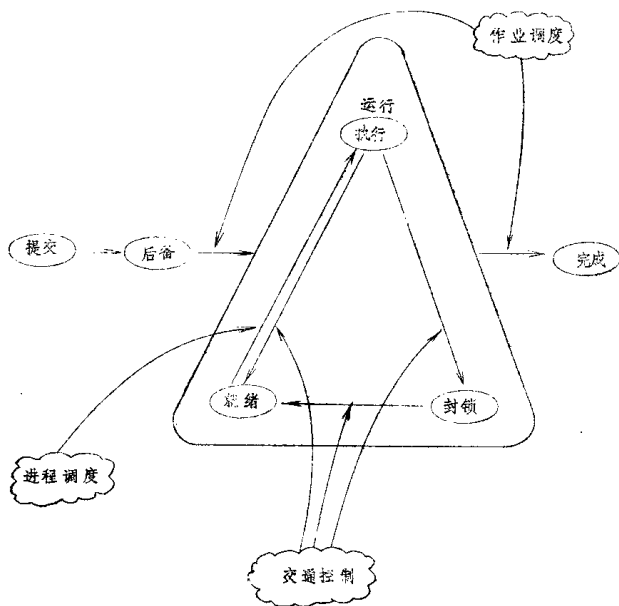
下面给出了成批处理作用的流程。



从上图中可以看出，一个成批处理作业从进入系统到退出系统分成四个阶段：（1）程序员将源程序信息交给机房；（2）由机房操作人员通过输入机把作业信息送入机内，并由系统将其安置到磁盘的某些盘区（称为输入#）中等待运行；（3）系统从这些等待运行的作

业中选择若干个作业，分配给它们所需的资源，投入运行；（4）作业运行结束后，系统收回该作业所占用的资源，并令其退出系统。

为了刻画作业在上图中所处的不同阶段，相应地引入四种状态，如下图所示：



多道程序中作业的信息流和状态转换