

第一章 操作系统引论

- 1.1 操作系统的目标和作用
- 1.2 操作系统的发展过程
- 1.3 操作系统的基本特性
- 1.4 操作系统的主要功能
- 1.5 OS结构设计

1.1 操作系统的目标和作用

操作系统的目标与应用环境有关。

➤ 查询系统：人一机交互性

在查询系统中所用的OS，希望能提供良好的人—机交互性



➤ 控制系统：要求具有实时性

对于应用于工业控制、武器控制以及多媒体环境下的OS，要求其具有实时性.



➤ 微机上配置的OS：注重其使用的方便性

比如控制汽车上的车载电子系统，就要求其简单易操作。



1.1.1 操作系统的目标

(1) 有效性：使得计算机易学易用

在早期(20 世纪 50~60 年代)，由于计算机系统非常昂贵，操作系统最重要的目标无疑是有效性。

操作系统的有效性可包含如下两方面的含意：

- 提高系统资源利用率。在未配置 OS 的计算机系统中，诸如 CPU、I/O 设备等各种资源，都会因它们经常处于空闲状态而得不到充分利用
- 提高系统的吞吐量。操作系统还可以通过合理地组织计算机的工作流程，而进一步改善资源的利用率，加速程序的运行，缩短程序的运行周期，从而提高系统的吞吐量

(2) 方便性：使得计算机易学易用

配置 OS 后可使计算机系统更容易使用。

- 一个未配置 OS 的计算机系统是极难使用的。用户要直接在计算机硬件上运行自己所编写的程序，就必须用机器语言书写程序；用户要想输入数据或打印数据，也都必须自己用机器语言书写相应的输入程序或打印程序。
- 如果我们在计算机硬件上配置了 OS，用户便可通过 OS 所提供的各种命令来使用计算机系统。比如，用编译命令可方便地把用户用高级语言书写的程序翻译成机器代码，大大地方便了用户，从而使计算机变得易学易用。

方便性和有效性是设计操作系统时最重要的两个目标。

(3) 可扩充性:

- 随着计算机技术的迅速发展, 计算机硬件和体系结构也随之得到迅速发展, 相应地, 它们也对 OS 提出了更高的功能和性能要求。
- 此外, 多处理机系统、计算机网络, 特别 Internet 的发展, 又对 OS 提出了一系列更新的要求。
- 因此, OS 必须具有很好的可扩充性, 方能适应计算机硬件、体系结构以及应用发展的要求。这就是说, 现代 OS 应采用新的 OS 结构, 如微内核结构和客户服务器模式, 以便于方便地增加新的功能和模块, 并能修改老的功能和模块

(4) 开放性:

- 自 20 世纪 80 年代以来, 由于计算机网络的迅速发展, 特别是 Internet 的应用的日益普及, 使计算机操作系统的应用环境已由单机封闭环境转向开放的网络环境。
- 为使来自不同厂家的计算机和设备能通过网络加以集成化, 并能正确、有效地协同工作, 实现应用的可移植性和互操作性, 要求操作系统必须提供统一的开放环境, 进而要求 OS 具有开放性。
- 开放性是指系统能遵循世界标准规范, 特别是遵循开放系统互连(OSI)国际标准。凡遵循国际标准所开发的硬件和软件, 均能彼此兼容, 可方便地实现互连。
- 开放性已成为 20 世纪 90 年代以后计算机技术的一个核心问题, 也是一个新推出的系统或软件能否被广泛应用的至关重要的因素。

1.1.2 操作系统的作用

➤ OS作为用户与计算机硬件系统之间的接口

含义：OS 处于用户与计算机硬件系统之间，用户通过 OS 来使用计算机系统。

➤ OS作为计算机系统资源的管理

在一个计算机系统中，通常都含有各种各样的资源。可将资源分为四类：**处理器、存储器、I/O 设备以及信息**。相应地，OS 的主要功能也正是针对这四类资源进行有效的管理

➤ OS实现了对计算机资源的抽象

OS隐藏了对硬件操作的细节，实现了对计算机硬件操作的多个层次的抽象

1. OS作为用户与计算机硬件系统之间的接口

操作系统处于用户与计算机硬件系统之间，用户通过OS使用计算机系统

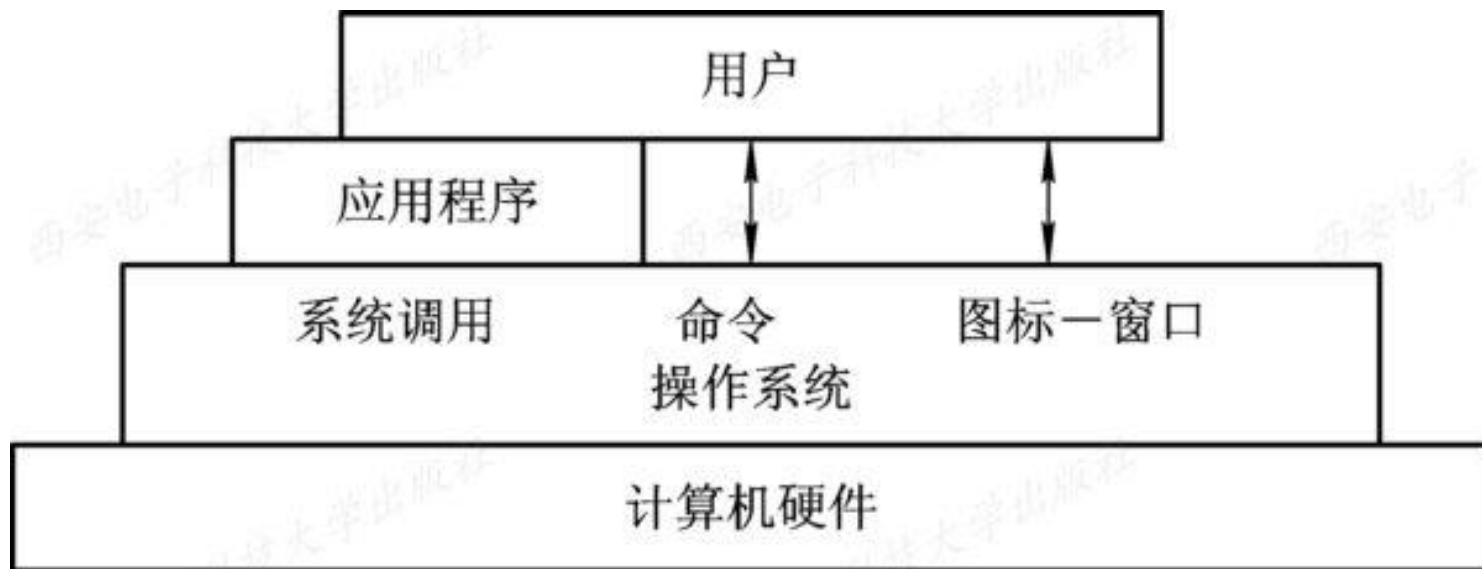


图1-1 OS作为接口的示意图

用户可通过以下三种方式使用计算机:

- **命令方式。**这是指由 OS 提供了一组联机命令接口，以允许用户通过键盘输入有关命令来取得操作系统的服务，并控制用户程序的运行。
- **系统调用方式。**OS 提供了一组系统调用，用户可在自己的应用程序中通过相应的系统调用，来实现与操作系统的通信，并取得它的服务。
- **图形、窗口方式。**这是当前使用最为方便、最为广泛的接口，它允许用户通过屏幕上的窗口和图标来实现与操作系统的通信，并取得它的服务。

2. OS作为计算机系统四类资源的管理者

处理机管理：分配和控制处理机

存储器管理：对内存进行分配与回收

I/O设备管理：负责I/O设备的分配回收

文件管理：文件的共享、存取与保护

3. OS实现了对计算机资源的抽象

裸机：完全无软件的计算机系统

扩充机器：覆盖了软件的机器，也称为虚拟机

- 对于一台完全无软件的计算机系统(即裸机)，由于它向用户提供的仅是硬件接口(物理接口)，因此，用户必须对物理接口的实现细节有充分的了解，这就致使该物理机器难于广泛使用。
- 为了方便用户使用I/O设备，人们在裸机上覆盖上一层I/O设备管理软件
- 如图1-2所示，由它来实现对I/O设备操作的细节，并向上将I/O设备抽象为一组数据结构以及一组I/O操作命令，如**read**和**write**命令，这样用户即可利用这些数据结构及操作命令来进行数据输入或输出，而无需关心I/O是如何具体实现的。

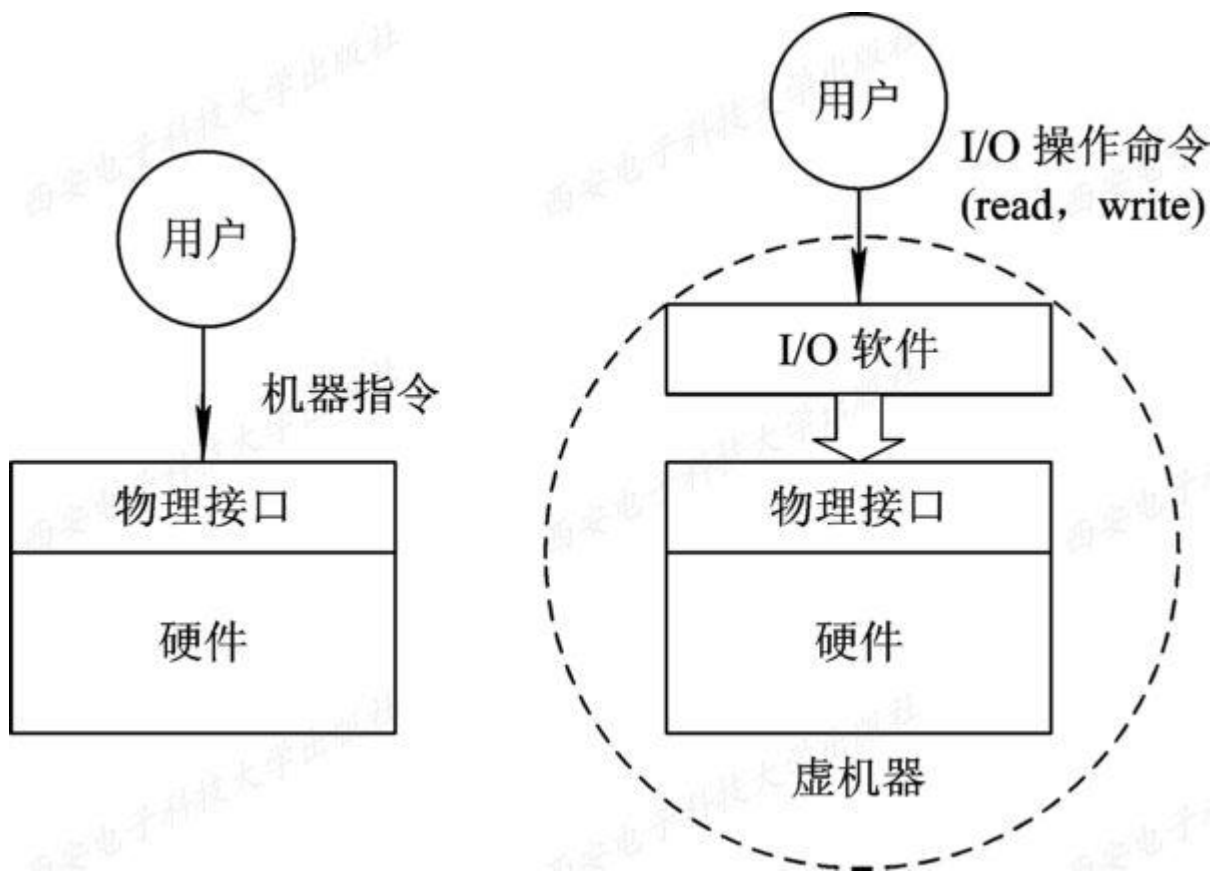


图1-2 I/O软件隐藏了I/O操作实现的细节

1.1.3 推动操作系统发展的主要动力

➤ 不断提高计算机资源利用率：

在计算机发展的初期，计算机系统特别昂贵，人们必须千方百计地提高计算机系统中各种资源的利用率，这就是 OS 最初发展的推动力。

➤ 方便用户：

当资源利用率不高的问题得到基本解决后，用户在上机、调试程序时的不方便性便又成为主要矛盾。于是人们又想方设法改善用户上机、调试程序时的环境，这又成为继续推动 OS 发展的主要因素。

➤ 器件的不断更新换代：

微电子技术的迅猛发展，推动着计算机器件的更新换代。如芯片从8位→16位→32位→64位。存储设备从磁盘→移动硬盘→闪存盘→扫描仪等。

➤ 计算机体系结构的不断发展：

- ❑ 例如，当计算机由单处理机系统发展为多处理机系统时，相应地，操作系统也就由单处理机 OS 发展为多处理机 OS。
- ❑ 又如，当出现了计算机网络后，配置在计算机网络上的网络操作系统也就应运而生，它不仅有效地管理好网络中的共享资源，而且还向用户提供了许多网络服务。

1.2 操作系统的发展过程

- OS 的形成迄今已有 50 多年的时间。在上世纪 50 年代中期出现了单道批处理操作系统
- 60 年代中期产生了多道程序批处理系统；不久又出现了基于多道程序的分时系统，与此同时也诞生了用于工业控制和武器控制的实时操作系统。
- 20 世纪 80 年代开始至 21 世纪初，是微型机、多处理机和计算机网络高速发展，同时也是微机 OS、多处理机 OS 和网络 OS 以及分布式 OS 的形成和大发展的年代。

1.2.1 未配置操作系统的计算机系统

从第一台计算机诞生(1945 年)到 20 世纪 50 年代中期的计算机,属于第一代计算机。此时的计算机是利用成千上万个真空管做成的,它的运行速度仅为每秒数千次,但体积却十分庞大,且功耗也非常高。这时还未出现 OS。此时操作方式有以下几种:

1. 人工操作方式

从第一台计算机诞生(1945 年)到 20 世纪 50 年代中期的计算机，属于第一代计算机。此时的计算机是利用成千上万个真空管做成的，但体积十分庞大，且功耗也非常高。这时还未出现 OS。计算机操作是由用户采用人工操作方式直接使用计算机硬件系统。

方式：纸带打孔放入输入机，经计算机处理后打印出结果

特点：用户独占全机，CPU等待人工操作能够（串行性），联机输入输出方式（输入输出都由主机控制）

问题：CPU等待人工操作，CPU和I/O设备间不匹配

突出的问题：

- 人工操作方式严重降低了计算机资源的利用率，此即所谓的人机矛盾。
- 随着 CPU速度的提高和系统规模的扩大，人机矛盾变得日趋严重。
- 此外，随着 CPU 速度的迅速提高而 I/O 设备的速度却提高缓慢，这又使 CPU 与 I/O 设备之间速度不匹配的矛盾更加突出。

2. 脱机输入/输出 (Off-Line I/O) 方式

目的： 解决人机矛盾及CPU和I/O设备之间速度不匹配的矛盾

方式： 事先将纸带装入输入机，在外围机的控制下，把纸带上的数据输入到磁带上。当CPU需要这些程序和数据时，再从磁带上高速地调入内存。

- 由于程序和数据的输入和输出都是在外围机的控制下完成的，或者说，它们是在**脱离主机**的情况下进行的，故称为脱机输入/输出方式
- 反之，在**主机的直接控制**下进行输入/输出的方式称为联机输入/输出(On-Line I/O)方式

关键设备：外围机、磁带

优点：

- (1) 减少了 CPU 的空闲时间。
装带(卡)、卸带(卡)以及将数据从低速 I/O 设备送到高速磁带(或盘)上，都是在脱机情况下进行的，并不占用主机时间，从而有效地减少了 CPU 的空闲时间，缓和了人机矛盾。
- (2) 提高了 I/O 速度。当 CPU 在运行中需要数据时，是直接从高速的磁带或磁盘上将数据调入内存的，不再是从低速 I/O 设备上输入，极大地提高了 I/O 速度，从而缓和了 CPU 和 I/O 设备速度不匹配的矛盾，进一步减少了 CPU 的空闲时间。

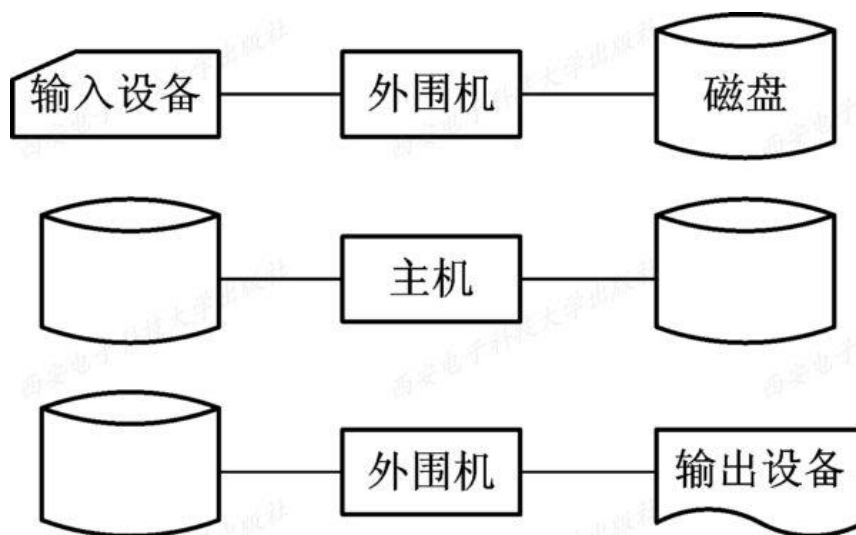


图1-3 脱机I/O示意图

1.2.2 单道批处理系统

- 上世纪 50 年代用晶体管替代真空管来制作计算机，，但计算机仍非常昂贵。为了能充分地利用它，应尽量让该系统连续运行，以减少空闲时间。
- 为此，通常是把一批作业以脱机方式输入到磁带上，并在系统中配上监督程序(Monitor)，在它的控制下使这批作业能一个接一个地连续处理

(1) **定义：** 系统对作业的处理是成批进行的，且在内存中始终只有一道作业

(2) **方式：** 把一批作业以脱机方式输入到磁带上，在监督程序(Monitor)的控制下，使这批作业能一个接一个地连续处理。

(3) **特征：**

- **自动性。** 在顺利情况下，在磁带上的一批作业能自动地逐个地依次运行，而无需人工干预。
- **顺序性。** 磁带上的各道作业是顺序地进入内存，各道作业的完成顺序与它们进入内存的顺序，在正常情况下应完全相同，亦即先调入内存的作业先完成。
- **单道性。** 内存中仅有一道程序运行，即监督程序每次从磁带上只调入一道程序进入内存运行，当该程序完成或发生异常情况时，才换入其后继程序进入内存运行。

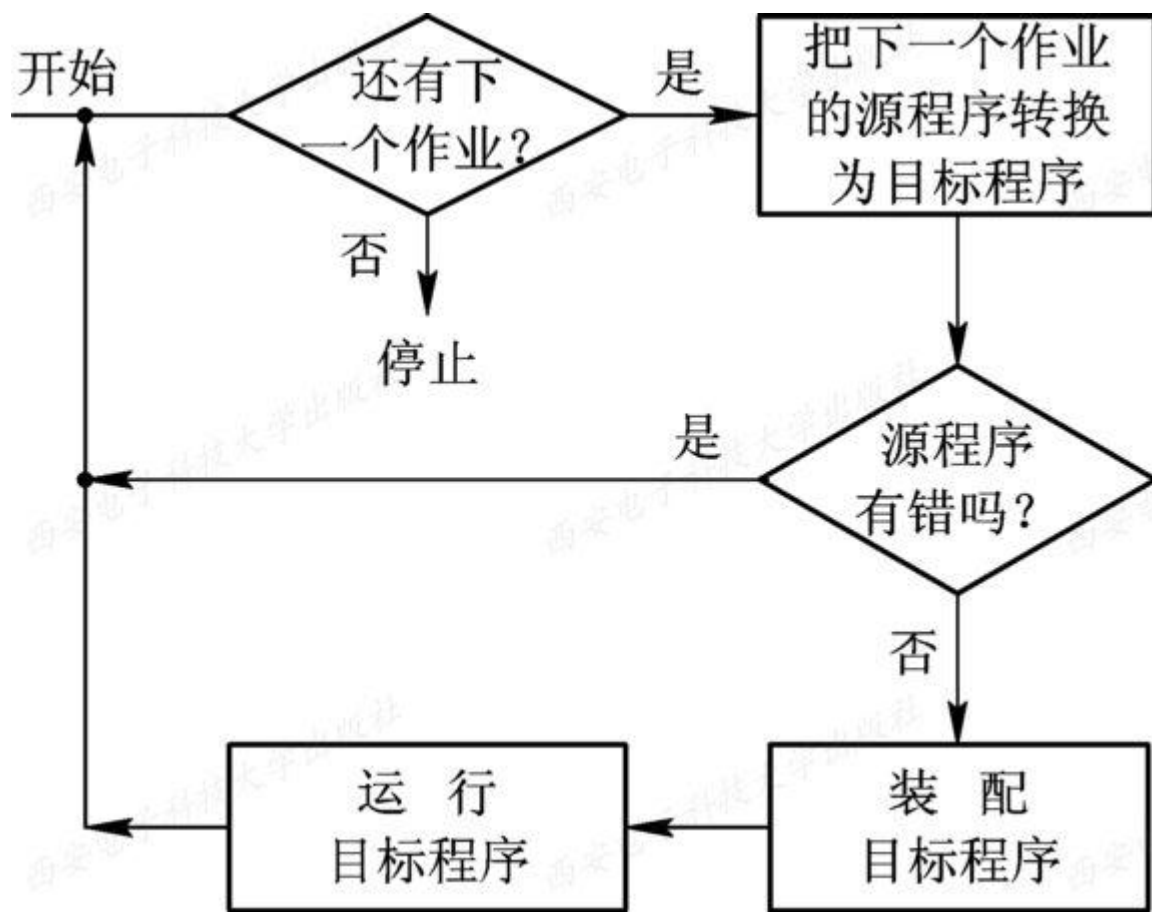


图1-4 单道批处理系统的处理流程

缺点：系统中的资源得不到充分的利用，当程序发出IO请求后，必须等待IO操作完成

从图可以看出：在 $t_2 \sim t_3$ 、 $t_6 \sim t_7$ 时间间隔内CPU空闲。

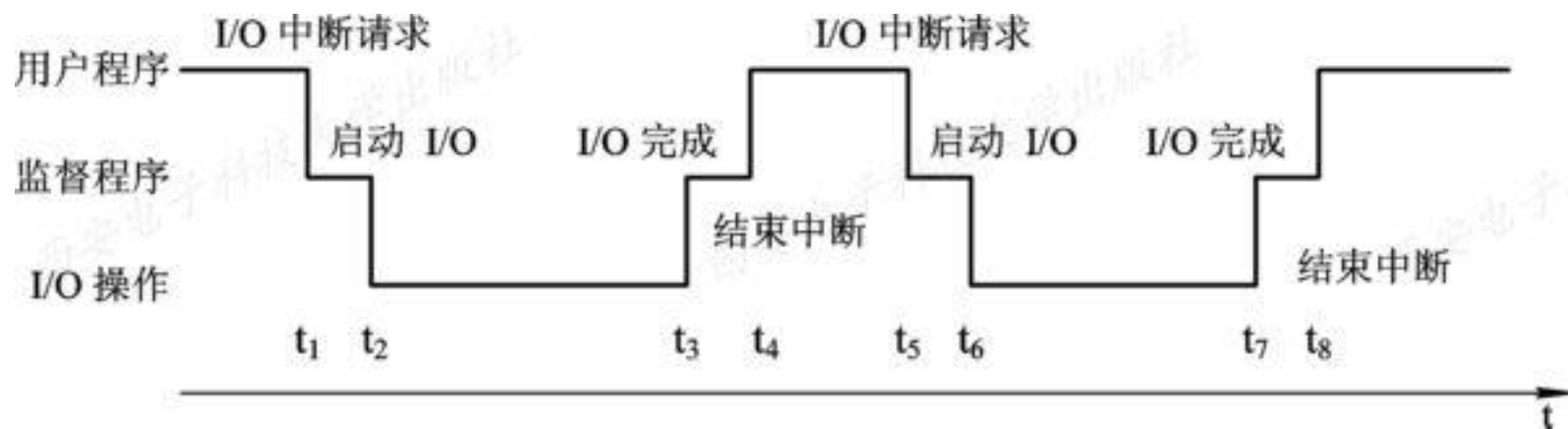


图1-5 单道程序的运行情况

1.2.3 多道批处理系统(Multiprogrammed Batch Processing System)

20 世纪 60 年代中期，人们开始利用小规模集成电路来制作计算机，生产出第三代计算机。

由 IBM 公司生产的第一台小规模集成电路计算机——360 机，较之于晶体管计算机，无论在体积、功耗、速度和可靠性上，都有了显著的改善。

虽然在开发 360 机器使用的操作系统时，为能在机器上运行多道程序而遇到了极大的困难，但最终还是成功地开发出能在一台机器中运行多道程序的操作系统 OS/360。

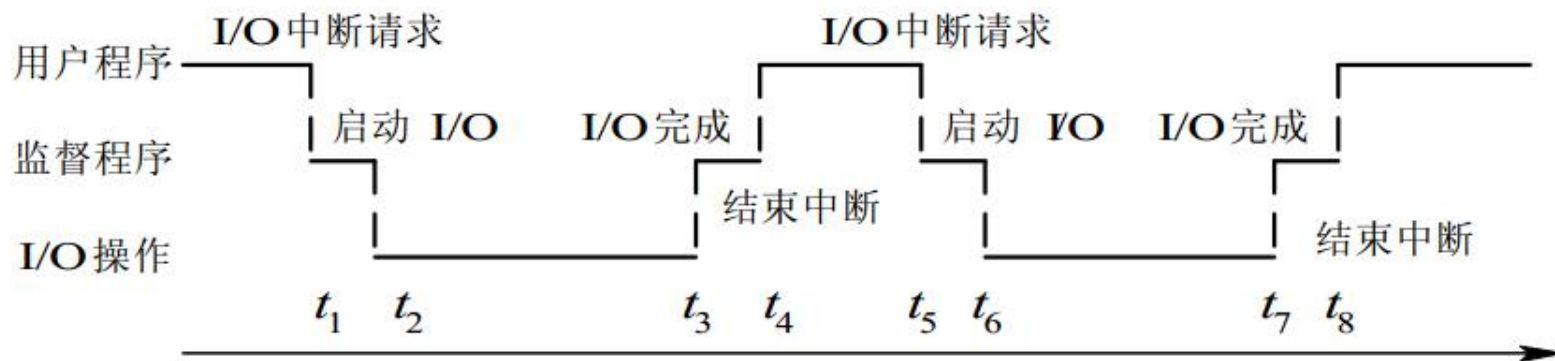
1.2.3 多道批处理系统 (Multiprogrammed Batch Processing System)

1. 多道程序设计的基本概念

目的： 进一步提高资源的利用率和系统吞吐量

- 在该系统中，用户所提交的作业都先存放在外存上并排成一个队列，称为“**后备队列**”；然后，由作业调度程序按一定的算法从后备队列中选择若干个作业调入内存，使它们共享 **CPU** 和系统中的各种资源

下面两图示出了四道程序时单道程序和多道程序的运行情况。



单道程序的运行情况

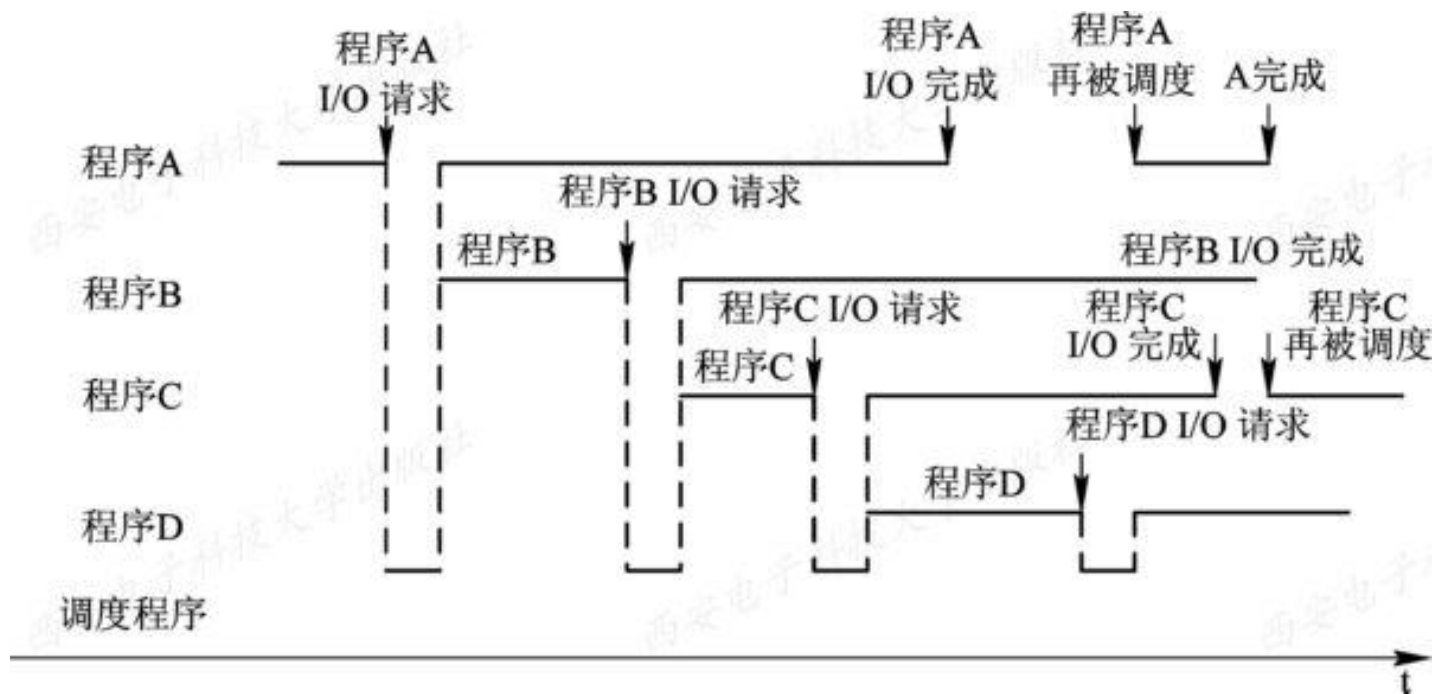


图1-6 多道程序的运行情况

2. 多道批处理系统的优缺点

优点:

➤ 资源利用率高。

程序交替运行，提高CPU利用率；
在内存中装入多道程序可提高内存的利用率；
提高I/O设备的利用率。

➤ 系统吞吐量大。

处理器及其他资源保持忙碌，切换较少，系统开销小，使得系统在单位时间内所完成的总工作量较大

缺点:

- 平均周转时间长。

作业要排队依次进行处理，因而作业的周转时间较长，通常需几个小时

- 无交互能力。

用户无法与程序交互，修改和调试程序极不方便

3. 多道批处理系统需要解决的问题

- **处理机管理问题。**在多道程序之间，应如何分配被它们共享的处理机，使 **CPU** 既能满足各程序运行的需要，又能提高处理机的利用率，以及一旦把处理机分配给某程序后，又应在何时收回等一系列问题，属于处理机管理问题。
- **内存管理问题。**应如何为每道程序分配必要的内存空间，使它们“各得其所”且不致因相互重叠而丢失信息，以及应如何防止因某道程序出现异常情况而破坏其它程序等问题，就是内存管理问题。

- **I/O 设备管理问题。**系统中可能具有多种类型的 I/O 设备供多道程序所共享，应如何分配这些 I/O 设备，如何做到既方便用户对设备的使用，又能提高设备的利用率，这就是 I/O 设备管理问题。
- **文件管理问题。**在现代计算机系统中，通常都存放着大量的程序和数据(以文件形式存在)，应如何组织这些程序和数据，才能使它们既便于用户使用，又能保证数据的安全性和一致性，这些属于文件管理问题。
- **作业管理问题。**对于系统中的各种应用程序，其中有的属于计算型，即以计算为主的程序；有的属于 I/O 型，即以 I/O 为主的程序；又有些作业既重要又紧迫；而有的作业则要求系统能及时响应，这时应如何组织这些作业，这便是作业管理问题

1.2.4 分时系统(Time Sharing System)

1. 分时系统的引入

分时系统(Time Sharing System)与多道批处理系统之间有着截然不同的性能差别，它能很好地将一台计算机提供给多个用户同时使用，提高计算机的利用率。它被经常应用于查询系统中，满足许多查询用户的需要。

分时系统的用户的需求具体表现在以下几个方面：

- **人-机交互。**每当程序员写好一个新程序时，都需要上机进行调试。由于新编程序难免有些错误或不当之处需要修改，因而希望能像早期使用计算机时一样对它进行直接控制，并能以边运行边修改的方式，对程序中的错误进行修改，亦即，希望能进行人-机交互。
- **共享主机。**在 20 世纪 60 年代计算机非常昂贵，不可能像现在这样每人独占一台微机，而只能是由多个用户共享一台计算机，但用户在使用机器时应能够像自己独占计算机一样，不仅可以随时与计算机交互，而且应感觉不到其他用户也在使用该计算机。

2. 分时系统实现中的关键问题

- **及时接收。**要及时接收用户键入的命令或数据并不困难，为此，只需在系统中配置一个**多路卡**。多路卡的作用是使主机能同时接收各用户从终端上输入的数据。此外，还须为每个终端配置一个缓冲区，用来暂存用户键入的命令(或数据)。
- **及时处理。**人机交互的关键，是使用户键入命令后能及时地控制自己作业的运行，或修改自己的作业。为此，各个用户的作业都必须在内存中，且应能频繁地获得处理机而运行；否则，用户键入的命令将无法作用到自己的作业上。
- **前面介绍的批处理系统是无法实现人机交互的。**因为通常大多数作业都还驻留在外存上，即使是已调入内存的作业，也经常要经过较长时间的等待后方能运行，因而使用户键入的命令很难及时作用到自己的作业上。

3. 分时系统的特征

- **多路性。**允许在一台主机上同时联接多台联机终端，系统按分时原则为每个用户服务。宏观上，是多个用户同时工作，共享系统资源；而微观上，则是每个用户作业轮流运行一个时间片。多路性即同时性，它提高了资源利用率，降低了使用费用，从而促进了计算机更广泛的应用。
- **独立性。**每个用户各占一个终端，彼此独立操作，互不干扰。因此，用户所感觉到的，就像是他一人独占主机。
- **及时性。**用户的请求能在很短的时间内获得响应。此时间间隔是以人们所能接受的等待时间来确定的，通常为 1~3 秒钟。
- **交互性。**用户可通过终端与系统进行广泛的人机对话。其广泛性表现在：用户可以请求系统提供多方面的服务，如文件编辑、数据处理和资源共享等。

1.2.5 实时系统(Real Time System)

实时系统(Real Time System): 指系统能及时(或即时)响应外部事件的请求, 在规定的时间内完成对该事件的处理, 并控制所有实时任务协调一致地运行。

1. 实时系统的类型:

- 工业控制系统: 生产控制、火炮的自动控制
- 信息查询系统: 飞机订票系统、股票交易系统
- 多媒体系统
- 嵌入式系统

2. 实时任务的类型

按任务执行时是否呈现周期性划分：

- **周期性任务：** 外部设备周期性地发出激励信号给计算机，要求它按指定周期循环执行，以便周期性地控制某外部设备。
- **非周期性任务：** 外部设备所发出的激励信号并无明显的周期性，但都必须联系着一个截止时间

根据对截止时间的要求划分：

- **硬实时任务：** 系统必须满足任务对截止时间的要求
- **软实时任务：** 系统对截止时间的要求不是非常严格

3. 实时系统与分时系统特征的比较

	实时OS	分时OS
多路性	对多路现场进行采样与控制	为多个终端用户服务
独立性	独立采集信息与控制对象	各用户独立操作，互不干扰
及时性	以被控对象的可接受时间为准	以人可接受时间为准
交互性	仅与特定服务程序交互	与人类用户进行全方位交互
可靠性	高度可靠性	在人类可容忍的误差范围内可靠

```
C:\>
C:\>dir

Volume in drive C is SYSTEM
Volume Serial Number is 3A70-4C95
Directory of C:\


OLD_DOS      1   <DIR>                03-16-09    9:36a
DOS          1   <DIR>                03-16-09    9:36a
COMMAND      COM       54,645    05-31-94    6:22a
WINA20      386        9,349    05-31-94    6:22a
CONFIG      OLD         152    01-02-11    4:05p
WINDOWS     <DIR>                01-02-11    8:16p
DOSIDLE     ASM       56,392    10-09-05    11:06p
DOSIDLE     EXE       12,214    10-09-05    11:06p
DOSIDLE     TXT       27,736    10-09-05    11:06p
FILE_ID     DIZ         408    10-09-05    11:06p
WHATSNEW    TXT       4,017    10-09-05    11:06p
AUTOEXEC    OLD         141    01-02-11    7:13p
CONFIG      SYS       166    01-02-11    8:16p
AUTOEXEC    BAT       152    01-02-11    8:16p

            14 file(s)              165,372 bytes
                        12,879,872 bytes free

C:\>
```

2. 单用户多任务操作系统

含义：只允许一个用户上机，但允许用户把程序分为若干个任务，使它们并发执行，从而有效地改善了系统的性能

代表系统： Windows

1985 年和 1987 年微软公司先后推出了 Windows 1.0 和 Windows 2.0 版本操作系统。

1990 年微软公司又发布了 Windows 3.0 版本，随后又宣布了 Windows 3.1 版本，它们主要是针对 386 和 486 等 32 位微机开发的，较之以前的操作系统有着重大的改进，引入了友善的图形用户界面，支持多任务和扩展内存的功能，使计算机更好使用，从而成为 386 和 486 等微机的主流操作系统

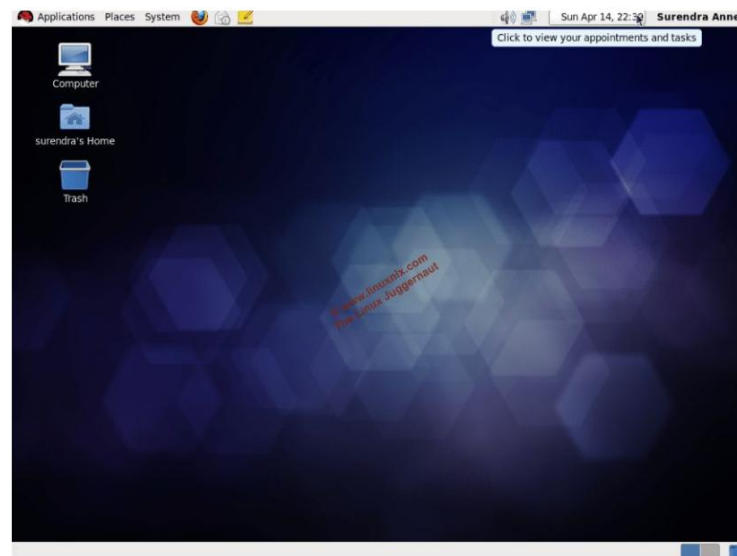


3. 多用户多任务操作系统

含义：允许多个用户通过各自的终端，使用同一台机器，共享主机系统中的各种资源，而每个用户程序又可进一步分为几个任务

代表系统：UNIX OS

UNIX OS 是美国电报电话公司的 Bell 实验室在 1969~1970 年期间开发的，1979 年推出来的 UNIX V.7 已被广泛应用于多种中、小型机上。随着微机性能的提高，人们又将 UNIX 移植到微机上。在 1980 年前后，将 UNIX 第 7 版本移植到 Motorola 公司的 MC 680xx 微机上，后来又将 UNIX V7.0 版本进行简化后移植到 Intel 8080 上，把它称为 Xenix。现在最有影响的两个能运行在微机上的 UNIX 操作系统的变型是 Solaris OS 和 Linux OS。



linux os

课后习题

- 1.简述操作系统的目的和作用
- 2.简述操作系统的发展过程