

项目一

学习操作系统基础知识

项目要点

- 操作系统的基本概念
- 操作系统基本类型
- 操作系统的结构

引言

操作系统是计算机系统中最重要软件之一，它由一组程序组成，这组程序能够有效地组织和管理计算机系统资源，合理地组织计算机的工作流程并控制程序的执行，使计算机高效地运行，并向用户提供各种服务功能，使用户能够方便地使用计算机。操作系统的目标就是管理好资源、方便用户和提高效率。本项目总体介绍了操作系统的发展、基本概念、结构、类型、特征和功能及其运行环境，最后介绍四种当前主流的操作系统。

任务：认识操作系统及硬件环境

任务描述

操作系统是方便用户管理和控制计算机软硬件资源的系统软件（或程序集合）。从用户角度看，操作系统可以看成是对计算机硬件的扩充；从人机交互方式来看，操作系统是用户与机器的接口；从计算机的系统结构看，操作系统是一种层次、模块结构的程序集合，属于有序分层法，是无序模块的有序层次调用。

任务分析

操作系统（Operating System，缩写为 OS）就是有效地管理计算机系统中的各种资源，合理地组织计算机的工作流程，以方便用户使用的一组软件的集合。

准备知识

1. 操作系统的定义

计算机系统由硬件和软件两部分组成。只有硬件设备，计算机系统还无法工作，计算机系统中必须有软件来发挥系统的效能并完成用户的各种应用需求。我们将计算机系统中的各种程序、数据和各种硬件设备统称为计算机系统资源，这样，用户程序的执行过程从宏观上看是在使用整个计算机系统，从微观上看是在使用计算机系统资源，为了使计算机系统能协调一致地工作，就需要对系统中的资源进行管理。由谁来管理计算机系统资源呢？承担这一任务的就是操作系统。

可以从以下 3 个方面来理解操作系统的概念。

①在计算机系统中，中央处理器、存储器、输入/输出设备等所有硬件均称为硬件资源，程序和数据等称为软件资源。使用计算机系统就是使用资源。当程序在系统中运行时，需要操作系统对程序运行所需要的资源进行调度和分配，以保证系统资源的有效利用。

②将计算机内运行的每一个程序称为一个工作流程，这样某段时间内在计算机系统内部可以有多个工作流程同时存在。各个工作流程之间可能存在协作（如程序之间的数据交换）或互斥（如两个程序要使用系统内的某个设备，而该设备只能被一个程序所占有）关系，因此要对程序的运行次序进行协调，以保证这种协作或互斥关系有序进行。如果不对程序的运行次序进行协调就会使系统产生错误或造成计算机系统死机。

拓展提高

计算机系统内各个工作流程运行次序的协调也是由操作系统来完成的。

③操作系统是一组软件构成的集合，在计算机系统中设置这组软件的目的在于方便用户，使计算机系统变得更加易于使用。

2. 操作系统的诞生

从第一台计算机诞生到现在，计算机无论在硬件方面还是在软件方面都取得了很大发展，操作系统也经历了从无到有的过程。

20 世纪 40 年代到 50 年代中期，是计算机的无操作系统时代。20 世纪 50 年代中期出现了第一个简单的批处理操作系统。20 世纪 60 年代中期产生了多道程序批处理系统，不久又出现了以多道程序为基础的分时系统。20 世纪 80 年代是微机和计算机局域网发展的年代，同时也是微机操作系统和局域网操作系统的形成和快速发展的时代。

(1) 人工操作方式

计算机诞生初期并没有操作系统，人们采用手工操作方式使用计算机，信息的输入/输出由人工在联机状态下进行。首先程序员将事先穿孔的纸带（或卡片）装入纸带输入机（或卡片输入机），通过纸带输入机（或卡片输入机）把程序和数据输入给计算机，然后启动计算机运行。程序运行完毕后，才让下一个用户上机。这种方式有以下两个缺点。

①资源独占：每次只允许一个用户使用计算机，一切资源全部由该用户占有，资源利用率低。

②CPU 等待人工操作：当用户进行装纸带（卡片）及卸纸带（卡片）等人工操作时，CPU 处于等待人工操作的空闲状态。

可见，人工操作方式严重降低了计算机资源的利用率，这种人工操作方式与机器利用率相矛盾的问题，随着 CPU 速度的提高以及系统规模的扩大，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，又使 CPU 与 I/O 设备之间速度不匹配的矛盾更加突出。为了缓和此矛盾，引入了脱机输入/输出方式。

(2) 脱机输入/输出技术

为了解决人机矛盾及 CPU 和 I/O 设备之间速度不匹配的矛盾，20 世纪 50 年代末出现了脱机输入/输出技术。该技术是指事先在一台外围机的控制下把纸带（卡片）上的数据（程序）输入到磁带上。当 CPU 需要这些程序和数据时再从磁带机高速输入到内存。



拓展提高

类似地，当 CPU 需要输出数据时直接送到磁带上，然后再在另一台外围机的控制下，将磁带上的结果通过相应的输出设备输出。图 1-1 表示的就是脱机输入/输出过程。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，脱离了主机的控制，故称为脱机输入/输出方式。相反，在主机的直接控制下进行输入/输出的方式为联机输入/输出方式。

脱机输入/输出方式的主要优点如下。

chapter
01chapter
02chapter
03chapter
04chapter
05chapter
06chapter
07

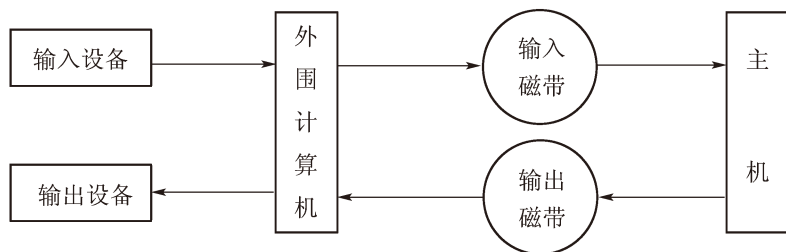


图 1-1 脱机输入 / 输出过程

①减少了 CPU 的空闲时间。当装带（卡）、卸带（卡），以及将数据从低速 I/O 设备送到高速的磁带（或磁盘）上时，都是在脱机情况下进行的。这些工作进行的时候不占用主机时间，不需要主机的干预，主机可以做其他工作，从而减少了 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，直接从高速的磁带或磁盘上将数据调入内存，不再是从低速 I/O 设备上输入，从而大大缓和了 CPU 和 I/O 设备速度不匹配的矛盾，提高了 CPU 的利用率。

（3）批处理技术

早期的计算机系统非常昂贵，为了能充分地利用它，应尽量让系统连续地运行，以减少作业转换产生的空闲时间。为此，通常是把一批作业以脱机输入方式输入到磁带（高速设备）上，并在系统中配备监督程序，在它的控制下一个个装入内存，一个个执行磁带上的作业，使这批作业能一个接一个地连续处理，直到把磁带上的所有作业全部处理完毕。在此期间，建立了以监督程序来管理和控制其他程序的方式，形成了操作系统的雏形。

这种由监督程序控制的系统称为单道批处理系统，它的优点是解决了作业间的自动转换问题，提高了 CPU 的利用率，但还没有真正形成对作业的控制和管理。

（4）多道程序设计技术

20 世纪 60 年代，硬件技术取得了两个方面的重大发展：一是中断技术的引进，二是通道技术的发展。这样，原来由 CPU 直接控制的输入 / 输出工作就转移给了通道，使得 CPU 全部用来进行主要的数据处理工作。过去，内存中只能存放一个用户作业在其中运行，CPU 等待通道传输数据的过程中，仍然因无工作可做而处于空闲状态。若在主存中同时存放多个作业，那么 CPU 在等待一个作业传输数据时，就可去执行内存中的其他作业，从而保证 CPU 以及系统中的其他设备尽可能得到充分的利用。



知识链接

为了提高批处理技术中程序的并行执行能力，提高资源的利用率，采用作业调度程序同时把几个作业放入内存，并允许它们交替执行，即多道程序设计技术。

单道程序与多道程序的执行过程分别如图 1-2 和图 1-3 所示。

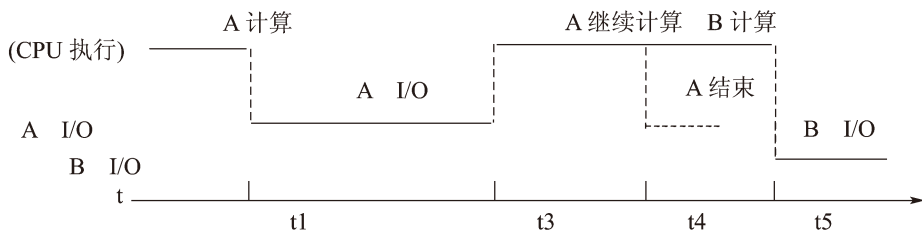


图 1-2 单道程序工作过程

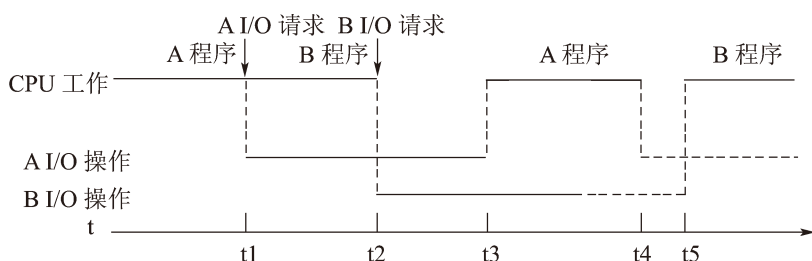


图 1-3 多道程序执行过程

在操作系统中引入多道程序设计技术以后，会使系统具有以下特征。

①多道性。在内存中可同时驻留多道程序，并允许它们并发执行，从而有效提高了资源的利用率和系统的吞吐量。

②无序性。多个作业完成的先后顺序与它们进入内存的顺序之间无严格的对应关系，即先进入内存的作业可能较后或最后完成，而后进入内存的作业可能先完成。

③宏观上并行、微观上串行。从宏观上看同时存在于内存中的多道作业都处于运行状态，它们先后开始了各自的运行，但又都未运行完毕，好像多道作业在并行运行。但从微观上看，由于我们讨论的是单 CPU 系统，内存中的多道作业轮流、交替地使用 CPU 系统，所以各作业仍是串行的。



知识链接

多道程序设计系统的出现标志着操作系统进入渐趋成熟的阶段，先后出现了作业调度管理、处理机管理、存储器管理、设备管理和文件系统管理等功能。

3. 操作系统的地位

操作系统在计算机系统中处于一个什么样的位置，与计算机硬件和其他的计算机软件之间又是什么关系？操作系统在整个计算机系统中所处的位置和它与计算机硬件以及其他计算机软件之间的关系如图 1-4 所示。

操作系统是现代计算机系统的重要组成部分，

用户
应用软件或应用系统
其他系统软件
操作系统
计算机硬件

图 1-4 计算机系统的抽象层次结构

- chapter 01
- chapter 02
- chapter 03
- chapter 04
- chapter 05
- chapter 06
- chapter 07

它是计算机系统运行和工作必不可少的软件，各种类型的计算机系统都离不开操作系统。可以看出，操作系统是在计算机硬件的基础上对硬件进行的第一层扩充，它是计算机系统中最核心的系统软件，其他的系统软件和应用软件都是在操作系统的基础上构建起来的。操作系统介于计算机硬件和计算机用户之间，它与计算机硬件、软件、应用系统，以至于计算机用户都有千丝万缕的联系。



拓展提高

一台没有任何软件配置和支持的计算机被称为“裸机”，要让裸机接受用户发出的命令，执行相应的操作是非常困难的。操作系统在硬件之上建立了一个服务体系，为操作系统以外的系统软件 and 用户应用软件提供了强大的支持，用户通过这个服务体系操作和使用计算机系统，面对的是一个非常友好的、方便的环境界面，因而用户面对的是一个更加易于使用的计算机系统。

4. 操作系统的特征

设置操作系统的目的在于提高系统的效率，增强系统的处理能力，充分发挥系统资源的利用率，方便用户使用。以多道程序设计为基础的现代操作系统具有以下主要特征。

(1) 并发性 (Concurrence)

在操作系统中，并发是指多个事件在同一时间间隔内发生。对计算机而言，并发是指在一段时间内，多道程序“在宏观上同时运行”。显然，多道和并发是同一个事物的两个方面，正是由于多道程序设计的实现才导致了多个程序的并发执行。而程序的并发执行导致了多个程序竞争一台计算机，使得并行运行中的任何一个程序都处于已开始运行但又未结束的状态。

现代操作系统是并发系统的管理机构，其本身就是与用户程序一起并发执行的。程序的并发执行带来了程序串行执行所没有的新问题，并导致操作系统对程序管理以及操作系统本身的复杂化。

(2) 虚拟性 (Virtual)

虚拟是指把一个物理实体映射为多个逻辑意义上的实体。前者是客观存在的，后者是虚构的，是一种感觉性的存在，即主观上的一种假象。例如，在多道程序系统中，虽然只有一个 CPU，每次只能执行一道程序，但采用多道程序技术后，在一段时间间隔内，宏观上有多个程序在运行。在用户看来，就好像有多个 CPU 在各自运行自己的程序。这种情况就是将一个物理的 CPU 虚拟为多个逻辑上的 CPU。逻辑上的 CPU 称为虚拟处理机，类似的还有虚拟存储器 and 虚拟设备等。

(3) 共享性 (Sharing)

操作系统是多道程序的管理机构。它使多个用户作业共享有限的计算机系统资源。由于资源是共享的，就必然会导致如何在多个作业之间合理地分配和使用资源，并且如何充分发挥计算机系统资源的利用效率的问题。从这个意义上讲，操作系统就是一

个计算机系统的资源管理程序。

计算机系统的所有资源都是共享的，但共享又分成两种不同的类型，即互斥共享和同时共享。所谓互斥共享是指资源的分配以作业（或进程）为单位，当一个作业未使用完这个资源前，别的作业不能同时使用。总的来说，这类资源毕竟是所有的作业（或进程）都可以使用的，故称为互斥共享。而同时共享，则是指多个作业“同时”使用资源，即当一个作业已开始使用某个资源但又尚未使用完毕，另一个作业也能使用。



知识链接

这种同时使用也是指在一段相当小的时间范围内允许多个作业轮流使用，因为一个物理设备是不可能在同一时刻真正同时完成多个任务的。

（4）不确定性（Nondeterministic）

所谓操作系统的不确定性，是指在操作系统控制下多道作业的执行顺序和每个作业的执行时间是不确定的。例如，有三个作业，两次或多次运行的执行序列可能不相同，每一个作业占有计算机的时间也可能不相同。

5. 操作系统的功能

从资源管理的角度出发，作为管理计算机系统资源并控制程序运行的操作系统，其功能可以简单归纳如下。

（1）处理机管理

在单道作业或单用户的环境下，处理机被一个作业或几个用户所独占，对处理机的管理十分简单。但在多道程序或多用户的环境下，要组织多个作业同时运行，就要解决处理机的管理问题。在多道程序环境下，处理机的分配和运行都是以进程为单位的，因而对处理机的管理可归结为对进程的管理，包括以下几个部分。

①进程控制。进程控制的主要任务是为作业创建进程，撤销已结束的进程以及控制进程在运行过程中的状态转换。

②进程调度。进程调度的任务就是从进程的就绪队列中，按照一定的算法选择一个进程，把处理机分配给它，并为它设置运行现场，使之投入运行。

③进程同步。为使系统中的进程有条不紊地运行，系统必须设置进程同步机制，以协调系统中各进程的运行。

④进程通信。系统中的各进程之间有时需要合作，这就需要进行通信来交换信息。

（2）存储管理

存储管理的主要任务是为多道程序的运行提供良好的环境，方便用户使用存储器，并提高主存的利用率。存储管理包括以下几类。

①地址重定位。在多道程序设计环境下，每个作业是动态装入主存的，作业的逻辑地址必须转换为主存的物理地址，这一转换称作地址重定位。

②存储分配。存储管理的主要目的是为每道程序分配内存空间，在作业结束时要收回它所占用的空间。

chapter
01chapter
02chapter
03chapter
04chapter
05chapter
06chapter
07

③存储保护。保证每道程序都在自己的主存空间运行，各道程序互不侵犯，尤其是不能侵犯操作系统空间。

④存储扩充。一般来说，主存的容量是有限的。在多道程序设计环境下往往感到主存容量不能满足用户作业的需要。为此，操作系统存储管理的任务是扩充主存容量，这种扩充通过建立虚拟存储系统来实现，是逻辑上的扩充。

(3) 设备管理

一个计算机系统的硬件，除了 CPU 和主存，其余几乎都属于外部设备，外部设备种类繁多，物理特性相差很大。因此，操作系统的设备管理往往很复杂。设备管理包括以下几类。

①缓冲管理。由于 CPU 和 I/O 设备的速度相差很大，为缓和这一矛盾，通常在设备管理中建立 I/O 缓冲区，而对缓冲区进行有效管理是设备管理的一项任务。

②设备分配。当用户发生 I/O 请求后，设备管理程序要依据一定的策略并根据系统中的设备情况，将所需设备分配给它。设备用完后要及时回收。

③设备处理。设备处理程序又称设备驱动程序，对于未设置通道的计算机系统，其基本任务通常是实现 CPU 和设备控制器间的通信。即由 CPU 向设备控制器发出 I/O 指令，要求它完成指定的 I/O 操作，并能够接受由设备控制器发来的中断请求，给予及时的响应和处理。对于设置了通道的计算机系统，设备处理程序还应能根据用户的 I/O 请求，自动构造通道程序。

④设备独立性和虚拟设备。设备独立性是指应用程序独立于物理设备，使用户编程与实际的物理设备无关。虚拟设备就是将一台物理设备映射为多台逻辑上的设备。

(4) 文件管理

文件管理主要对系统中的软件资源进行管理，即对用户文件和系统文件进行管理。文件管理包括以下几项。

①目录管理。为方便用户在文件存储器中找到所需文件，通常由系统为每一文件建立一个目录项，包括文件名、属性以及存放位置等，若干目录项又可构成一个目录文件。目录管理的任务是为每个文件建立目录项，并对目录项施以有效的组织，以方便用户按名存取。

②文件读、写管理。文件读、写管理是文件管理的最基本功能之一。文件系统根据用户给出的文件名去查找文件目录，从中得到文件在文件存储器上的位置，然后利用文件读、写指针，对文件进行读、写操作。

③文件存取控制。为了防止系统中的文件被非法窃取或破坏，在文件系统中应建立有效的保护机制，以保证文件系统的安全性。

④文件存储空间管理。所有的系统文件和用户文件都存放在文件存储器上，文件存储空间管理的任务是为新建文件分配存储空间，文件被删除后应及时回收所占用的空间。文件存储空间管理的目标是提高文件存储空间的利用率，并提高文件系统的工作效率。

(5) 用户接口

操作系统除了对资源进行管理，还为用户提供相应的接口，通过使用这些接口达到方便使用计算机的目的。

①命令接口。命令接口也称作业级接口，分为联机命令接口和脱机命令接口。联机命令接口是为联机用户提供的，它由一组键盘命令及其解释程序组成，当用户在终端或控制台输入一条命令后，系统便自动转入命令解释程序，对该命令进行解释并执行。在完成指令操作后，控制又返回到终端或控制台，等待接受用户输入下一条命令，这样用户可通过不断键入不同的命令达到控制作业的目的。



知识链接

脱机命令接口是为批处理系统的用户提供的，在批处理系统中，用户不直接与自己的作业进行交互，而是使用作业控制语言的语句，将用户对其作业控制的意图写成作业说明书，然后将作业说明书连同作业一起，以卡片的形式提交给系统。当系统调度该作业时，通过解释程序对作业说明书进行逐条解释并执行。这样，作业一直在作业说明书的控制下运行，直到遇到作业结束语句时系统停止该作业的执行。这种工作方式已成为历史。

②程序接口。程序接口是用户获取操作系统服务的唯一途径。程序接口由一组系统调用组成。每一个系统调用都是一个完成特定功能的子程序。早期的操作系统，系统调用都是用汇编语言写成的，因而只有在汇编语言写的应用程序中才可以直接调用，而在高级语言及 C 语言中，往往提供与系统调用一一对应的库函数，应用程序通过调用库函数来使用系统调用。近年来推出的操作系统中，如 UNIX 系统，系统调用是用 C 语言编写，并以函数形式提供的，从而可在 C 语言编写的程序中直接调用。

③图形接口。图形接口不需要记忆命令，图形接口的目标是对出现在屏幕上的对象直接进行操作，以控制和操纵程序的运行。这种图形用户接口大大减低了用户记忆的工作量，将受用户的欢迎。图形用户接口的主要构件是：窗口、菜单和对话框。

6. 操作系统的基本类型

不同的应用领域、不同的应用目的，对操作系统提出了不同的要求。如果让一个操作系统满足所有要求，那么这个操作系统将庞大无比，难有高效。因此，人们根据不同的环境，分别设计了不同类型的操作系统。

(1) 批处理系统

批处理系统也称为作业流处理系统，主要用在科学计算的大中型机上。它的特点是采用脱机技术将众多的作业送入计算机系统，然后由批处理系统按批选择作业进行处理。

chapter
01chapter
02chapter
03chapter
04chapter
05chapter
06chapter
07



知识链接

在作业处理的过程中不需要用户的控制和干预，它所追求的目标是系统吞吐量，作业周转时间短，资源使用效率高等。批处理系统可以分为单道批处理、多道批处理。

①单道批处理。在早期计算机系统中配置的一种操作系统类型（类似于 DOS 下的批处理文件），其特征如下。

作业依照在外存中排定的次序依次进入系统，不需作业调度和进程调度。

内存中仅有一道作业在运行。

作业完成次序依赖于进入系统的次序，即按顺序进行。

②多道批处理操作系统结合多道程序设计技术的批处理系统，其特征如下。

作业进入系统并执行需经过二级调度，即作业调度和进程调度。

内存中可同时驻留多道作业，这些作业的运行，在宏观上并行、微观上串行。

作业完成次序与进入系统的次序无关。

（2）分时系统

分时系统是多用户共享系统，一般使用一台计算机连接多个终端，各用户通过相应的终端使用计算机。其特点是人一机交互性，即用户通过终端控制台向计算机主机提出处理请求，在主机上运行的操作系统检查用户提出的请求的合法性，检查通过后对该请求进行处理，然后将处理结果反馈给终端上的用户。

当多个用户同时工作时，在操作系统的管理下将 CPU 的执行时间划分成若干个时间片，轮流分给每一个终端上的用户服务，以合理的响应时间满足用户的要求，使得每一个用户都感觉自己独占一台计算机，分时工作方式如图 1-5 所示。

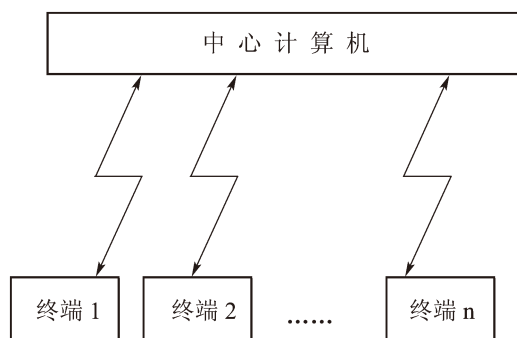


图 1-5 分时工作方式

同时性：若干用户可以同时操作，共同使用系统同一资源。

独立性：用户在各自的终端上工作互不干扰，由于时间片很短，尽管系统轮流服务于多个用户，但每个用户都可以得到响应，没有明显等待的感觉，因此每个用户都认为自己独占计算机系统。

及时性：用户的请求能得到及时响应。

交互性：系统以对话方式为各个终端用户服务，用户在终端上通过与系统交互会话请求系统的服务，直接控制作业的运行。

（3）实时系统

实时系统一般总是以专用系统的身份出现，可分为实时控制系统和实时信息处理系统两种类型。

①实时控制系统。当把计算机用于生产过程的控制，并形成以计算机为中心的控制系统时，系统要求能实时采集现场数据，并对所采集的数据进行及时的处理，进而自动地控制相应的执行机构，使某些参数如温度、压力及方位等能按预定的规律变化，以保证产品的质量，提高产量。



知识链接

类似地，也可将计算机用于武器的控制，如火炮的自动控制系统、飞机的自动驾驶系统以及导弹的制导系统等。通常把要求进行实时控制的系统统称为实时控制系统。

②实时信息处理系统。通常，把要求对信息进行实时处理的系统，称为实时信息处理系统。该系统由一台或多台通过通信线路连接的远程终端组成，计算机接受从远程终端发来的服务请求，根据用户提出的问题，对信息进行检索和处理，并在很短的时间内为用户做出正确的回答。典型的实时信息处理系统有：飞机订票系统及情报检索系统等。

实时控制系统和实时信息处理系统统称为实时系统。所谓“实时”，是表示“及时”“即时”。而实时系统是指系统能及时（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。

（4）网络操作系统

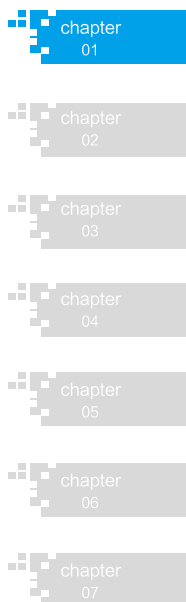
简单地说，网络操作系统就是在计算机网络环境下具有网络功能的操作系统。

所谓计算机网络，就是一个数据通信系统，通过它把地理上分散的计算机和终端设备连接起来，以达到数据通信和资源共享的目的。



在计算机网络系统中，网络上各节点机的硬件特性不同，数据表示格式及其他方面的要求也不同。为了能够正确地进行相互通信并相互理解通信内容，各通信方之间应有许多约定，这些约定称作协议。因此，网络操作系统(Network Operating System, NOS)是使网上各计算机能方便有效地进行数据通信和资源共享，为网络用户提供所需的各种网络服务的软件集合。

网络操作系统首先是一个操作系统，因此它应具有通常操作系统具有的处理机管理、存储管理、设备管理和文件管理的功能。除此之外，作为网络操作系统，它还具有以下功能。



实现网络中各节点机之间的通信。

实现网络中硬、软件资源的共享。

提供多种网络服务软件。

提供网络用户的应用程序接口。

分布式操作系统。

分布式计算机系统（简称分布式系统）是由多台计算机组成的系统，且满足以下条件。

①系统中任意两台计算机之间可以利用通信来交换信息。

②系统中各台计算机之间无主次之分，既无控制整个系统的主机，也无受制于他机的从机。

③系统中的资源为系统中的所有用户共享。用户往往只需了解系统中是否具有自己所需要的资源，而无须了解该资源位于哪台计算机上。

④系统中的若干台计算机可以相互合作，共同完成同一任务。分布式操作系统的主要特点是各节点的自治性，资源共享的透明性，各节点间的协同性，系统的坚定性。分布式操作系统和单机的集中式操作系统的主要区别表现在通信、资源管理和系统结构3个方面。

分布式操作系统的主要特点如下。

①系统状态的不精确性。在分布式系统中，系统内的各节点是自治的，它们并不向外界报告它们的状态信息。因此，在系统中很难获得完整的系统状态信息。另一方面，系统在收集各节点的状态信息时，因为信息在网上传输需要一定的时间，因而收到的信息可能是过时的。这些信息不能确切反映系统的当前状态。

②控制结构的复杂性。在分布式系统中，各节点机之间不存在主从关系或层次关系，因而增加了控制机构的复杂性。首先，由于各节点的自治性，它们之间发生冲突的概率要高得多，使同步问题变得复杂，死锁问题也难以处理。其次，由于系统的透明性要求，使得系统故障的检测和用户操作的检查都增加了难度。

③通信开销引起性能下降。网络通信开销是分布式系统开销的一个重要组成部分。在某些系统中，由于实时通信开销过大而被迫放弃。

7. 操作系统的结构

操作系统是由很多相关模块组成的一组程序，它的每个模块的质量固然重要，它的整体结构更是有着决定性的作用。操作系统结构的研究引起了很多人的注意。人们通过软件开发的实践逐步认识到，软件结构是影响软件质量的内在因素，良好的结构可以提高软件的正确性、可维护性和工作效率。

（1）无序模块式

早期的操作系统大多采用无序模块式结构，现在的一些小型操作系统仍采用这种结构。在这些系统中每个模块都有定义良好的接口，相互间的调用不受约束。这种方法的主要优点是结构紧密，组合方便，灵活性大。主要缺点是模块独立性差，系统结构不清晰，难以保证可靠性。操作系统规模庞大，结构复杂。



知识链接

它的研制周期很长，从提出需求、明确规格说明起到投入运行，一般历经数年。一个大型操作系统的研制，往往需要数千人年的工作量，而且很难保证正确性。操作系统模块之间的关系十分重要，如果不分层次、不加约束，就不能有效地工作。

(2) 分层式

荷兰科学家狄克斯特拉 (E.F.Dijkstra) 在 1968 年首次提出层次结构法。这种结构法将操作系统的模块分成多个层次，各层次之间是单向依赖关系。单向依赖是指只允许上层的模块调用下层的模块，不能反向调用。典型的操作系统的分层结构如图 1-6 所示，层次结构的中心是裸机本身所提供的各种功能，向外扩展的每一层都提供一种功能，这种功能只依赖该层以内的各个层次，这种结构的各层依次组成了一系列虚拟机，紧挨裸机的是操作系统内核，而最外层就是具备用户所需要功能的虚拟机。

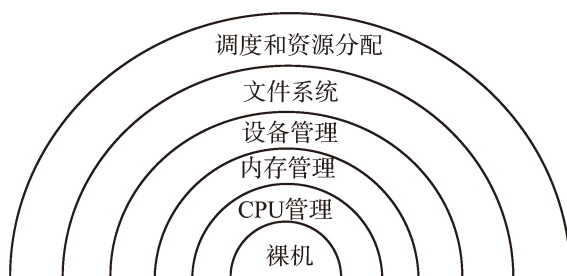


图 1-6 操作系统分层结构

实际运行的操作系统很少具有这样清晰的层次结构，因为这些操作系统开始设计时，人们对操作系统的结构原理并不像现在这样清楚，而且往往在设计阶段结束后会因某种需要而加进新的内容，以至损害了操作系统原来的层次结构。

(3) 客户/服务器模型

采用客户/服务器模型构造的操作系统的基本思想是把操作系统划分为若干进程，其中每个进程实现单独的一套服务（功能）。例如文件服务、进程服务、处理机调度服务及存储服务。每一个服务对应一个服务器，每个服务器都运行在用户态，并执行一个循环，在执行循环过程中不断检查是否有客户请求该服务器提供的某种服务。



知识链接

客户可以是一个应用程序，也可以是另一操作系统程序。它通过发送一条消息给服务器请求一项服务。运行在核心态下的操作系统内核把消息传送给服务器。

由服务器执行具体操作，其结果经由内核以消息的形式返回给用户，如图 1-7 所示。

chapter
01

chapter
02

chapter
03

chapter
04

chapter
05

chapter
06

chapter
07

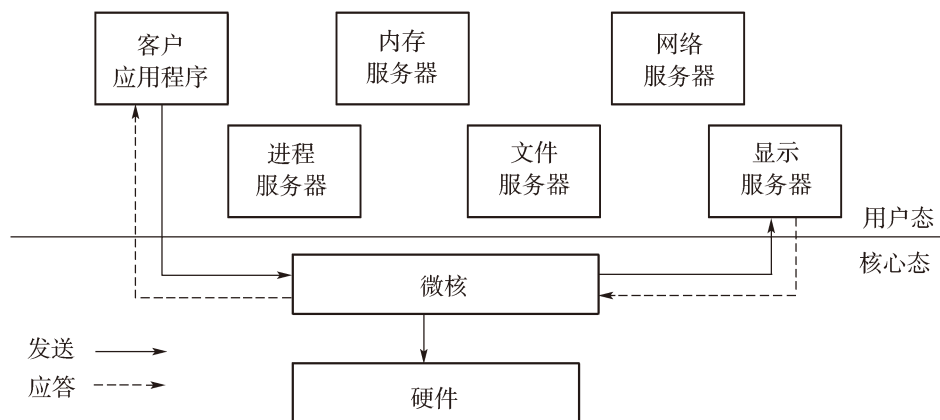


图 1-7 客户 / 服务器操作系统

在这种模型中，内核只执行很少的任务，称为微核或微内核。采用客户 / 服务器模型的好处有如下几点。

简化操作系统核心。因为可以把很多功能作为独立的服务器进程移出核心。
改进可靠性和独立性。每个服务器在自己的地址空间中独立运行，因而防止了受其他进程的影响。此外，由于服务器运行在用户态，它们不直接访问硬件或者修改执行体的内存空间。

完全适宜于分布式计算模型。由于联网的计算机是以客户 / 服务器模型为基础，并且使用消息来通信，本地服务器可以较容易地发送消息给客户应用程序的远程计算机。客户不需要知道某些请求是在本地得到服务还是在远程得到服务。

任务实施

任何系统软件都是建立在硬件基础上的，离不开硬件的支持。操作系统是加在裸机上的第一层软件，它的功能与运行直接依赖于硬件环境，与硬件的关系尤为密切。

(1) CPU 与外设并行工作

为了获得 CPU 与外设之间更高的并行能力，也为了让种类繁多、物理特性各异的外设能以标准的接口方式连到系统中，计算机系统中引入了各自独立的通道结构。

通道又称 I/O 处理机，它能完成主存和外设间的信息传输，并与中央处理器并行操作。采用通道技术实现了 I/O 操作的独立性和各部件工作的并行性。通道把 CPU 从繁琐的输入输出操作中解放出来。采用通道技术后，不仅能够实现 CPU 与通道的并行操作，而且通道与通道之间也能实现并行操作，各通道上的外设也能实现并行操作，从而可达到提高计算机系统工作效率的目的。



知识链接

通常一个 CPU 主存可以连接若干个通道，一个通道可以连接若干个控制器，一个控制器又可以连接若干个设备，即所谓四级连接。三级控制是指 CPU 执行 I/O 指令实施对通道的控制，通道执行通道命令对控制器实施控制，控制器控制设备执行相应的输入 / 输出操作。

(2) I/O 中断

什么是中断？从操作系统的观点来看，中断是指对异步或例外事件的一种响应，该响应自动保存 CPU 状态，自动转入中断处理程序。中断的引进，最初是为了实现外部设备和 CPU 的并行工作，这一概念后来被扩展了。

现在，在系统中发生的需要处理机暂停正在执行的程序转而去进行所要干预的所有事件，都要通过中断机构进行处理。这些事件（称为中断源）的类型与计算机的系统结构以及操作系统的设计关系重大。在计算机系统中，一般将中断分为如下几种类型。

- ① I/O 中断。这是外部设备完成了预定 I/O 操作或 I/O 操作中出错所引起的中断。
- ② 程序中断。这是程序中的错误引起的中断。
- ③ 硬件故障中断，或称机器检验中断。
- ④ 外中断。外中断来自外部信号，这些信号可能来自其他的机器。外中断还包括时钟中断，以及来自键盘的中断。
- ⑤ 访管中断。这是由机器中的访管指令引起的中断。

用户程序在执行过程中，当中断发生时，硬件的中断机构首先判断发生的中断类型，保护当前的处理机状态，以便将来进行恢复。其次，根据中断类型自动转入预先规定好的中断处理程序。最后，当中断处理程序结束时，恢复现场并返回被中断的程序。中断的处理过程如图 1-8 所示。图中的 PSW 为程序状态字，用来控制指令的执行顺序、保存和指示与当前正在执行的程序有关的系统状态，即所说的处理现场。

知识链接

每一类型的中断都有与之相关的两个程序状态字，即旧 PSW 和新 PSW。中断前的程序状态字称为旧 PSW，中断后的程序状态字称为新 PSW。

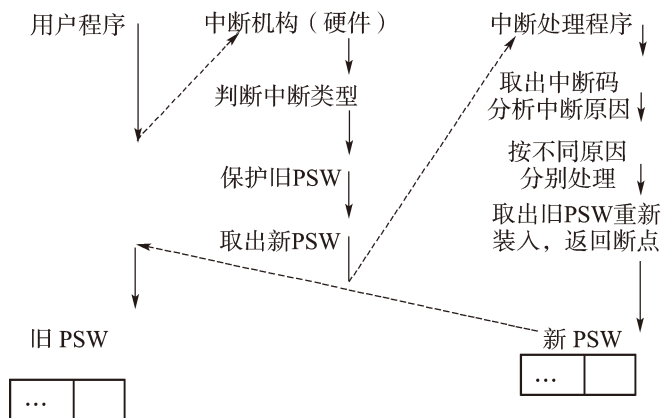


图 1-8 中断的处理过程

(3) 管态与目态

为了便于构造安全可靠的操作系统，现代计算机硬件都提供了两种处理机状态：

- chapter 01
- chapter 02
- chapter 03
- chapter 04
- chapter 05
- chapter 06
- chapter 07

管态和目态。该状态由一位触发器标识，通常属于程序状态字 PSW 的一部分，即 PSW 中的一位。管态也称核心态或系统态。机器处于管态时程序可执行硬件所提供的全部指令，包括特权指令和非特权指令。由于利用特权指令可修改程序状态字，因而在管态下可以修改机器状态，通常操作系统程序在管态下运行。

目态也称常态或用户态。机器处于目态时程序只能执行硬件机器指令系统的一个子集，即非特权指令集合。通常，用户程序在目态下运行。如果用户程序在目态执行特权指令，硬件将产生中断，由操作系统获得控制，特权指令的执行被制止。设置管态和目态的目的是赋予操作系统特权执行某些特殊指令，以保证计算机系统的安全工作。

(4) 存储保护

在多道程序计算机系统中，除了操作系统驻留内存外，可允许多个用户作业同时驻留内存。为了防止执行中的用户程序对其他用户程序或操作系统造成有意或无意的破坏，计算机系统必须提供存储保护机构，用以限定程序有权访问的内存地址范围，这是保证系统安全运行的基本条件之一。有两种常用的硬件存储保护技术。

① 界限寄存器

界限寄存器方法是在 CPU 中设置一对界限寄存器，分别存放现行程序在内存中的下限地址和上限地址，每当执行内存访问操作时，硬件将自动检查被访问的内存地址是否处于寄存器所限定的地址范围内，若超出范围便产生地址越界中断，表示这是非法访问。只有操作系统可以访问全内存。

② 存储保护键

存储保护键是由若干二进制组成的标志。一些计算机系统将内存划分成若干定长的存储块，并赋予每个存储块一个附加的不在编址范围内的存储保护键。当有作业进入内存时，操作系统赋予它一个唯一的保护键码，并将分配给该作业的各存储块也设置成同样的保护键码。



知识链接

当该作业被调度到 CPU 上执行时，操作系统同样将其保护键码置入现行 PSW 中“键”字段中。此后每当执行内存访问操作时，硬件将先检查该存储块的保护键码与现行 PSW 中键值是否匹配。若匹配才允许访问。对操作系统程序通常赋予一个特殊的保护键码，如二进制组成的全“0”或全“1”码。它赋予操作系统可以访问全内存的特权。

项目小结

本项目主要讲解了计算机操作系统的相关知识，主要包括操作系统的形成与发展、操作系统的概念、操作系统的结构、硬件环境等主要内容。

项目考核



填空题

1. 20 世纪 80 年代是 _____ 和 _____ 发展的年代，同时也是微机操作系统和局域网操作系统的形成和快速发展的时代。

2. 各个工作流程之间可能存在 _____ 或 _____ 关系，因此要对程序的运行次序进行协调，以保证这种协作或互斥关系有序进行。



判断题

1. 20 世纪 40 年代到 50 年代中期，是计算机的无操作系统时代。20 世纪 50 年代中期出现了第一个简单的批处理操作系统。（ ）

2. 计算机系统由硬件和软件两部分组成。只有硬件设备，计算机系统还无法工作，计算机系统中必须有软件来发挥系统的效能并完成用户的各种应用需求。（ ）

3. 操作系统是由很多相关模块组成的一组程序，它的每个模块的质量固然重要，它的整体结构更是有着决定性的作用。操作系统结构的研究引起了很多人的注意。（ ）



问答题

1. 什么是操作系统？操作系统的基本特征是什么？
2. 操作系统在计算机系统中处于什么地位？具有哪些功能？
3. 操作系统具有哪些基本类型？
4. 操作系统提供哪些接口？它们的作用是什么？
5. 操作系统的结构在发展过程中发生了哪些变化？
6. 什么是通道？通道的作用是什么？
7. 什么是管态和目态？为什么设置管态和目态？

chapter
01chapter
02chapter
03chapter
04chapter
05chapter
06chapter
07