

从层次结构上来看,计算机系统自下而上可粗略分为 4 个部分:硬件、操作系统、应用程序和用户。前面的章节中主要从硬件底层的角度讲述了计算机系统的组成和工作原理,如中央处理器、内存、输入输出设备等,它们提供了基本的计算资源。然而在信息时代,软件被称为计算机系统的灵魂。作为软件核心的操作系统,已经与现代计算机系统密不可分、融为一体。操作系统管理各种计算机硬件,为应用程序提供基础,并充当计算机硬件与用户之间的中介。应用程序,如字处理程序、电子制表软件、编译器、网络浏览器等,规定了按何种方式使用这些资源来解决用户的计算问题。操作系统控制和协调各用户的应用程序对硬件的分配与使用。在计算机系统的运行过程中,操作系统提供了正确使用这些资源的方法。

综上所述,操作系统(Operating System, OS)是指控制和管理整个计算机系统的硬件和软件资源,并合理地组织调度计算机的工作和资源的分配,以提供给用户和其他软件方便的接口及环境的程序集合。计算机操作系统是随着计算机研究和应用的发展逐步形成并发展起来的,它是计算机系统中最基本的系统软件。

5.1 操作系统的目标和作用

操作系统是管理计算机硬件与软件资源的程序,同时也是计算机系统的内核与基石。操作系统是控制其他程序运行,管理系统资源并为用户提供操作界面的系统软件的集合。操作系统负责诸如管理与配置内存、决定系统资源供需的优先次序、控制输入与输出设备、操作网络与管理文件系统等基本事务。首先我们来看一下操作系统的定义,计算机用户都曾使用过各种各样的操作系统,什么是操作系统?不同的人对操作系统有着不同的看法。但总体来说,操作系统是一个管理计算机系统资源、控制程序运行的系统软件,它为用户提供了一个方便、安全、可靠的工作环境和界面。作为计算机系统的控制和指挥中心,操作系统不仅是一个软件(Software),而且是一个系统软件(System Software),所谓系统软件是指控制和协调计算机及外部设备,支持应用软件开发和运行的系统,是无须用户干预的各种程序的集合,主要功能是调度、监控和维护计算机系统;负责管理计算机系统中各种独立的硬件,使得它们可以协调工作。系统软件使得计算机使用者和其他软件将计算机当作一个整体而不需要顾及底层每个硬件是如何工作的。

系统软件的运行既依赖计算机系统的硬件(Hardware),又要管理计算机系统的一切硬件设施。在操作系统运行过程中,需要硬件强力的支持,而且有一部分功能是由硬件直接完成,从这个意义上讲,操作系统又不完全是软件,而是由软硬件结合的有机体,在软硬件的配合下,共同完成操作系统所应完成的任务(如图 5-1 所示)。它的主要特点如下。

- **方便:** 操作系统使计算机更易于使用。

- **有效**：操作系统允许以更有效的方式使用计算机系统资源。
- **可扩充性**：在操作系统中,允许有效地开发、测试和引进新的系统功能。
- **开放性**：实现应用程序的可移植性和互操作性,要求具有统一的开放环境。

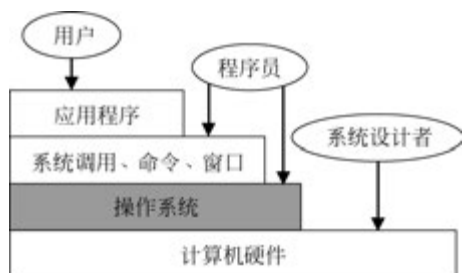


图 5-1 操作系统在计算机系统的位置

5.1.1 作为用户/计算机接口的操作系统

所谓操作系统的用户接口,就是操作系统提供给用户,使用户可通过它们调用系统服务的手段。任何操作系统管理计算机资源的目的都在于将计算机资源提供给用户使用,用户通过用户接口使用操作系统,用户接口是操作系统的五大功能之一,为用户提供统一的接口是操作系统的目标之一。在一般情况下,一个完整的操作系统在启动后就会给用户提供一个对计算机进行操作的界面。例如,DOS 操作系统会在显示器上显示一个字符操作界面;Windows 操作系统会显示一个图形界面。这样,用户就可以按照输入命令的方式来使用操作系统的某种功能。用户接口主要分为如下三类。

- (1) 命令接口,如图 5-2 所示,是以联机命令方式提供的用户接口。
- (2) 图形接口,如图 5-3 所示,是以图形方式提供的用户接口。
- (3) 程序接口,是以程序调用形式提供的用户接口。

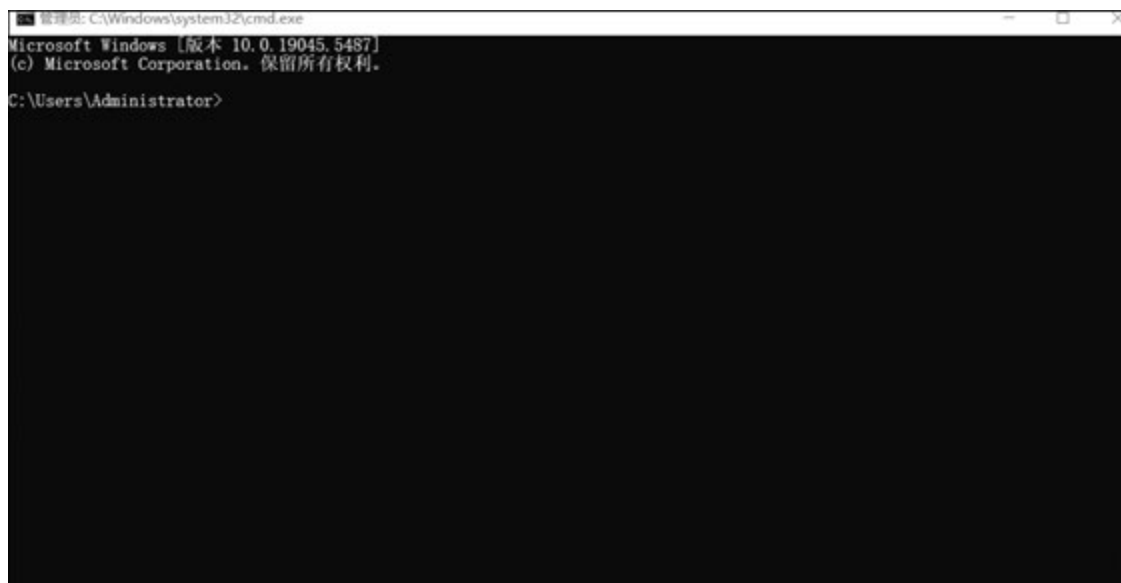


图 5-2 Windows 命令窗口

如图 5-3 所示,操作系统通常提供以下几方面的服务。

- (1) **程序开发**：操作系统提供各种各样的工具和服务,如编辑器和调试器,用于帮助程序



图 5-3 Windows 7 系统的窗口

员开发程序。通常,这些服务以实用工具程序的形式出现,严格来说并不属于操作系统核心的一部分;它们由操作系统提供,称作应用程序开发工具。

(2) **程序运行**: 运行一个程序需要很多步骤,包括必须把指令和数据载入内存、初始化 I/O 设备和文件、准备其他一些资源。操作系统为用户处理这些调度问题。

(3) **I/O 设备访问**: 每个 I/O 设备的操作都需要特有的指令集或控制信号,操作系统隐藏这些细节并提供了统一的接口,因此程序员可以使用简单的读和写操作访问这些设备。

(4) **文件访问控制**: 对操作系统而言,关于文件的控制不仅必须详细了解 I/O 设备(磁盘驱动器、磁带驱动器)的特性,而且必须详细了解存储介质中文件数据的结构。此外,对有多个用户的系统,操作系统还可以提供保护机制来控制对文件的访问。

(5) **系统访问**: 对于共享或公共系统,操作系统控制对整个系统的访问以及对某个特殊系统资源的访问。访问功能模块必须提供对资源和数据的保护,以避免未授权用户的访问,还必须解决资源竞争时的冲突问题。

(6) **错误检测和响应**: 计算机系统运行时可能发生各种各样的错误,包括内部和外部硬件的错误,如存储器错误、设备失效或故障,以及各种软件错误,如算术溢出、试图访问被禁止的存储器单元、操作系统无法满足应用程序的请求等。对每种情况,操作系统都必须提供响应以清除错误条件,使其对正在运行的应用程序影响最小。响应可以是终止引起错误的程序、重试操作或简单地给应用程序报告错误。

(7) **记账**: 一个好的操作系统可以收集对各种资源使用的统计信息,监控诸如响应时间之类的性能参数。在任何系统中,这个信息对于预测功能提升的需求以及调整系统以提高性能都是很有用的。对多用户系统,这个信息还可用于记账。

5.1.2 作为资源管理器的操作系统

一台计算机就是一组资源,这些资源用于对数据的移动、存储和处理,以及对这些功能的控制。而操作系统负责管理这些资源。

那么是否可以说是操作系统在控制数据的移动、存储和处理呢?从某个角度来看,答案是肯定的:通过管理计算机资源,操作系统控制计算机的基本功能,但是这个控制是通过一种不寻常的方式来实施的。通常,我们把控制机制想象成在被控制对象之外或者至少与被控制对象有一些差别和距离(例如,住宅供热系统是由自动调温器控制的,它完全不同于热产生和热发送装置)。但是,操作系统却不是这种情况,作为控制机制,它有以下两方面不同之处。

(1) 操作系统与普通的计算机软件作用相同,它也是由处理器执行的一段程序或一组程序。

(2) 操作系统经常会释放控制,而且必须依赖处理器才能恢复控制。

操作系统实际上不过是一组计算机程序,与其他计算机程序类似,它们都给处理器提供指令,主要区别在于程序的意图。操作系统控制处理器使用其他系统资源,并控制其他程序的执行时机。但是,处理器为了执行任何类似的任务,都必须停止执行操作系统程序,而去执行其他程序。因此,这时操作系统释放对处理器的控制,让处理器去做其他一些有用的工作,然后用足够长的时间恢复控制权,让处理器准备好做下一项工作。

操作系统中有一部分在内存中,其中包括内核程序(kernel,或称 nucleus)和当前正在使用的其他操作系统程序,内核程序包含操作系统中最常使用的功能。内存的其余部分包含用户程序和数据,它的分配由操作系统和处理器中的存储管理硬件联合控制。操作系统决定在程序运行过程中何时使用 I/O 设备,并控制文件的访问和使用。处理器自身也是一个资源,操作系统必须决定在运行一个特定的用户程序时,可以分配多少处理器时间,在多处理器系统中,这个决定要传到所有的处理器。

5.1.3 操作系统的易扩展性

一个重要的操作系统应该能够不断发展,其原因如下。

- **硬件升级和新型硬件的出现:** 举一个例子,早期运行 UNIX 和 Macintosh 的处理器没有“分页”(paging)的硬件机制,因此这两个操作系统也没有使用分页机制,而较新的版本经过修改,具备了分页功能。同样,图形终端和页面式终端替代了行滚动终端,这也将影响操作系统的设计,例如,此类终端允许用户通过屏幕上的窗口同时查看多个应用程序,这就要求在操作系统中提供更复杂的支持。
- **新的服务:** 为适应用户的要求或满足系统管理员的需要,需要扩展操作系统以提供新的服务。例如,如果发现用现有的工具很难保持较好的性能,操作系统就必须增加新的度量和控制工具。
- **纠正错误:** 任何一个操作系统都有错误,随着时间的推移,这些错误逐渐被发现并会引入相应的补丁程序。当然,补丁本身也可能会引入新的错误。

由于操作系统经常性变化,因此对其设计提出了一定的要求。一个非常明确的观点是,在构造系统时应该采用模块化的结构,清楚地定义模块间的接口,并备有说明文档。对于像现代操作系统这样的大型程序,简单的模块化是不够的,也就是说,不能只是简单地把程序划分为模块,还需要做更多的工作。在本章的后续部分将继续讨论这个问题。

5.2 操作系统的发展过程

102

5.2.1 手工操作(无操作系统)

从 1946 年第一台计算机诞生到 20 世纪 50 年代中期,还未出现操作系统,计算机工作采用手工操作方式。程序由人工编制二进制代码,校验之后,通过卡片或者纸带(如图 5-4 所示)输入计算机,完成之后按下控制台运行键,命令计算机开始运行。计算机运行的结果,通过卡片、纸带以及各种显示状态的氖灯显示出来。如此完全依赖人工运行方式的电子数字计算机,其运行效率是非常低的。

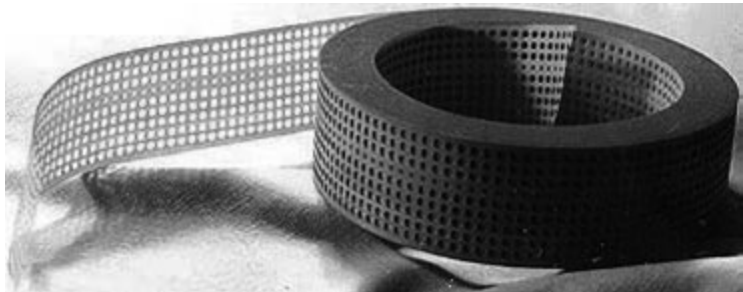


图 5-4 计算机纸带

程序员将对对应应用程序和数据的已穿孔的纸带(或卡片)装入输入机,然后启动输入机,把程序和数据输入计算机内存,接着通过控制台开关启动程序针对数据运行;计算完毕,打印机输出计算结果;用户取走结果并卸下纸带(或卡片)后,才让下一个用户上机。

手工操作方式有以下两个特点。

(1) 用户独占全机。不会出现因部分资源已被其他用户占用而等待的现象,但资源的利用率低。

(2) CPU 等待手工操作。CPU 的利用不充分。

20 世纪 50 年代后期,出现人机矛盾:手工操作的慢速度和计算机的高速度之间形成了尖锐矛盾,手工操作方式已严重损害了系统资源的利用率(使资源利用率降为百分之几,甚至更低)。唯一的解决办法:只有摆脱人的手工操作,才能实现作业的自动过渡。这样就出现了成批处理。

5.2.2 简单批处理系统

批处理系统是加载在计算机上的一个系统软件,在它的控制下,计算机能够自动地、成批地处理一个或多个用户的作业(该作业包括程序、数据和命令)。首先出现的是联机批处理系统,即作业的输入输出由 CPU 来处理。

主机与输入机之间增加一个存储设备——磁带,在运行于主机上的监督程序的自动控制下,计算机可自动完成:成批地把输入机上的用户作业读入磁带,依次把磁带上的用户作业读入主机内存并执行,然后把计算结果向输出机输出。完成了上一批作业后,监督程序又从输入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。

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接,减少了作业建立时间和手工操作时间,有效克服了人机矛盾,提高了计算机的利用率。

但是,在作业输入和结果输出时,主机的高速 CPU 仍处于空闲状态,等待慢速的输入输

出设备完成工作：主机处于“忙等”状态。为克服与缓解高速主机与慢速外设的矛盾，提高 CPU 的利用率，又引入了脱机批处理系统，即输入输出脱离主机控制。

这种方式的显著特征是：增加一台不与主机直接相连而专门用于与输入输出设备打交道的卫星机。其功能是：

- (1) 从输入机上读取用户作业并放到输入磁带上。
- (2) 从输出磁带上读取执行结果并传给输出机。

这样，主机不是直接与慢速的输入输出设备打交道，而是与速度相对较快的磁带机发生关系，有效缓解了主机与设备的矛盾。主机与卫星机可并行工作，二者分工明确，可以充分发挥主机的高速计算能力。

然而，这种方式仍存在不足之处：每次主机内存中仅存放一道作业，每当它运行期间发出输入输出(I/O)请求后，高速的 CPU 便处于等待低速的 I/O 完成状态，致使 CPU 空闲。为改善 CPU 的利用率，又引入了多道程序系统。

5.2.3 多道批处理系统

多道批处理系统中主要采用了多道程序设计技术。所谓多道程序设计技术，就是指允许多个程序同时进入内存并运行。即同时把多个程序放入内存，并允许它们交替在 CPU 中运行，它们共享系统中的各种硬软件资源。当一道程序因 I/O 请求而暂停运行时，CPU 便立即转去运行另一道程序。

单道程序的运行过程如图 5-5 所示。在 A 程序计算时，I/O 空闲，A 程序 I/O 操作时，CPU 空闲(B 程序也是同样)；必须 A 工作完成后，B 才能进入内存中开始工作，两者是串行的，全部完成共需时间 = $T_1 + T_2$ 。



图 5-5 单道程序运行过程示例

多道程序的运行过程如图 5-6 所示。将 A、B 两道程序同时存放在内存中，它们在系统的控制下，可相互穿插、交替地在 CPU 上运行：当 A 程序因请求 I/O 操作而放弃 CPU 时，B 程序就可占用 CPU 运行，这样 CPU 不再空闲，而 A 此时在进行 I/O 操作时，I/O 设备也不空闲，显然，CPU 和 I/O 设备都处于“忙”状态，大大提高了资源的利用率，从而也提高了系统的效率，A、B 全部完成所需时间 $\ll T_1 + T_2$ 。

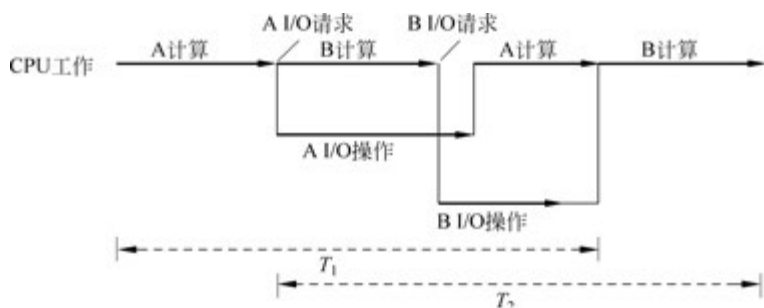


图 5-6 多道程序运行过程示例

多道程序设计技术不仅使 CPU 得到充分利用,而且改善了 I/O 设备和内存的利用率,从而提高了整个系统的资源利用率和系统吞吐量(单位时间内处理作业(程序)的个数),最终提高了整个系统的效率。

单处理器系统中多道程序运行时的特点如下。

(1) 多道: 计算机内存中同时存放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: 实际上,各道程序轮流使用 CPU,并交替运行。

多道程序系统的出现,标志着操作系统渐趋成熟的阶段,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。20 世纪 60 年代中期,在前述的批处理系统中,引入多道程序设计技术后形成多道批处理系统(批处理系统)。

多道批处理系统有以下两个特点。

(1) 多道: 系统内可同时容纳多个作业。这些作业放在外存中,组成一个后备队列,系统按一定的调度原则每次从后备作业队列中选取一个或多个作业进入内存运行,运行作业结束、退出运行和后备作业进入运行均由系统自动实现,从而在系统中形成一个自动转接的、连续的作业流。

(2) 成批: 在系统运行过程中,不允许用户与其作业发生交互作用,即作业一旦进入系统,用户就不能直接干预其作业的运行。

批处理系统的追求目标是提高系统资源利用率和系统吞吐量,以及作业流程的自动化,然而它的一个重要缺点是不提供人机交互能力,给用户使用计算机带来不便。虽然用户独占全机资源,并且直接控制程序的运行,可以随时了解程序运行情况。但这种工作方式因独占全机造成资源效率极低。

多道批处理系统的一些具体应用实例如下。

- IBM System/360 OS

360 OS 是 IBM 为其 System/360 系列大型计算机开发的操作系统,支持多道批处理。它能够同时处理多个作业,并通过高级的作业调度和内存管理技术,提高系统资源的利用率。360 OS 包括三个主要版本: PCP(Primary Control Program)、MFT(Multi-Programming with a Fixed number of Tasks)和 MVT(Multi-Programming with a Variable number of Tasks)。

- Control Data Corporation (CDC) SCOPE

SCOPE 是 CDC 为其 6600 系列超级计算机开发的多道批处理操作系统。SCOPE 支持多任务处理和高级作业调度,通过优化资源分配和作业管理,提高了系统的整体性能。

这些多道批处理系统通过高效的作业调度、资源管理和内存管理技术,显著提高了计算机系统的资源利用率和整体性能,推动了大型计算机系统的发展和应用。

5.2.4 分时系统

由于 CPU 速度不断提高,一台计算机可同时连接多个用户终端,而每个用户可在自己的终端上联机使用计算机,好像自己独占机器一样。分时技术正是得以解决上述问题的重要手段,即把处理器的运行时间分成很短的时间片,按时间片为单位轮流把处理器分配给各联机作业使用。

若某个作业在分配给它的时间片内不能完成其计算,则该作业暂时中断,把处理器让给另一作业使用,等待下一轮时再继续运行。由于计算机速度很快,作业运行轮转得很快,给每个

用户的印象是,好像它独占了一台计算机。而每个用户可以通过自己的终端向系统发出各种操作控制命令,在充分的人机交互情况下,完成作业的运行。

具有上述特征的计算机系统称为分时系统,它允许多个用户同时联机使用计算机,如图 5-7 所示。

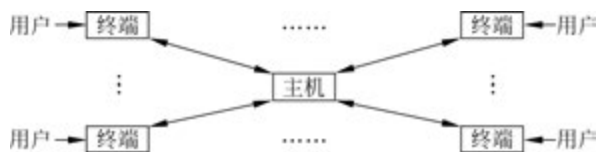


图 5-7 分时系统

分时系统的特点如下。

(1) 多路性。若干用户同时使用一台计算机。微观上看是各用户轮流使用计算机;宏观上看是各用户并行工作。

(2) 交互性。用户可根据系统对请求的响应结果,进一步向系统提出新的请求。这种能使用户与系统进行人机对话的工作方式,明显地有别于批处理系统,因而,分时系统又被称为交互式系统。

(3) 独立性。用户之间可以相互独立操作,互不干扰。系统保证各用户程序运行的完整性,不会发生相互混淆或破坏现象。

(4) 及时性。系统可对用户的输入及时作出响应。分时系统性能的主要指标之一是响应时间,它是指从终端发出命令到系统予以应答所需的时间。

分时系统的主要目标是能够对用户命令及时进行响应,即不至于用户等待每一个命令的处理时间过长。分时系统可以同时接纳数十个甚至上百个用户,由于内存空间有限,往往采用对换(又称交换)方式的存储方法。即将未“轮到”的作业放入磁盘,一旦“轮到”,再将其调入内存;而时间片用完后,又将作业存回磁盘(俗称“滚进”“滚出”法),使同一存储区域轮流为多个用户服务。多用户分时系统是当今计算机操作系统中最普遍使用的一类操作系统。

分时系统的一些具体应用实例如下。

- CTSS

CTSS(Compatible Time-Sharing System)是由麻省理工学院(MIT)的计算机科学与人工智能实验室(CSAIL)于 1961 年开发的世界上第一个分时系统。它运行在 IBM 7094 计算机上,允许多个用户通过终端同时访问计算资源,并进行程序开发、调试和运行。

- UNIX

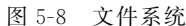
UNIX 最早由 AT&T 贝尔实验室的 Ken Thompson 和 Dennis Ritchie 在 20 世纪 60 年代末和 70 年代初开发。UNIX 采用了分时技术,使多个用户可以通过终端同时使用系统资源。UNIX 的设计简洁且功能强大,成为许多现代操作系统(如 Linux 和 BSD)的基础。

这些分时系统通过引入时间片轮转和多任务处理技术,使多个用户可以同时高效地共享计算资源,推动了计算机科学的发展,并为现代多任务操作系统奠定了基础。

5.2.5 文件系统

分时操作系统中的文件系统扮演着管理和组织存储设备数据的关键角色。它通过执行文件的创建、删除、读取、写入,以及管理文件命名和路径等功能,有效地支持用户和应用程序对数据的存取和管理。文件系统通常采用层次化结构,使用目录和子目录组织文件,并提供标准接口供文件操作使用。早期的分时操作系统如 CTSS 和 Multics 通过引入分层目录结构和权

当用户或应用程序请求操作文件时,文件系统首先通过目录结构查找文件的路径和元数据(如文件名、大小、权限等)。接着,它负责管理存储空间的分配和回收,确保文件数据的正确存储和读取。在进行读写操作时,文件系统利用缓存和预读技术优化性能,并通过权限控制确保只有授权用户可以访问文件。此外,文件系统在系统崩溃或硬件故障时保护数据的完整性,确保数据不会丢失。通过这些步骤,文件系统高效地管理和组织存储设备上的数据,实现了数据的可靠存取和管理功能。具体过程如图 5-8 所示。



这是微软为 Windows NT 系列操作系统开发的文件系统,提供更高的安全性、压缩、加密

和大文件支持。NTFS 支持文件权限、磁盘配额和数据恢复功能,是现代 Windows 操作系统(如 Windows 10、Windows Server)的默认文件系统。

- ext (Extended File System)

ext2/ext3/ext4 是 Linux 操作系统中广泛使用的文件系统。ext2 是最早的扩展文件系统,ext3 引入了日志功能以提高数据可靠性,而 ext4 则进一步增强了性能和扩展性,支持更大的文件和分区,并优化了数据存取速度。

这些文件系统在设计 and 功能上各有特点,适用于不同的操作系统和应用场景,提供从个人计算设备到企业级存储解决方案。

5.3 操作系统的基本特性

5.3.1 并行性与并发性

并行性和并发性(Concurrency)是相似又有区别的两个概念,并行性是指两个或多个事件在同一时刻发生;而并发性是指两个或多个事件在同一时间间隔内发生。在多道程序环境下,并发性是指在一段时间内,宏观上有多个程序在同时运行,但在单处理器系统中,每一时刻却仅能有一道程序执行,故微观上这些程序只能是分时地交替执行。倘若在计算机系统中有多个处理器,则这些可以并发执行的程序便可被分配到多个处理器上,实现并行执行,即利用每个处理器来处理一个可并发执行的程序,这样,多个程序便可同时执行。

5.3.2 共享

在操作系统环境下,所谓共享(Sharing)是指系统中的资源可供内存中多个并发执行的进程(线程)共同使用。由于资源属性的不同,进程对资源共享的方式也不同,目前主要有互斥共享和同时共享两种资源共享方式。

5.3.3 虚拟技术

操作系统中的所谓“虚拟”,是指通过某种技术把一个物理实体变为若干逻辑上的对应物。物理实体(前者)是实的,即实际存在的;而后者是虚的,是用户感觉上的东西。相应地,用于实现虚拟的技术,称为虚拟技术。在操作系统中利用了多种虚拟技术,分别用来实现虚拟处理器、虚拟内存、虚拟外部设备和虚拟信道等。

在虚拟处理器技术中,是通过多道程序设计技术,让多道程序并发执行的方法,来分时使用一台处理器的。此时,虽然只有一台处理器,但它能同时为多个用户服务,使每个终端用户都认为是有一个 CPU 在专门为他服务。利用多道程序设计技术,把一台物理上的 CPU 虚拟为多台逻辑上的 CPU,也称为虚拟处理器,我们把用户所感觉到的 CPU 称为虚拟处理器。

类似地,可以通过虚拟存储器技术,将一台计算机的物理存储器变为虚拟存储器,以便从逻辑上来扩充存储器的容量。此时,虽然物理内存的容量可能不大(如 64MB),但它可以运行比它大得多的用户程序(如 256MB)。这使用户所感觉到的内存容量比实际内存容量大得多,认为该计算机的内存至少也有 256MB。当然这时用户所感觉到的内存容量是虚的。我们把用户所感觉到的存储器称为虚拟存储器。

我们还可以通过虚拟设备技术,将一台物理 I/O 设备虚拟为多台逻辑上的 I/O 设备,并允许每个用户占用一台逻辑上的 I/O 设备,这样便可使原来仅允许在一段时间内由一个用户

访问的设备(即临界资源),变为在一段时间内允许多个用户同时访问的共享设备。例如,原来的打印机属于临界资源,而通过虚拟设备技术,可以把它变为多台逻辑上的打印机,供多个用户“同时”打印。此外,也可以把一条物理信道虚拟为多条逻辑信道(虚信道)。在操作系统中,虚拟的实现主要是通过分时使用的方法。显然,如果 n 是某物理设备所对应的虚拟的逻辑设备数,则虚拟设备的平均速度就是物理设备速度的 $1/n$ 。

5.3.4 异步性

在多道程序环境下,允许多个进程并发执行,但只有进程在获得所需的资源后方能执行。在单处理器环境下,由于系统中只有一个处理器,因而每次只允许一个进程执行,其余进程只能等待。当正在执行的进程提出某种其他资源要求时,如打印请求,而此时打印机正在为其他某进程打印,由于打印机属于临界资源,因此正在执行的进程必须等待,且放弃处理器,直到打印机空闲,并再次把处理器分配给该进程时,该进程方能继续执行。可见,由于资源等因素的限制,使进程的执行通常都不是“一气呵成”,而是以“停停走走”的方式运行的。

内存中的每个进程在何时能获得处理器运行,何时又因提出某种资源请求而暂停,以及进程以怎样的速度向前推进,每道程序总共需多少时间才能完成,等等,都是不可预知的。由于各用户程序性能的不同,例如,有的侧重于计算而较少需要 I/O;而又有有的程序其计算少而 I/O 多,这样,很可能是先进入内存的作业后完成;而后进入内存的作业先完成。或者说,进程是以人们不可预知的速度向前推进,此即进程的异步性。尽管如此,但只要运行环境相同,作业经多次运行,都会获得完全相同的结果。因此,异步运行方式是允许的,是操作系统的一个重要特征。

5.3.5 持久性与可靠性

操作系统的持久性主要通过文件系统和存储管理来实现。文件系统负责将数据组织和存储在持久性存储设备(如硬盘)上,并提供读取和写入数据的接口。操作系统使用缓存技术来优化文件访问速度,并通过写入确认机制确保数据成功写入存储设备,从而防止数据丢失或损坏。

操作系统通过多种手段保证系统的稳定性和可用性。这包括错误检测和处理程序,例如通过硬件和软件的自检功能来发现和纠正故障。操作系统还实施了容错设计,如数据镜像和磁盘冗余阵列(RAID)技术,用于在硬件故障时恢复数据。此外,操作系统提供备份和恢复策略,定期备份数据并允许用户从备份中恢复,以应对意外数据丢失或系统崩溃的情况。这些措施共同确保了操作系统在面对各种异常情况时能够保持高可靠性,为用户和应用程序提供稳定、安全的计算环境。

5.3.6 实时性

实时操作系统通过高效的任务调度算法和优先级管理机制,确保关键任务在规定的时间内完成。根据时间要求的严格程度,实时操作系统分为硬实时系统和软实时系统。硬实时系统要求任务必须在指定时间内完成,如工业控制和航空航天领域;软实时系统则较为灵活,能处理大部分任务的时间限制。

设计上,实时操作系统注重最小化响应时间,通过快速中断处理、优先级继承和阻塞等机制,确保任务按时完成。这些系统不仅具备高度的可预测性和可靠性,还需避免过度延迟或资源争用,确保能够准确响应和处理实时事件。

5.4 操作系统的主要功能

操作系统管理和控制计算机系统中的所有硬件、软件资源,合理地组织计算机工作流程,并为用户提供一个良好的工作环境和友好的接口。计算机系统的主要硬件资源有处理器、存储器、外部设备,软件资源以文件形式存在外存储器上。因此,从资源管理和用户接口的观点上看,操作系统具有处理器管理、存储管理、设备管理、文件管理和提供用户接口的功能。

5.4.1 进程管理

操作系统首先通过创建进程来分配必要的资源,如内存和 CPU 时间,并初始化进程控制块(Process Control Block,PCB),记录每个进程的状态和相关信息,以便全面掌握所有进程的情况,便于进行统一管理。其次,操作系统通过各种调度算法(如先来先服务、最短作业优先和时间片轮转)来分配 CPU 时间给各个进程。调度算法的选择和执行直接关系到系统资源的高效利用和整体性能。操作系统需要根据当前系统负载和各进程的需求,动态调整调度策略,以达到最佳的系统性能。

此外,操作系统还通过进程同步和通信机制(如信号量和消息传递)来协调多个进程的操作,避免竞争条件和死锁问题。这些机制确保了进程间能够安全、可靠地共享资源和数据,从而实现协同工作。操作系统还负责维护和转换进程状态。一个进程在执行过程中可能处于不同的状态,如就绪、运行、等待和终止。操作系统通过跟踪每个进程的状态,并在适当时刻进行状态转换,确保进程能够顺利执行。如图 5-9 所示为当一个进程完成 I/O 操作后,操作系统会将其从等待状态转换为就绪状态,以便重新参与调度。最后,操作系统提供进程优先级机制,通过设定不同进程的优先级,确保关键任务能够及时执行,从而实现系统的实时性和响应能力。这对于某些对时间敏感的应用场景尤为重要,如实时系统和多媒体应用。

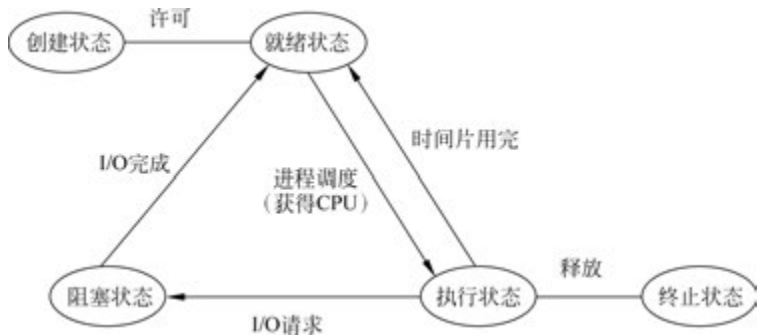


图 5-9 操作系统进程管理流程

5.4.2 处理器管理

计算机系统中处理器(CPU)是最宝贵的系统资源,处理器管理的目的是要合理地安排时间和次序,保证多个作业能够顺利完成并且尽量提高 CPU 的效率,使用户等待的时间最少。操作系统对处理器管理策略不同,提供作业处理方式也就不同,如批处理方式、分时处理方式和实时处理方式。

在大型操作系统中,一般存在多个处理器,可同时管理多个作业。此时,操作系统更要考虑怎样选出其中一个作业放入主存中运行,怎样为这个作业分配处理器时间等。处理器管理

模块要对系统中各个处理器的状态进行登记,还要登记各个作业对处理器的要求,管理模块还要用一个优化算法来实现最佳调度规则。把所有的处理器分配给各个用户作业使用的最终目的是提高处理器的利用率。

5.4.3 存储管理

存储管理的主要工作是对内存存储器(主存储器)进行合理分配、有效保护和扩充。存储管理模块对内存的管理分为三个步骤。如图 5-10 所示,首先,为各个用户分配内存空间;其次是保护已占内存空间的作业不被破坏;最后,是结合硬件实现信息的物理地址到逻辑地址的变换。这样,用户在操作中不必担心信息究竟在哪个具体空间——即实际物理地址,就可以操作,这样就方便了用户对计算机的使用和操作。内存管理模块对内存的管理是使用一种优化算法对内存管理进行优化处理的,以提高内存的利用率。

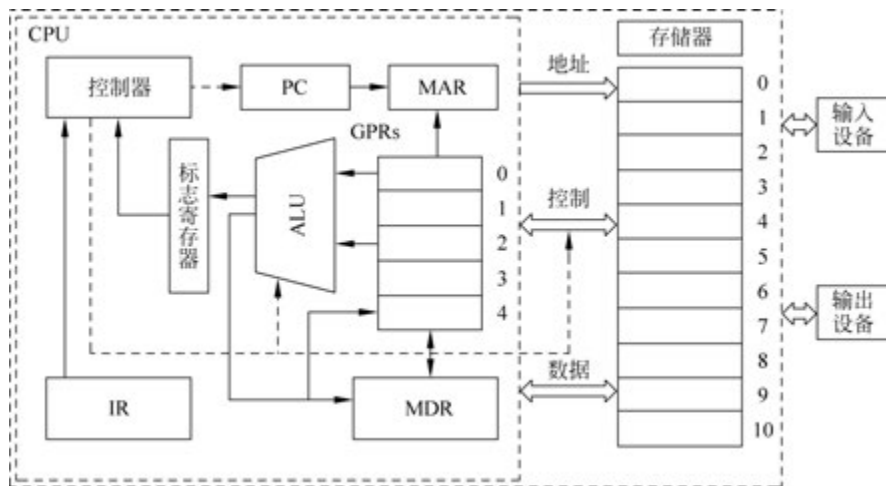


图 5-10 操作系统存储管理流程

内存保护是操作系统的关键功能之一,通过防止进程之间的相互干扰,确保系统的稳定性和安全性。每个进程只能访问其合法的内存区域,操作系统利用硬件支持的内存管理单元(Memory Management Unit,MMU)实现这一功能。此外,内存共享机制允许多个进程共享某些内存区域,如代码段或数据段,提高资源利用效率。

当进程结束或不再需要某些内存时,操作系统会回收这些内存,并将其重新分配给其他进程。垃圾回收机制和内存紧缩技术帮助操作系统有效管理和回收内存,防止内存泄漏和碎片化,从而保持系统的整体性能和稳定性。

5.4.4 设备管理

由于计算机的不断发展,其应用领域越来越广泛,应用技术越来越高,应用方法越来越简便,与用户的界面越来越友好,随之外部设备的种类也日益增多,功能不断提高,档次日渐升级。因此,操作系统的设备管理模块的功能也必须跟上外部设备的发展而不断发展、不断升级以适应外部设备日益发展的需要。

当用户程序要使用外部设备时,设备管理控制(或调用)驱动程序使外部设备工作,并随时对该设备进行监控,处理外部设备的中断请求等。设备管理模块的任务是当用户要求使用某种设备时,应马上分配给用户所要求使用的设备,并按用户要求驱动外部设备以供用户应用。对外部设备的中断请求,设备管理模块要及时响应并处理中断。

5.4.5 信息管理(文件系统管理)

以上三种管理都是针对计算机硬件资源的管理,文件系统的管理是对软件资源的管理,即对信息存储的管理。为了管理庞大的系统软件资源及用户提供的程序和数据,操作系统将它们组织成文件的形式,操作系统对软件的管理实际上就是对文件系统的管理。

操作系统对文件的管理主要是通过文件管理模块来实现的。文件管理模块管理的范围包括文件目录、文件组织、文件操作和文件保护。

5.4.6 用户接口

计算机用户与计算机的交流是通过操作系统的用户接口(也称用户界面)完成的。操作系统为用户提供的接口有两种,一是操作界面;二是操作系统的功能服务界面。用户接口的形式一般包括面向终端用户的命令行接口、图形界面接口以及面向程序员开发的系统调用接口等。

5.5 常见操作系统及分类

5.5.1 Windows 操作系统

Microsoft Windows 是微软公司制作和研发的一套桌面操作系统,它问世于 1985 年,起初仅仅是 MS-DOS 模拟环境,由于微软不断地更新升级,后续的系统版本不但易用,也慢慢地成为家家户户最喜爱的操作系统。

Windows 采用了图形用户界面(GUI),比起从前的 DOS 需要键入指令使用的方式更为人性化。随着计算机硬件和软件的不断升级,微软的 Windows 也在不断升级,从架构的 16 位、32 位再到 64 位,系统版本从最初的 Windows 1.0 到大家熟知的 Windows 95、Windows 98、Windows 2000、Windows XP、Windows Vista、Windows 7、Windows 8、Windows 10、Windows 11 和 Server 服务器企业级操作系统,不断持续更新,如图 5-11 所示,微软一直在致力于 Windows 操作系统的开发和完善。



图 5-11 Windows 操作系统

5.5.2 Linux 操作系统

Linux(如图 5-12 所示)是一个最早期的计算机系统也是一个廉价的系统,不需要任何的

认证就可以使用,而且就在这个信息时代飞速发展时,Linux 越来越多地受到大家的认同,并且它与现在非常流行的嵌入式系统联系紧密,受到了越来越多人的重视。



图 5-12 Linux 操作系统

Linux 是一类 UNIX 计算机操作系统的统称。Linux 操作系统的内核的名字也是“Linux”。Linux 操作系统也是自由软件和开放源代码发展中最著名的例子。严格来讲,Linux 这个词本身只表示 Linux 内核,但实际上人们已经习惯了用 Linux 来形容整个基于 Linux 内核,并且使用 GNU 工程的各种工具和数据库的操作系统。

Linux 是最受欢迎的自由计算机操作系统内核。它是一个用 C 语言和汇编语言写成,符合 POSIX 标准的类 UNIX 操作系统。Linux 最早是由林纳斯·托瓦兹(Linus B. Torvalds)为尝试在 Intel x86 架构上提供自由免费开源的类 UNIX 操作系统而开发的。该计划开始于 1991 年,林纳斯·托瓦兹当时在 Usenet 新闻组 comp.os.minix 登载帖子,这份著名的帖子标志着 Linux 计划的正式开始。

简单地说,Linux 是一套免费使用和自由传播的类 UNIX 操作系统,它主要用于基于 x86 系列 CPU 的计算机上。这个系统是由世界各地成千上万的程序员共同设计和实现的。其目的是建立不受任何商品化软件的版权制约的、全世界人人都能自由使用的 UNIX 兼容产品。

Linux 的基本思想有两点:第一,一切都是文件;第二,每个软件都有确定的用途。其中第一条详细来讲就是系统中的所有都归结为一个文件,包括命令、硬件和软件设备、操作系统、进程等对于操作系统内核而言,都被视为拥有各自特性或类型的文件。至于说 Linux 是基于 UNIX 的,很大程度上也是因为这两者的基本思想十分相近。

Linux 的突出特点如下。

(1) 完全免费。Linux 是一款免费的操作系统,用户可以通过网络或其他途径免费获得,并可以任意修改其源代码。这是其他操作系统做不到的。正是由于这一点,来自全世界的无数程序员参与了 Linux 的修改、编写工作,程序员可以根据自己的兴趣和灵感对其进行改变,这让 Linux 吸收了无数程序员的精华,不断壮大。

(2) 完全兼容 POSIX1.0 标准。这使得可以在 Linux 下通过相应的模拟器运行常见的 DOS、Windows 的程序。这为用户从 Windows 转到 Linux 奠定了基础。许多用户在考虑使用 Linux 时,就想到以前在 Windows 下常见的程序是否能正常运行,这一点就消除了他们的疑虑。

(3) 多用户、多任务。Linux 支持多用户,各个用户对于自己的文件设备有自己特殊的权利,保证了各用户之间互不影响。多任务则是现在计算机最主要的一个特点,Linux 可以使多个程序同时并独立地运行。

(4) 良好的界面。Linux 同时具有字符界面和图形界面。在字符界面用户可以通过键盘输入相应的指令来进行操作。它同时也提供了类似 Windows 图形界面的 X-Window 系统,用

户可以使用鼠标对其进行操作。在 X-Window 环境中就和在 Windows 中相似,可以说是一个 Linux 版的 Windows。

(5) 支持多种平台。Linux 可以运行在多种硬件平台上,如具有 x86、680x0、SPARC、Alpha 等处理器的平台。此外 Linux 还是一种嵌入式操作系统,可以运行在掌上电脑、机顶盒或游戏机上。2001 年 1 月份发布的 Linux 2.4 版内核已经能够完全支持 Intel 64 位芯片架构。同时 Linux 也支持多处理器技术。多个处理器同时工作,使系统性能大大提高。

5.6 本章小结

本章从用户和系统资源管理两个不同的角度讲解了操作系统的实现目标、功能和作用。回顾操作系统的发展历史,从手工操作,到批处理系统,再到现在的多用户分时系统以及专用的实时系统,操作系统随着计算机技术的整体发展变得更加高效、便利。在功能方面,操作系统的核心包括进程管理、内存管理、文件系统管理、设备管理、安全与保护、用户接口。这些功能共同确保系统资源的高效利用、数据的持久性和安全性,并为用户和应用程序提供稳定的运行环境。本章最后介绍了 Windows 和 Linux 两种最为常见和典型的操作系统,两种系统平台孰优孰劣之争从 20 世纪延续到现在,相信读者经过了解和学习后,也有自己的判断,通过进一步尝试也许就能确定和选择自己心中最合适的操作系统来使用了。

习 题 5

一、选择题

1. 一个完整的计算机系统是由_____组成的。
A. 硬件 B. 软件 C. 硬件和软件 D. 用户程序
2. 操作系统的基本职能是_____。
A. 控制和管理系统内各种资源,有效地组织多道程序的运行
B. 提供用户界面,方便用户使用
C. 提供方便的可视化编辑程序
D. 提供功能强大的网络管理工具
3. 以下著名的操作系统中,属于多用户、分时系统的是_____。
A. DOS 系统 B. Windows NT 系统
C. UNIX 系统 D. OS/2 系统
4. 为了使系统中所有的用户都能得到及时的响应,该操作系统应该是_____。
A. 多道批处理系统 B. 分时系统
C. 实时系统 D. 网络系统
5. 关于操作系统的叙述,_____是不正确的。
A. “管理资源的程序” B. “管理用户程序执行的程序”
C. “能使系统资源提高效率的程序” D. “能方便用户编程的程序”
6. 操作系统的发展过程是_____。
A. 设备驱动程序组成的原始操作系统、管理程序、操作系统
B. 原始操作系统、操作系统、管理程序
C. 管理程序、原始操作系统、操作系统

- D. 管理程序、操作系统、原始操作系统
7. 用户程序中的输入输出操作实际上是由_____完成的。
A. 程序设计语言 B. 编译系统 C. 操作系统 D. 标准库程序
8. _____不是基本的操作系统。
A. 批处理操作系统 B. 分时操作系统
C. 实时操作系统 D. 网络操作系统
9. 以下_____是批处理操作系统。
A. CTSS B. 360 OS C. FreeRTOS D. UNIX
10. _____不是分时系统的基本特征。
A. 同时性 B. 独立性 C. 实时性 D. 交互性
11. 操作系统的基本职能是_____。
A. 控制和管理系统内各种资源,有效地组织多道程序的运行
B. 提供用户界面,方便用户使用
C. 提供方便的可视化编辑程序
D. 提供功能强大的网络管理工具
12. 操作系统是一种_____。
A. 系统软件 B. 系统硬件 C. 应用软件 D. 支援软件

二、简答题

1. 操作系统的主要功能包括哪些?
2. 操作系统的主要设计目标有哪些?
3. 你更喜欢用哪种操作系统?谈谈你的理由。