

## 操作系统概述

操作系统 (operating system, OS) 的出现、应用和发展是计算机软件所取得的一个重大进展,它经历了从无到有、从小到大、从简单到复杂、从原始到先进的发展历程。它为人们使用计算机提供了方便的接口,使计算机系统的各种资源得到了充分利用。它的发展与计算机用户的需求和计算机制造技术的进步以及计算机理论科学的发展密切相关。它获得了多种美誉,如“计算机系统的管家”、“人机间的桥梁”、“软件的基石”等。

### 【学习目标】

- ① 操作系统的定义。
- ② 操作系统的作用。
- ③ 操作系统的种类。
- ④ 操作系统的特征。
- ⑤ 操作系统的功能。
- ⑥ Windows 和 Linux 操作系统简介。

## 1.1 操作系统的定义和作用

操作系统利用一个或多个处理机的硬件资源,为系统用户提供一组服务,它还代替用户来管理辅助存储器 and 输入/输出(input/output, I/O)设备。

### 1.1.1 操作系统的定义

计算机软件分为系统软件和应用软件两大类。系统软件用于管理计算机本身和应用程序,应用软件是为满足用户特定需求而设计的软件。操作系统和系统工具软件构成了系统软件。

操作系统在计算机中的位置如图 1-1 所示。它运行在裸机之上,允许用户运行其他程序,如 Web 浏览器、电子书阅读器和视频播放器等。

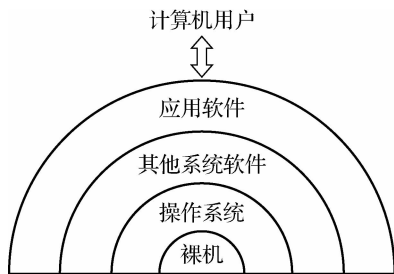


图 1-1 操作系统在计算机中的位置

操作系统是一种运行在内核态的软件。操作系统执行两个基本独立的任务,即为应用程序提供一个资源集的清晰抽象,并管理这些硬件资源。无论从哪个角度看待操作系统,它都完成以下 3 个目标。

- (1)方便。操作系统使计算机易于使用。
- (2)有效。操作系统允许以更有效的方式使用计算机系统资源,包括硬件资源和软件资源。
- (3)扩展的能力。在构造操作系统时,应该允许在不妨碍服务的前提下有效地开发、测试和引进新的系统功能。

下面从两方面介绍操作系统的这 3 个目标,以便更好地理解操作系统。

(1)作为用户与计算机接口的操作系统。为用户提供各种应用的硬件和软件可以看做是一种层次结构,如图 1-2 所示。应用程序的用户(最终用户)通常并不关心计算机的硬件细节,因此最终用户把计算机系统看做是一组应用程序。一个应用程序可以用一种程序设计语言描述,它是由程序员开发的。如果需要用一组完全负责硬件的机器指令开发应用程序,将会是一件非常复杂和困难的任务。为简化这个任务,需要提供一些系统程序,其中一部分称为实用工具,它们实现了在创建程序、管理文件和控制 I/O 设备中经常使用的功能。程序员在开发应用程序时将使用这些软件;应用程序在运行时,将调用这些实用工具以实现特定的功能。最重要的系统程序是操作系统,操作系统为程序员屏蔽了硬件细节,并为程序员使用系统提供了方便的接口。它可以作为中介,使程序员和应用程序更容易地访问和使



用这些功能与服务。

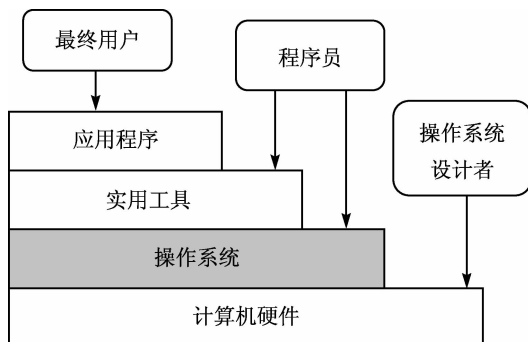


图 1-2 计算机系统的层次结构

(2)作为资源管理者的操作系统。把操作系统看做是向应用程序提供基本接口,这是一种自顶向下的观点。按照另一种自底向上的观点,操作系统则用来管理一个复杂系统的各个部分。现代操作系统包含处理机、存储器、时钟、磁盘、鼠标、网络接口、打印机以及许多其他设备。从这个角度看,操作系统的任务是在相互竞争的程序之间有序地控制对处理机、存储器以及其他 I/O 接口设备的分配。

当一个计算机有多个用户时,管理和保护存储器、I/O 设备以及其他资源的需求变得强烈起来,因为用户间可能会互相干扰。另外,用户通常不仅要共享硬件,还要共享信息(文件、数据库等)。简而言之,操作系统的这一观点认为,操作系统的主要任务是记录哪个程序在使用什么资源,对资源请求进行分配,评估使用代价,并且为不同的程序和用户调解互相冲突的资源请求。

综上所述,操作系统是计算机系统上的系统软件,管理和控制计算机系统上的硬件和软件资源,合理地组织计算机的工作流程,以便有效利用这些资源为用户提供一个功能强、使用方便的工作环境,从而在计算机与用户之间起到接口的作用。

### 1.1.2 操作系统的作用

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。操作系统是一个庞大的管理控制程序,管理着系统中的各种资源并为用户提供接口来操作计算机。

#### 1. 管理系统中的各种资源

在计算机系统中,所有硬件部件(如 CPU、存储器、I/O 设备等)称为硬件资源,而程序和数据等信息称为软件资源。操作系统对每一种资源的管理都必须进行以下几项工作。

(1)监视资源。监视资源包括需要知道该资源有多少,资源的状态如何,它们都在哪里,谁在使用,可供分配的又有多少,资源的使用历史等。

(2)决定分配资源策略。决定分配资源策略包括选择哪种资源分配策略,决定谁有权限可以获得这种资源,何时可以获得,可以获得多少,如何退回资源等。

(3)分配资源。分配资源是指按照已决定的资源分配策略,对符合条件的申请者分配某种资源,并进行相应的管理事务处理。

(4)回收资源。回收资源是指在使用者放弃某种资源之后,对该种资源进行善后处理,



如果是可重复使用的资源,则进行回收、整理,以备再次使用。

## 2. 为用户提供友好的界面

操作系统必须为最终用户和系统用户的各种工作提供友好的界面,以方便用户的工作。典型的操作系统界面有以下两类。

(1)命令行界面。如 UNIX 和 MS-DOS 操作系统,在这种操作系统环境下,用户通过键盘输入完成某种工作的命令,操作系统分析命令的正确性,然后执行命令并返回执行结果。

(2)图形化界面。如 Windows 操作系统,在其提供的图形化界面环境下,用户可通过单击某个按钮或选择某个菜单来完成相应的命令。

## 1.2 操作系统的发展历史

操作系统与计算机系统的发展息息相关。操作系统从最早的批处理阶段开始,经历了分时机制、多处理机等阶段,后来又添加了多处理机协调功能,甚至是分布式系统的协调功能。

### 1. 无操作系统的计算机系统

第一台真正的数字计算机是英国数学家 Charles Babbage 设计的。尽管 Babbage 花费了几乎一生的时间和财产,试图建造他的“分析机”,但是他始终未能让机器正常运转,因为它是一台纯数字计算机,他所在时代的技术不能生产出他所需要的高精度机械设备。毫无疑问,这台“分析机”没有操作系统。

从 Babbage 失败之后一直到第二次世界大战期间,数字计算机的建造几乎没有什么进展,这次战争引发了有关计算机研究的爆炸性展开。衣阿华州立大学的 John Atanasoff 教授和他的学生 Clifford Berry 建造了第一台可工作的数字计算机,该机器使用了大约 300 个真空管。同时,其他一些大学院校也建造了各自的数字计算机。这些机器非常庞大,往往使用数万个真空管,占据几个房间,但其运算速度却不如现在最便宜的个人计算机。

在计算机出现的早期,每台机器都有一个小组专门来设计、制造、编程、操作和维护。编程全部采用机器语言,通过在一些插板上的硬连线来控制其基本的功能,这时没有编程语言,操作系统更是闻所未闻。

到 20 世纪 50 年代早期,出现了穿孔卡片,这时就可以不用插板,而是将程序写在卡片上,然后读入计算机。

### 2. 单道批处理系统

20 世纪 50 年代发明的晶体管极大地改变了计算机的状况。这时的计算机已经很可靠,厂商可以成批地生产计算机并卖给客户,客户可以长时间地使用它来完成一些有用的工作。至此,第一次将设计人员、生产人员、操作员、程序员和维护人员分开。

这个时期的计算机被安置在空调房间里,并配有专人操作。由于其价格昂贵,仅有少数大公司、主要的政府部门和大学才买得起。运行一个作业(一个或一组程序)时,程序员首先将程序写在纸上(使用 FORTRAN 或汇编语言),然后用穿孔机制成卡片,最后将这些卡片交给操作员。



操作员从卡片上读入任务,计算机运行完成当前任务后,其计算结果从打印机上输出,操作员从打印机上取得运算结果并送到输出室,程序员就可以从输出室得到运算结果,然后,操作员再从卡片上读入另一个任务。当操作员在机房里走来走去时,许多机时被浪费掉了。

由于当时计算机非常昂贵,很自然地,人们开始想办法减少机时的浪费,批处理系统由此产生。其思想是:在作业输入室收集到较多的作业后,使用一台相对廉价的计算机将它们读到磁带上,另外用较昂贵的计算机来完成真正的计算。

在收集到一批作业之后,输入磁带被送到机房里并装到磁带机上。操作员随后装入一个特殊的程序(现代操作系统的前身),它从磁带上将第一个作业读入并运行,其输出写到第一盘输出磁带上,而不是打印出来。每个作业结束后,此特殊的程序自动地读入下一个作业运行。当这一批作业完全结束后,操作员取下输入和输出磁带,将输入磁带换成写有下一批作业的输入磁带,然后把输出磁带拿到一台机器上进行脱机打印。

此时的批处理系统称为单道批处理系统,又称为监督程序或管理程序,用于管理应用程序的运行。

### 3. 多道批处理系统

由于在单道批处理系统中,一个作业单独进入内存并独占系统资源,直到运行结束后下一个作业才能进入内存,当作业进行 I/O 操作时,CPU 只能处于等待状态,因此,CPU 利用率较低,尤其是对于 I/O 操作时间较长的作业。为了提高 CPU 的利用率,在单道批处理系统的基础上引入了多道程序设计(multiprogramming)技术,这就形成了多道批处理系统,即在内存中可同时存在若干道作业,作业执行的次序与进入内存的次序无严格的对应关系,因为这些作业是通过一定的作业调度算法来使用 CPU 的,一个作业在等待 I/O 处理时,CPU 调度另外一个作业运行,因此 CPU 的利用率得到了显著地提高。

多道批处理系统的优点是,由于系统资源为多个作业所共享,其工作方式是作业之间自动调度执行,并在运行过程中用户不干预自己的作业,从而大大提高了系统资源的利用率和作业吞吐量。其缺点是无交互性,用户一旦提交作业就失去了对其运行的控制能力,而且是批处理的,作业周转时间长,用户使用不方便。

### 4. 分时系统

为了克服多道批处理系统的不足,引入了分时操作系统。分时处理,又称会话型处理,是在多道程序设计基础上发展起来的一种处理方式。它把时间分隔技术应用到 CPU 的调度上,形成了一种新的操作系统。第一个真正的分时处理系统是美国麻省理工学院研制的 CTSS(compatible time sharing system),它支持位于不同终端的多个用户同时使用一台计算机,彼此独立互不干扰,从而使用户感到好像整台计算机全为他所用。

### 5. 实时系统

分时系统为交互式作业提供了快速的响应服务,但还不能满足某些对响应时间要求非常严格的任务需要。例如,炼钢/炼油控制系统、航空网络售票系统等,对任务的响应时间有更严格的要求。

为了满足对响应时间要求严格的任务的需求,人们研发出了实时系统。实时操作系统指使计算机能及时响应外部事件的请求,在严格规定的时间内完成对该事件的处理,并控制所有实时设备和实时任务协调一致地工作的操作系统。





实时操作系统要追求的目标是对外部请求在严格时间范围内作出反应,有高可靠性和完整性。其主要特点是资源的分配和调度首先要考虑实时性,然后才是效率。此外,实时操作系统应有较强的容错能力。

## 1.3 操作系统的种类

不同的应用领域和应用目的,对操作系统提出了不同的要求。根据各操作系统具备的功能、特征、规模和提供的应用环境等方面的差别,可划分为不同的类型。最基本的类型有三种:批处理操作系统、分时操作系统和实时操作系统。随着计算机体系结构的发展,又出现了许多新型操作系统。

### 1. 批处理操作系统

批处理操作系统是创建于 20 世纪 50 年代末期的第一代操作系统,主要是受到了早期系统效率低的启发。由于早期系统在装入、汇编和执行程序的每一步都需要操作人员的手工辅助,消耗了大量时间,从而导致了极其昂贵的硬件设备的低效使用。

解决方法是使用一个中央控制程序对标准的装入、汇编、执行的过程进行自动化。这个控制程序可以发现和装入所需的系统程序——汇编器、编译器、链接器或例程序等,并能处理作业到作业的自动切换。这样可以递交多个作业由系统同时处理,典型方式是使用一批穿孔卡片。这个控制程序被称为批处理操作系统。

新的硬件技术极大地促进了这些系统的发展。真空管被晶体管所取代,使得机器体积显著减小,并更可靠、更便宜。I/O 处理也得到了明显的改进。最重要的一个硬件创新是 I/O 处理机或 I/O 通道,它类似于 CPU 的处理机,但它是一种带有专用 I/O 处理的较小的指令集合。I/O 处理机把主处理机(CPU)从频繁的低层交互(与 I/O 设备的交互)中解脱出来。当收到来自 CPU 的高层 I/O 请求时,I/O 处理机有效地提高了 CPU 和许多 I/O 设备的并行操作,进而通过并行化改善了系统的整体性能。

起初,CPU 周期性地询问 I/O 处理机的状态,看它是处于忙碌态还是发生了某种错误。但很快人们就发现,如果采用一种机制,仅当 I/O 处理机需要 CPU(或者报告问题以及接收新的命令)时才通知 CPU,则系统可以更加高效地运转。解决方法是引入了中断,这是一种硬件信号,由 I/O 处理机发送给 CPU 以得到 CPU 的立即注意。后来把中断扩展为允许 CPU 可以快速响应多种不同的情况,如除数为零、无效的操作符、违反内存保护等。

为进一步流线化批处理系统的操作,人们编写编译器以产生可重定位的而不是绝对的代码。后来编写了链接器,它能组合之前单独编译后的程序(包括例程序),而无须重新编译。通过以上方法,仅当管理 I/O 批量作业、设置非标准任务和系统失败需要采取修正行为时,系统才需要用户的维护。程序员逐渐被禁止进入机房,那里成为计算机管理员(一个新的专家)的独占领域。

术语“批”至今仍在广泛使用,但其含义已经发展为表示非交互计算。一个批处理作业是指它不需要任何用户交互,把它提交给系统,意味着系统会选择在一个最合适的时间执行它。例如,系统一般会在一个比交互式作业低的优先级下执行一个批处理作业。系统也可以选择推迟它的执行,直到系统具有足够的空间/时间或直到它已经累积了一批合适的执行



作业。另外,用户可以通过高作业级别使得作业可以在以后某个时间(如晚上)开始。

## 2. 分时操作系统

分时操作系统的工作方式是:一台主机连接了若干个终端,每个终端有一个用户在使用。用户交互式地向系统提出命令请求,系统接受每个用户的命令,采用时间片轮转方式处理服务请求,并通过交互方式在终端上向用户显示结果。用户根据上步结果发出下步命令。分时操作系统将 CPU 的时间划分成若干个片段,称为时间片。操作系统以时间片为单位,轮流为每个终端用户服务。每个用户轮流使用一个时间片而使每个用户并不感到有别的用户存在。分时系统具有多路性、交互性、独占性和及时性的特征。多路性是指同时有多个用户使用一台计算机,宏观上看是多个人同时使用一个 CPU,微观上是多个人在不同时刻轮流使用 CPU;交互性是指用户根据系统响应结果进一步提出新请求(用户直接干预每一步);独占性是指用户感觉不到计算机为其他人服务,就像整个系统为他所独占;及时性是指系统对用户提出的请求及时响应,它支持位于不同终端的多个用户同时使用一台计算机,彼此独立互不干扰。

常见的通用操作系统是分时系统与批处理系统的结合。其原则是:分时优先,批处理在后。“前台”响应需频繁交互的作业,如终端的要求等;“后台”处理时间性要求不强的作业。

## 3. 实时操作系统

实时操作系统的特征是将时间作为关键参数。例如,在工业过程控制系统中,工厂中的实时计算机必须收集生产过程的实时数据并用有关数据控制机器;汽车在装配线上移动时,必须在限定的时间内进行规定的操作,如果焊接机器人焊接得太早或太迟,都会毁坏汽车。如果某个运行必须绝对地在规定的时刻(或规定的时间范围)发生,这是硬实时系统。可以在工业过程控制、民用航空、军事以及类似应用中看到很多这样的系统。这些系统必须提供绝对保证,让某个特定的动作在给定的时间内完成。

除了硬实时系统还有软实时系统,在这种系统中,偶尔违反截止时间是不希望的,但可以接受,并且不会引起任何永久性的损害。数字音频或多媒体系统就是这类系统。数字电话也是软实时系统。

## 4. 网络操作系统

网络操作系统是基于计算机网络的,是在各种计算机操作系统上按网络体系结构协议标准开发的软件,包括网络管理、通信、安全、资源共享和各种网络应用。其目标是相互通信及资源共享。在网络操作系统支持下,网络中的各台计算机能互相通信和共享资源。其主要特点是与网络的硬件相结合来完成网络的通信任务。

## 5. 分布式操作系统

大量的计算机通过网络连接在一起,可以获得极高的运算能力及广泛的数据共享,这种系统称做分布式操作系统。它在资源管理、通信控制和操作系统的结构等方面都与其他操作系统有较大的区别。由于分布式计算机系统的资源分布于系统的不同计算机上,操作系统对用户的资源需求不能像一般的操作系统那样等待有资源时直接分配,而是要在系统的各台计算机上搜索,找到所需资源后才可进行分配。对于有些资源,如具有多个副本的文件,还必须考虑一致性。所谓一致性是指若干个用户对同一个文件同时读出的数据是一致的。为了保证一致性,操作系统必须控制文件的读/写操作,使得多个用户可同时读一个文



件,而任一时刻最多只能有一个用户在修改文件。分布式操作系统的通信功能类似于网络操作系统。由于分布式操作系统不如网络操作系统分布得广,同时分布式操作系统还要支持并行处理,因此它提供的通信机制和网络操作系统提供的有所不同,它要求通信速度高。分布式操作系统的结构也不同于其他操作系统,它分布于系统的各台计算机上,能并行地处理用户的各种需求,有较强的容错能力。

## 6. 嵌入式操作系统

有一种特殊的计算机应用,它们把计算机嵌入在某种设备中,对其进行实时监控和管理,在这种应用中,所有的计算机软件都事先保存在 ROM 里,不允许用户在应用时安装软件。这样的系统称为“嵌入式系统”,运行在嵌入式系统中的操作系统,称为“嵌入式操作系统”。

早期的嵌入式系统主要应用在航空航天等领域。随着信息技术的发展,嵌入式系统逐渐大众化,进入人们的家庭和日常生活中。例如,微波炉、电冰箱、手机等都是典型的嵌入式系统的应用。

嵌入式操作系统除了具备一般操作系统最为基本的功能(如任务调度、同步机制、中断处理、文件功能等)以外,通常还要求设计得非常小巧而有效,它必须面向用户、面向产品、面向应用。嵌入式操作系统具有如下特点。

(1)实时性。嵌入式系统广泛应用于过程控制、数据采集、通信、多媒体信息处理等要求获得快速响应的设备中。因此嵌入式操作系统都应该是实时操作系统。

(2)可靠性。嵌入式系统一旦开始运行,用户就很少干预。因此负责系统管理的嵌入式操作系统必须稳定、可靠,不能有任何差错。

(3)微型性。在数字化设备中,由于其体积限制和实时性要求,不可能提供像微型机那样大的内、外存空间,因此为了能在有限空间中运行,嵌入式系统必然不能像大型机甚至微型机系统那样庞大。嵌入式系统中的软件都被固化在存储器芯片或单片机中。

(4)专业性。嵌入式的硬件平台多种多样,应用更是繁多。但是对某一个体的应用场合而言,一旦确定,就不能轻易改变。

(5)可剪裁性。大部分的数字化设备用途比较单一,在系统运行过程中其处理的计算也相对简单,且不同用户对设备功能的要求也不同,因此嵌入式系统应该能根据用户需求进行自行定制剪裁,以使用最小的软件集合实现最符合用户需求的系统。同时,嵌入式系统的可剪裁性也能将满足用户要求的系统控制在最小体积。

(6)高可移植性。嵌入式系统通常会在一系列功能相似的嵌入式硬件上运行。为了满足不同硬件或不同应用场合的特殊需求,嵌入式系统还应该能在简单修改后就可以在不同的环境中正确、有效地运行,即嵌入式系统应具有可移植性,不依赖于特定硬件。

最常见的嵌入式系统有 Palm OS、Windows CE 和 Linux。尤其是在复杂嵌入式应用领域, Linux 系统由于具有开源、外围软件多等特点而被广泛应用。

## 1.4 操作系统的基本特征

操作系统种类繁多,但却有一些共同特征,这些特征也是操作系统这一层次软件与应用





软件的区别所在。现代操作系统都具有并发性、共享性、虚拟性和异步性,其中并发性是操作系统最重要的特征,其他三个特性均基于并发性而存在。

### 1. 并发性

在说明什么是并发性之前,首先要区分两个概念:并发和并行。如果在一个时间段内发生了一个以上的事件,则称这几个事件具有并发性;而并行性指的是这多个事件在同一时刻点发生。

在不同类型的操作系统中,并发性的含义有一定的区别。在单处理机系统中,每个特定时刻只能有一个程序占用 CPU。但一个较长的时间段可以被分为多个小的时间碎片,这些碎片可以按照一定的原则分发给多个不同的程序,使得在这个时间段内有多个程序得到一定程度的执行。这些程序是具有并发性,而不具有并行性的。而在多处理机系统中,每个特定时刻有多个 CPU 可以使用,在这样的时刻点,多个可以独立运行的程序就能够并行执行。

### 2. 共享性

操作系统中的共享指的是多个并发执行的程序能够按照一定的规则共同使用操作系统所管理的软硬件资源。由于这些资源具有不同的使用要求,因此其共享方式也有所不同。操作系统所管理的软硬件资源按照使用方式可以分为同时访问方式和互斥访问方式。

(1)同时访问共享。在一些系统资源中,允许在同一段时间内,有多个程序同时访问数据。这时的“同时”,指的仍然是宏观上的。微观上,这些程序仍然是交替对资源进行访问,由操作系统把任务分给处理机。

(2)互斥访问共享。系统中的有些资源是无法共享的,如打印机等。它们的共享势必会造成任务的混乱,所以需要规定它们在同一段时间内只能由一个程序进行访问。

并发与共享密不可分,互相依存。并发是为了资源的共享,而共享实现的条件和底层机制基础又是并发。如果操作系统没有很好地处理并发和共享,势必会影响程序的正常运行,甚至无法运行。

### 3. 虚拟性

这里所谓的虚拟性并不是虚拟机,而是将计算机体系结构中的各种物理设备映射为多个逻辑设备。这种映射通常是利用分时共享的方式实现的,被映射的物理设备有多种,如内存、外设、CPU 等。每个不同的设备,由于其工作模式不同,所使用的映射方法也不尽相同。

使用虚拟存储器技术,可以将较小的物理内存“扩充”为较大的虚拟内存。这种方法的核心思想就是仅将程序当前运行所需的数据和代码载入内存,当这个程序的一个相对独立的功能模块运行完成或暂时无法继续进行,这个模块所对应的数据和代码将被暂存到外存的指定区域,其释放的内存空间将被重新分配给本程序的其他功能模块或其他应用程序的功能模块使用。使用这种方法可以确保对空间有较大要求的应用程序也能正常运行于小内存的机器上。

类似地,虚拟处理器技术也利用了分时方法,且使用多道程序设计技术保证多道程序并发执行。在一段时间内,CPU 将分时间片处理多个程序请求,但提出这些请求的各个用户并不会感觉到有其他人和自己共用 CPU 时间,而是感觉自己独占了资源。在硬件能力快速发展的今天,这种方法能够最大限度地发挥联网机器的效用,提高处理效率。

对设备的虚拟则是将一个物理设备映射为多个虚拟的逻辑设备,尽管进行不同程序处



理的仍然是一台机器,但在用户感知上则是有多台机器同时进行数据处理。这种映射的实现仍然依靠分时共享方法。

可以看出,虚拟特性的实现主要依靠分时共享方法和多道程序设计技术,通过虚拟操作系统可以将一个设备映射为多个,将一个设备的能力均分到不同的逻辑设备上,以便多用户共享资源。虚拟方法不会造成多个用户长期等待其他用户的情况,同时也极大地提高了资源利用率。

#### 4. 异步性

异步性也称随机性,在多道程序设计中,由于并发性的实现机制,任务通常并不是“一气呵成”的,而是时而中断,时而运行。例如,一个进程在运行中,需要访问指定资源或者某个事件的发生才能进行下一步操作。在等待的过程中,它会被操作系统暂停,将 CPU 抽取出去执行其他的进程。于是,有新的问题留给了操作系统,如程序什么时候开始执行,什么时候暂停,以怎样的速度执行,总共要执行多长时间等,这些都需要操作系统能合理的处置。

并发性给操作系统带来了潜在的危险,它们在异步的过程中,会产生时间相关的错误,因此确保能捕捉到任何一种事件序列是操作系统的一项重要任务。

## 1.5 操作系统的主要功能

现代操作系统的主要任务就是维护一个优良的运行环境,以便多道程序能够有序、高效地获得执行,而在运行的同时,还应该尽可能地提高资源利用率和系统响应速度,并保证用户操作的方便性。因此,操作系统的基本功能包括处理机管理、存储管理、设备管理和文件管理。此外,为了给用户提供一个统一、方便、有效地使用系统能力的手段,现代操作系统还需要提供一个友好的人机接口。在互联网不断发展的今天,操作系统通常还具备基本的网络服务功能和信息安全防护等方面的支持。

### 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管理存储器资源,为多道程序运行提供有力的支撑,便于用户使



用存储资源,提高存储空间的利用率。存储管理包括以下几项。

(1)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。

(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。

(3)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,这一转换称做地址重定位。

(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 和主存外,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)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。

(2)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。

(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

前面三种管理是针对计算机硬件资源的管理,文件管理则是针对系统中的信息资源的管理,主要包括以下几项。

(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外,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。用户接口主要包括以下三种。

(1)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,其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对应的库函数,应用程序通过调用库函数来使用系统调用。

(2)命令接口。命令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输入不同的命令达到控制作业的目的。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才停止该作业的执行。

(3)图形接口。图形接口的主要构件是窗口、菜单和对话框,操作简单、直观,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。

## 1.6 主流操作系统简介

在计算机发展过程中,产生了许多不同应用类型的计算机,出现了许多具有不同功能和特点的操作系统。

### 1.6.1 Windows 操作系统

Windows 是微软公司推出的一系列操作系统,它问世于 1985 年,当时是 DOS 之下的操作环境,其后续版本逐渐发展成为个人计算机和服务器用户设计的操作系统,并最终获得了世界个人计算机操作系统软件的垄断地位。Windows 可以在几种不同类型的平台上运行,如个人计算机、服务器和嵌入式系统等,其中在个人计算机领域应用最为普遍。

Windows 具有以下优点。

(1)直观、高效的面向对象的图形用户界面,易学易用。例如,要打开一个文档,首先用鼠标选择该文档,然后从右键菜单中选择“打开”操作即可。这种操作方式模拟了现实世界的行为,易于理解、学习和使用。

(2)用户界面统一、友好。Windows 应用程序拥有相同的或相似的基本外观,包括窗口、菜单、工具条等。用户只要掌握其中一个,就不难学会其他软件,从而降低了用户培训、学习的费用。

(3)丰富的、设备无关的图形操作。Windows 的图形设备接口(GDI)提供了丰富的图形



操作函数,可以绘制出线、圆、框等几何图形,并支持各种输出设备。设备无关意味着在针式打印机上和高分辨率的显示器上都能显示出相同效果的图形。

(4)多任务。Windows 是一个多任务的操作环境,它允许用户同时运行多个应用程序,或在一个程序中同时做几件事情。每个程序在屏幕上占据一块矩形区域,这个区域称为窗口,并且窗口是可以重叠的。用户可以移动这些窗口,在不同的应用程序之间进行切换,并可以在程序之间进行手工和自动的数据交换和通信。

当前,最新的个人计算机 Windows 版本是 Windows 7(桌面如图 1-3 所示),最新的服务器版本 Windows 是 Windows Server 2008 R2。2011 年 9 月,微软公司正式公布了 Windows 8 预览版,该系统除了具备微软公司传统视窗显示外,还配合触控屏幕的平板电脑设计,使用了类似 Windows Phone 7 操作系统中的动态方块(live tiles)界面。

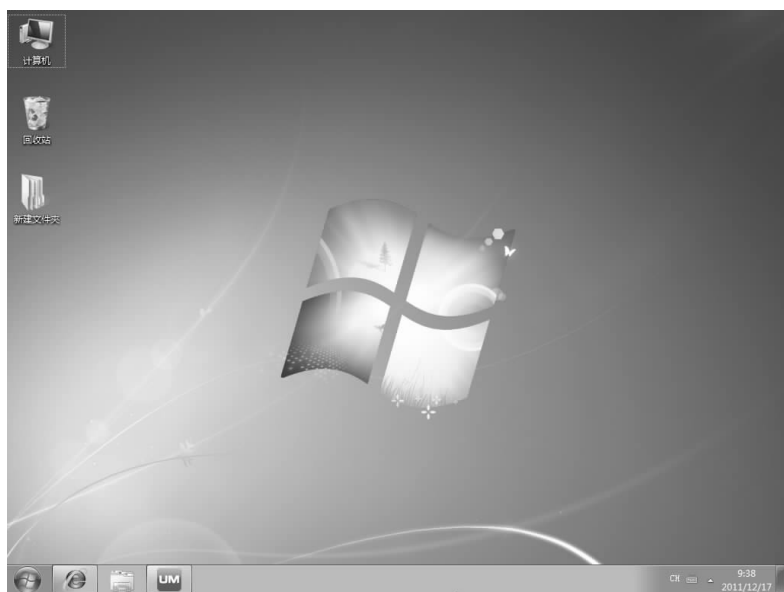


图 1-3 Windows 7 桌面

## 1.6.2 UNIX 和 Linux 操作系统

### 1. UNIX 操作系统

UNIX 于 20 世纪 60 年代末期在贝尔实验室问世,是一个可以在个人计算机、小型计算机、中型计算机、巨型计算机和工作站上安装的操作系统。UNIX 操作系统的问世和发展对计算机硬件和软件的发展作出了巨大贡献并产生了深刻影响。

UNIX 具有以下 4 个显著的特点。

(1)UNIX 是可移植的操作系统。它可以不经过较大的改动而方便地从一个平台移植到另一个平台。在 UNIX 操作系统下开发的应用程序移植也同样方便,其根本原因是 UNIX 主要部分是用 C 语言编写的(而不是用于特定操作系统的机器语言),采用 C 语言编写操作系统使得操作系统的兼容性非常好。

(2)UNIX 拥有一套功能强大的工具(命令)集。它们能够组合起来解决许多问题,而这





一工作在其他操作系统中则需要通过编程来完成。

(3)UNIX 对设备的操作与具体设备无关。用户不必关心设备和特性,因为操作系统本身就包含了设备的驱动程序,这意味着它可以方便地配置和运行任何设备。

(4)操作系统是开放式的。任何人只要遵守系统的标准都可以给操作系统编写驱动程序和系统代码,并且 UNIX 的源代码是公开的。

概括地说,UNIX 具有强大的操作系统所拥有的一切特点,包括多道程序、虚拟内存、开放性和兼容性、非常优秀的文件系统和目录系统等。

## 2. Linux 操作系统

Linux 是全世界影响力最大的免费使用和自由传播的类似于 UNIX 的操作系统,它的最初作者是芬兰学生、计算机业余爱好者 Linus Torvalds。经过数十年全球各国编程爱好者的努力,Linux 发展迅猛,成为除 Windows 以外使用最广的操作系统。

Linux 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设计和实现的,其目的是建立不受任何商品化软件版权制约的、全世界都能自由使用的 UNIX 兼容产品。Linux 属于自由软件,任何用户不用支付任何费用就可以获得该操作系统和它的源代码,并且可以根据资金的需要对它进行自由的拼装和必要的修改,无偿地使用它,无约束地继续传播。它具有 UNIX 操作系统的全部功能,任何会使用 UNIX 操作系统的用户不必重新学习和培训都可以直接使用 Linux 操作系统,想要学习 UNIX 操作系统的人也可以从学习 Linux 中得到帮助。Linux 操作系统的整体设计是为了让操作系统在微处理机上更有效地运行。

Linux 可以运行在各种硬件平台上。目前,Linux 主要应用于网络服务器领域,因为它免费、开源、稳定、安全,所以非常适合网络应用。例如,现在最常见的建站程序 PHP,就是基于 Linux 系统开发的语言。Linux 还应用于嵌入式系统,如机顶盒、移动电话等。

大多数普通用户可使用的 Linux 操作系统都是相关公司或爱好者基于 Linux 内核开发的 Linux 发行版。主流 Linux 发行版中通常有 Red Hat、OpenLinux、SUSE、TurboLinux、Ubuntu 等。

图 1-4 所示为 Red Hat Linux 9 默认的 GNOME 界面。

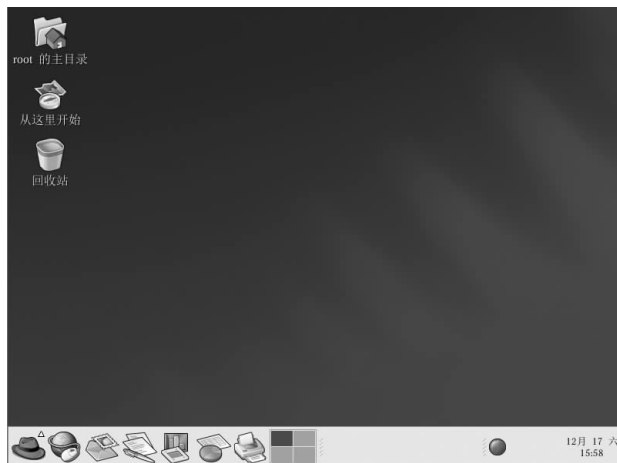


图 1-4 Red Hat Linux 9 默认的 GNOME 界面



## 1.7 习 题

### 一、选择题

1. 操作系统是一种( )。  
A. 通用软件            B. 系统软件            C. 应用软件            D. 软件包
2. 操作系统是对( )进行管理的软件。  
A. 系统软件            B. 系统硬件            C. 计算机资源            D. 应用程序
3. 操作系统中采用多道程序设计技术,以提高 CPU 和外部设备的( )。  
A. 利用率            B. 可靠性            C. 稳定性            D. 兼容性
4. 计算机系统中配置操作系统的目的是提高计算机的( )和方便用户使用。  
A. 速度            B. 利用率            C. 灵活性            D. 兼容性
5. ( )操作系统允许多个用户在其终端上同时交互地使用计算机。  
A. 批处理            B. 实时            C. 分时            D. 多道批处理

### 二、填空题

1. 按功能划分,软件可分为\_\_\_\_\_软件和\_\_\_\_\_软件两种。
2. 操作系统是在\_\_\_\_\_上加载的第一层软件,是对计算机硬件系统功能的首次扩充。
3. 操作系统的主要作用是\_\_\_\_\_和\_\_\_\_\_。
4. 如果一个操作系统在用户提交作业后,不提供交互能力,只追求计算机资源的利用率、大吞吐量和作业流程的自动化,则属于\_\_\_\_\_操作系统。
5. 现代操作系统都具有并发性、\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_,其中并发性是操作系统最重要的特征,其他三个特性均基于并发性而存在。

### 三、简答题

1. 什么是操作系统? 操作系统的主要功能是什么?
2. 简述多道批处理系统的优缺点。
3. 嵌入式操作系统除了具备一般操作系统最为基本的功能外,还有哪些特点?
4. 简述 Windows 系列操作系统与 Linux 操作系统的主要区别。