

第 1 章 操作系统概述

计算机科学与技术发展到今天,从个人计算机到大型计算机,无一例外地都配置了一种或多种操作系统。如果要让用户去使用一台没有操作系统的计算机,那将是难以想象的。那么,什么是操作系统(Operating System, OS)呢?本章主要介绍操作系统的概念、操作系统的形成与发展、操作系统的特征与功能,最后介绍几种现代主流操作系统。

通过本章的学习使学生了解操作系统的发展过程,熟悉操作系统的功能,掌握操作系统的概念和特征。

1.1 操作系统的概念

1.1.1 计算机系统

一个完整的计算机系统,不论是大型机、小型机还是微型机,都由两大部分组成:计算机硬件和计算机软件,如图 1-1 所示。它们组成一个统一整体,各个组成部分相互联系、相互作用,共同完成所分配的各项任务。

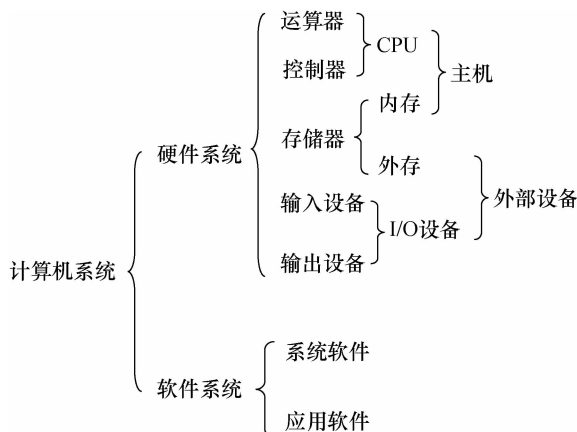


图 1-1 计算机系统的组成

1. 计算机硬件

计算机硬件是指构成计算机系统所必须配置的各种设备,是“看得见,摸得着”的物理部件,它是组成计算机系统的物质基础。计算机硬件主要由运算器、控制器、存储器、输入设备

和输出设备组成,它们之间的关系如图 1-2 所示。

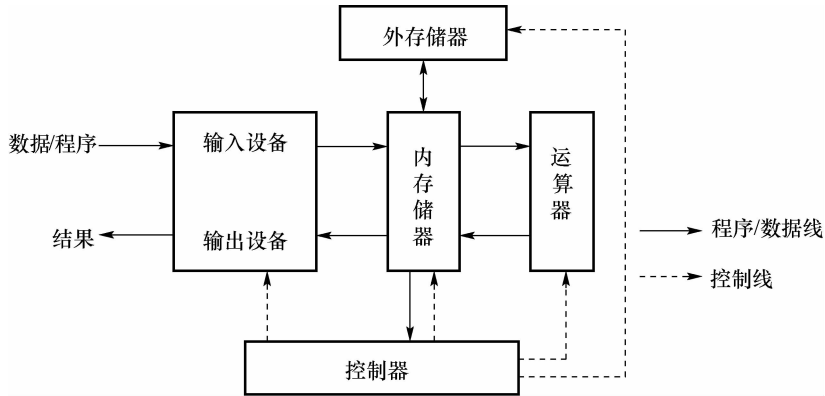


图 1-2 计算机硬件之间的关系

2. 计算机软件

计算机软件是指由计算机硬件执行以完成一定任务的程序及其数据。计算机软件分为系统软件和应用软件。计算机用户通过应用软件让计算机为自己服务,而应用软件又是通过系统软件来管理和使用计算机硬件。

系统软件是支持和管理计算机硬件的软件,是服务于硬件的,与具体的应用领域无关。它提供的是一个平台,如编译软件和操作系统等。应用软件是完成用户某项要求的软件,是服务于特定用户的,它满足某一个应用领域的要求。

计算机硬件和计算机软件在计算机系统中是相辅相成、缺一不可的,它们共同组成了计算机系统。计算机硬件是计算机的躯体和基础,计算机软件是计算机的头脑和灵魂,没有软件的计算机和缺少硬件的计算机都不能称为完整的计算机系统。

1.1.2 什么是操作系统

计算机系统由硬件和软件两部分构成。操作系统属于软件中的系统软件,是紧挨着硬件的第一层软件,是对硬件功能的首次扩充,其他软件则是建立在操作系统之上的。通过操作系统对硬件功能进行扩充,并在操作系统的统一管理和支持下运行各种软件。操作系统与硬软件的关系如图 1-3 所示。

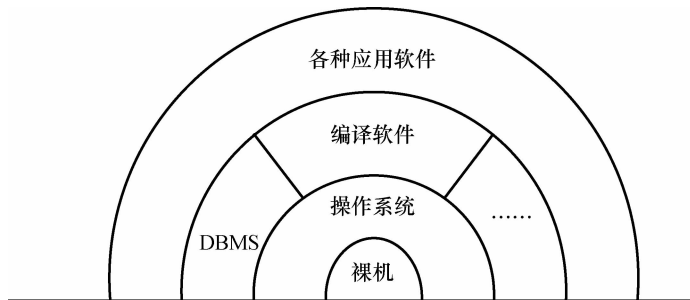


图 1-3 操作系统与硬软件的关系

操作系统是运行在计算机上的最基本的程序。在操作系统的支持下,可以维持计算机上的所有组成部分,如键盘、显示器、内存、硬盘、CPU 及应用软件等共同工作;可以控制外围设备,如磁盘驱动和打印机等。操作系统还提供了一个执行其他应用程序的软件平台,而不管其硬件情况。对于大型的分布式系统,操作系统可以控制同时运行的不同程序和用户。此外,操作系统通过各种方式可以支持计算机和网络的安全性。

因此,操作系统在计算机系统中占据着一个非常重要的地位,它不仅是硬件与所有其他软件之间的接口,而且任何数字电子计算机,从微处理器到巨型计算机都必须在其硬件平台上加载相应的操作系统之后,才能构成一个可以协调运转的计算机系统。只有在操作系统的指挥控制下,各种计算机资源才能被分配给用户使用。也只有操作系统的支撑下,其他系统软件如各类编译系统、程序库、运行支持环境才得以取得运行条件。没有操作系统,任何应用软件都无法运行。可见,操作系统实际上是一个计算机系统中硬、软件资源的总指挥部。操作系统的性能高低,决定了整体计算机的潜在硬件性能能否发挥出来。

从前面的介绍我们了解了操作系统在计算机系统中的地位,它是计算机硬件和其他软件以及计算机用户之间的联系纽带,如果没有操作系统,用户几乎无法使用计算机系统。那么,什么是操作系统呢?不同计算机使用者的看法可能不同,下面从不同角度来讨论操作系统的概念。

1. 用户环境的观点

从用户的角度来看,操作系统是用户与计算机硬件系统之间的接口,用户通过操作系统来使用计算机系统,即用户在操作系统的支持下,能够方便、快捷、安全、可靠地操纵计算机硬件资源,运行自己的程序。用户可通过以下三种方式使用计算机:其一,直接使用操作系统提供的键盘命令或 Shell 命令语言;其二,利用鼠标单击窗口中的图标,以执行相应的应用程序,如 Windows 操作系统的图形用户接口;其三,在应用程序中调用操作系统的内部功能模块,即系统调用接口。这些接口和各种应用程序为用户开发和运行应用软件提供了便利的环境和手段。

2. 资源管理的观点

把操作系统看做是系统资源的管理者,是目前关于操作系统描述的主要观点。现代计算机系统通常包括各种各样的资源,总体上可分为处理器、存储器、I/O 设备和文件四类。因此,操作系统的功能就是负责对计算机的这些软、硬件资源进行控制、调度、分配和回收,协调系统中各程序对资源使用请求的冲突,保证各程序都能顺利运行完成。

3. 虚拟机观点

通常把覆盖了软件的机器称为虚拟机,一台完全无软件的计算机称为“裸机”,即使其功能再强,也是难于使用的。从这一观点来看,操作系统为用户使用计算机提供了许多服务功能和良好的工作环境,用户不再直接使用“裸机”,而是通过操作系统来控制和使用计算机,从而把计算机扩充为功能更强、使用更加方便的虚拟计算机。

综上所述,我们把操作系统定义如下:操作系统是一组控制和管理计算机硬件和软件资源,合理地组织计算机工作流程,以及方便用户使用的程序的集合。

1.1.3 操作系统的目标

目前,操作系统的种类繁多,不同类型的操作系统其实现目标也不尽相同。但是,要设计和编制一个操作系统,必须实现以下目标:

1. 方便性

从图 1-3 可以看出,操作系统最终是要为用户服务的。所以,设计操作系统时必须考虑用户能否方便地操作计算机。用户的操作包括直接使用计算机完成各种命令,也包括通过设计程序让计算机完成各种任务。

2. 有效性

在未配置操作系统的计算机系统中,诸如 CPU、I/O 设备等各类资源,都会经常处于空闲状态而得不到充分利用;内存及外存中所存放的数据由于无序而浪费了存储空间。配置了操作系统后,可使 CPU 和 I/O 设备由于能保持忙碌状态而得到更为有效的利用,且由于使内存和外存中存放的数据有序而节省了存储空间。此外,操作系统要合理地组织计算机的工作流程,提高系统资源的利用率,增加系统的吞吐量,从而使有限的资源完成更多的任务。

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,它们对操作系统提出了更高的功能和性能要求。因此,操作系统必须具有很好的可扩充性才能适应发展的要求。这就是说,在设计操作系统的体系结构时,要采用合理的结构使其能够不断地扩充和完善。

4. 开放性

操作系统的主要功能是管理计算机硬件,它必须适应和管理不同的硬件。随着计算机硬件技术的发展,不同厂家的新型的、集成化的硬件不断涌现出来。为了使这些硬件产品能够正确、有效地协同工作,就必须实现应用程序的可移植性和互操作性,因而要求计算机系统具有统一的开放环境,其中首先是要求操作系统具有开放性。

1.2 操作系统的形成与发展

操作系统是由于客观的需要而产生的,它伴随着计算机技术本身及其应用的日益发展而逐渐发展和不断完善。它的功能由弱到强,在计算机系统中的地位不断提高。至今,它已成为计算机系统中的核心,没有计算机系统是不配置操作系统的。

1.2.1 推动操作系统发展的动力

操作系统的形成迄今已有 50 多年的时间。在 20 世纪 50 年代中期出现了第一个简单的批处理操作系统。到 20 世纪 60 年代中期产生了多道程序批处理系统,不久又出现了基

于多道程序的分时系统。20世纪70年代出现了微机和局域网络,同时也产生了微机操作系统和网络操作系统,之后又出现了分布式操作系统。在这短短的50多年中,操作系统取得如此巨大的进展,其主要动力可以归结为以下4个方面:

1. 不断提高计算机资源利用率的需要

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,这就推动了人们不断发展操作系统的功能,由此产生了批处理系统。它能自动地对一批作业进行处理。

2. 方便用户操作

当资源利用率不高的问题得到基本解决后,用户在上机操作、调试程序时的不方便就成为主要矛盾。于是,人们就想方设法改善用户上机和调试程序的条件,这又成为继续推动操作系统发展的主要因素,随之便形成了允许人机交互的分时系统,或称为多用户系统。

3. 硬件的不断更新换代

计算机硬件在不断的更新,从电子管到晶体管,到集成电路,到大规模集成电路,使得计算机的性能不断提高,从而推动了操作系统的性能和功能的不断改进和完善。

4. 计算机体系结构的不断发展

计算机体系结构的发展也不断地推动着操作系统的发展,并产生了新的操作系统。例如,当计算机由单处理器系统发展为多处理器系统时,操作系统也从单处理器操作系统发展为多处理器操作系统。又如,当计算机网络出现后,也就产生了网络操作系统。

1.2.2 操作系统的形成

操作系统从无到有,从小到大,从弱到强,其发展大致经历了以下4个阶段。

1. 无操作系统

无操作系统的计算机系统,其资源管理和控制由人工负责,它采用两种方式:人工操作方式和脱机输入输出方式。

(1) 人工操作方式

从第一台电子计算机 ENIAC(埃尼阿克)诞生到20世纪50年代中期的计算机都没有出现操作系统,这时计算机资源的管理是由操作员采用人工方式直接控制的。用户既是程序员又是操作员。上机完全是手工操作:程序员先将事先已穿孔(对应于程序和数据)的纸带(或卡片)装入纸带输入机(或卡片输入机),然后启动输入机将程序和数据输入到计算机中,接着启动计算机运行。当程序运行完毕,用户卸下并取走纸带(或卡片)后,才让下一个用户上机。

这种人工操作方式的特点是:

- ① 用户独占全机。一台计算机的全部资源只能由一个用户独占。
- ② CPU 等待人工操作。当用户进行装带(卡)、卸带(卡)等人工操作时,CPU 是空闲的。

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。随着计算机

CPU 速度的提高,人工操作的低速率与计算机主机运行的快速运算之间速度不匹配的矛盾日趋严重。为了解决上述矛盾,引入了脱机输入输出方式。

(2)脱机输入输出方式

为了解决 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。

脱机输入输出技术是指事先将装有用户程序 and 数据的纸带(或卡片)装入纸带(或卡片)输入机,在一台外围机的控制下把纸带(卡片)上的数据(程序)输入到磁盘(带)上。当计算机主机需要这些程序和数据时,再从磁盘(带)上高速地调入主存。类似地,当计算机主机需要输出时,可由计算机主机直接高速地把数据从内存送到磁盘(带)上,然后再在另一台外围机的控制下,将磁盘(带)上的结果通过相应的输出设备输出。

简单地讲,脱机输入输出方式是指程序和数据的输入输出是在外围机的控制下,而不是在主机控制下完成的。脱机输入输出方式示意图如图 1-4 所示。

脱机输入输出技术减少了计算机主机的空闲等待时间,提高了 I/O 设备的处理速度。如果输入输出是在主机的控制下完成的则称为联机输入输出。

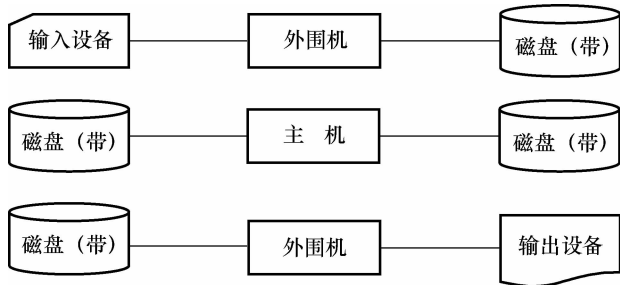


图 1-4 脱机输入输出方式

2. 批处理系统

批处理系统主要采用了批处理技术。批处理技术是计算机系统对一批作业自动进行处理的一种技术。批处理系统有单道批处理系统和多道批处理系统两种形式。

(1)单道批处理系统

单道批处理系统是 20 世纪 50 年代 General Motors 研究室在 IBM701 计算机上实现的第一个操作系统。如果把用户在一次解题或一个事务处理过程中要求计算机系统所做的工作的集合称为作业的话,通常是把一批作业以脱机输入方式输入到磁盘(带)上,并在系统中配上监督程序,在监督程序的控制下使这批作业能一个接一个地连续处理。

自动批处理过程:由监督程序将磁盘(带)上的第一个作业调入主存,并把运行控制权交给该作业;该作业处理完后,又将控制权交给监督程序;监督程序再将磁盘(带)上的第二个作业调入主存,并把运行控制权交给该作业;如此反复,直到磁盘(带)上的所有作业全部完成。

由于系统对作业的处理都是成批地进行的,且在内存中始终只保持一道作业,故称单道批处理系统。图 1-5 给出了单道程序运行的工作情况。

单道批处理系统的特点是：

- ①自动性。磁盘(带)上的一批作业能自动地逐个作业依次执行,而无需人工干预。
- ②顺序性。磁盘(带)上的作业是顺序地进入内存的,先调入内存的作业先完成。
- ③单道性。内存中仅有一个程序并使之运行。

单道批处理系统大大减少了人工操作的时间,提高了机器的利用率。但是,在单道批处理作业运行时,主存中仅存放了一道程序,每当程序发出 I/O 请求时,CPU 便处于等待 I/O 完成状态,致使 CPU 空闲,特别是 I/O 设备的低速性,使 CPU 的利用率降低。

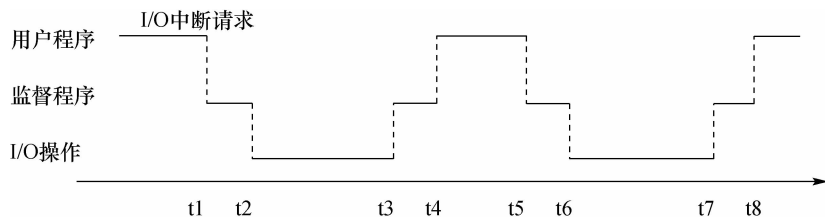


图 1-5 单道程序运行情况

(2) 多道批处理系统

多道批处理系统是在 20 世纪 60 年代设计的。为了改善 CPU 的利用率,提高机器的使用效率,在单道批处理系统中引入了多道程序设计技术,形成了多道批处理系统,它使 CPU 与外设可以并行工作。多道程序设计技术是指同时把多个作业放入内存并允许它们交替执行,共享系统中的各类资源,当某个程序因某种原因而暂停执行时,CPU 立即转去执行另一道程序。图 1-6 给出了四道程序运行的工作情况。

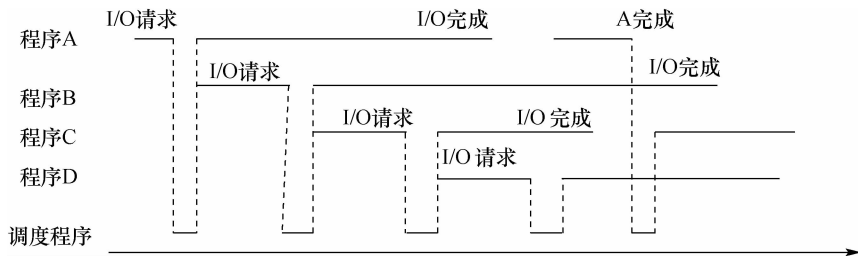


图 1-6 四道程序运行情况

多道批处理系统的特点是：

①多道性。在内存中可以同时驻留多道程序,并允许它们并发执行,从而有效地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先后顺序没有严格的对应关系,即先进入内存的作业不一定先完成,后进入内存的作业不一定最后完成。

③调度性。作业从提交给系统开始直至完成,需要经过两次调度：

一是作业调度,它是按照一定的作业调度算法,从外存的后备作业队列中选择若干个作业调入内存。

二是进程调度,它是按照一定的进程调度算法,从内存中已有的作业中选择一个作业,将处理器分配给该作业,使之运行。

多道批处理系统的优点:

①资源利用率高。由于在内存中装入了多道程序,它们共享资源,使资源尽可能处于忙碌状态,从而提高了资源的利用率。

②系统吞吐量大。系统吞吐量是指系统在单位时间内所完成的工作总量。能提高系统吞吐量的原因可归结为:第一,CPU 和其他资源保持“忙碌”状态;第二,仅当作业完成时或运行不下去时才进行切换,系统开销小。

多道批处理系统的不足:

①平均周转时间长。作业的平均周转时间是指从作业装入系统开始,到完成并退出系统所经过的时间。在批处理系统中,由于作业要排队,要经过两次调度依次进行处理,因而作业的周转时间长。

②无交互能力。用户一旦把作业提交给系统后,直至作业完成,用户都不能与自己的作业进行交互,这对修改和调试程序都是极不方便的。

3. 分时操作系统

(1) 分时系统的产生

如果说,推动多道批处理系统形成和发展的主要动力是提高资源利用率和系统吞吐量,那么,推动分时系统形成和发展的主要动力,则是用户的需要。具体地说,用户的需要表现在以下几个方面:

①人一机交互。对于一程序员来说,希望能方便地上机调试、控制、修改程序。

②共享主机。在 20 世纪 60 年代,计算机十分昂贵,只能多个用户共享一台计算机,用户希望在用机时能够像自己独占计算机一样,不仅可以随时与计算机交互,而且感觉不到其他用户在使用该计算机。

③便于用户上机。用户希望能通过本地终端直接将作业传送到机器上进行处理,并能对本地作业进行控制。

分时系统恰是为了满足上述的用户需要所形成的一种新型操作系统。它与多道批处理系统有着截然不同的性能。由上述不难得知,分时系统是指一台主机上连接了多个带有显示器和键盘的终端,同时允许多个用户以分时方式共享主机中的资源,每个用户都可以通过本地终端以交互的方式使用计算机,如图 1-7 所示。

所谓分时技术就是把处理器的运行时间划分成很短的时间片,根据时间片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,

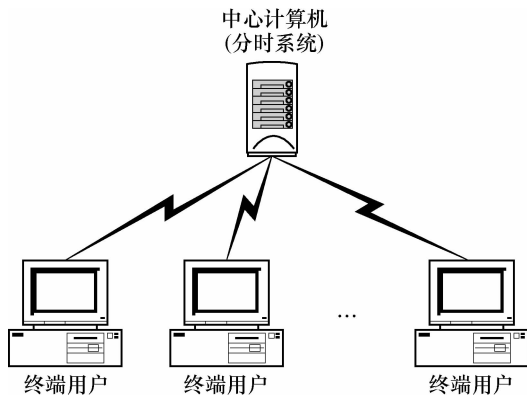


图 1-7 分时系统

则该作业被暂时中断,把处理器让给下一个作业使用,被中断作业等待下一轮时再继续运行。由于计算机的速度很快,作业运行轮转得也很快,这样给每个用户的感受就像独占了一台计算机一样。

(2) 分时系统实现中的关键问题

①及时接收。要及时接收用户键入的命令或数据并不困难,只需在系统中配置一个多路卡。此外,还需为每个终端配置一个缓冲区,用来暂存用户键入的命令。

②及时处理。人机交互的关键是使用户键入命令后,能及时地控制或修改本地作业。为此,要让所有的用户作业直接进入主存,在不长的时间内(如3秒)使每个作业运行一次,从而使用户的作业得到及时处理。

(3) 分时系统的特征

分时系统的特征有多路性、独立性、及时性和交互性。

①多路性。允许在一台主机上同时连接多台终端、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;微观上,则是每个用户轮流运行一个时间片。多路性也称同时性,它提高了资源的利用率,从而促使计算机更广泛地应用。

②独立性。每个用户占用一个终端,彼此独立操作、互不影响。因此,每个用户会感觉到独自占用了主机。

③及时性。用户的请求能在很短的时间内获得响应,此时的时间间隔是根据人们能接受的等待时间来确定的,通常为2~3秒钟。

④交互性。用户可以通过终端与系统进行广泛的对话。其广泛性表现在:用户可以请求系统提供各方面的服务,如文件编辑、数据处理和资源共享等。

4. 实时系统

多道批处理系统和分时系统使资源的利用率得以提高,系统的响应时间缩短,从而使计算机的应用范围日益扩大。但在实时控制和实时信息处理中,要求系统的响应时间更短,这就产生了实时系统。

(1) 实时系统的概念

实时系统是指系统能及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。

(2) 实时系统的类型

根据控制对象的不同,实时系统分为实时控制系统和实时信息处理系统。

①实时控制系统。实时控制系统包含以计算机为中心的生产过程控制系统和武器控制系统,又称为计算机控制系统。系统要求能及时采集现场数据,并对采集的数据进行及时处理,进而自动控制相应的执行机构,使某些参数能按预定的规律变化,以保证产品的质量和提高产量。通常用于工业控制、军事控制等领域。如飞机自动驾驶系统、火箭飞行控制系统、导弹制导系统等。

②实时信息处理系统。实时信息处理系统是指对信息进行实时处理的系统。在该系统中,计算机能及时接收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,并在很短的时间内向用户做出正确应答。典型的实时信息处理系统有机票订购系统、情报检索系统等。

(3) 实时系统的特征

实时系统的特征有多路性、独立性、及时性、交互性和可靠性。

①多路性。是指系统能对多个现场进行数据采集,并对多个对象或多个执行机构进行控制。

②独立性。是指信息的采集和对象的控制操作互不干扰。

③及时性。是以控制对象所要求的开始时间和截止时间来确定的,高于分时系统,一般为秒级、毫秒级,甚至微秒级。

④交互性。是指用户可访问系统中某些特定的专用服务程序,其交互性弱于分时系统。

⑤可靠性。是指采用多级容错技术来保证系统的安全性和数据的安全性。其可靠性高于分时系统。

(4) 实时系统与分时系统的主要区别

①系统的设计目标不同。分时系统的设计目标是提供一种随时可供多个用户使用的通用性很强的系统;而实时系统则大多数都是具有某种特殊用途的专用系统。

②响应时间的长短不同。分时系统的响应时间通常为秒级;而实时系统的响应时间通常为毫秒级,甚至微秒级。

③交互性的强弱不同。分时系统的交互性强,而实时系统的交互性相对较弱。

批处理系统、分时系统和实时系统是3种基本的操作系统类型。而一个实际的操作系统,可能兼有三者或其中两者的功能,则称该操作系统为通用操作系统。

1.2.3 操作系统的进一步发展

1. 微机操作系统

微机操作系统是指配置在微机上的操作系统。最早出现的微机操作系统是8位微机上的CP/M操作系统。微机操作系统可分为单用户单任务操作系统、单用户多任务操作系统和多用户多任务操作系统。

(1) 单用户单任务操作系统

单用户单任务操作系统是指只允许一个用户上机,且只允许用户程序作为一个任务运行。这是一种最简单的微机操作系统,主要配置在8位微机和16位微机上。具有代表性的单用户单任务操作系统是CP/M和MS-DOS。

①CP/M。CP/M是Control Program Monitor的缩写,它是Digital Research公司于1975年推出的8位微机操作系统。它具有较好的层次结构、可适应性、可移植性和易学易用性。使它在8位微机中占据了统治地位,成为8位微机操作系统的标准。

②MS-DOS。DOS是Disk Operating System的缩写,它是MS公司于1981年推出的16位机操作系统。它在CP/M系统上进行了较大的扩充,增加了许多内部命令和外部命令。该操作系统具有较强的功能和性能优良的文件系统,占据了16位微机操作系统的统治地位,成为16位微机操作系统的标准。

(2) 单用户多任务操作系统

单用户多任务操作系统是指只允许一个用户上机,但允许一个用户程序分为多个任务并发执行,从而有效地改善系统的性能。它主要配置在 32 位微机上,最具代表性的单用户多任务操作系统是 OS/2 和 MS-Windows。

①OS/2。OS/2 是 IBM 公司于 1987 年推出的 16/32 位机操作系统。

②MS-Windows。Windows 是 MS 公司于 1990 年推出的 32 位机操作系统。它具有易学易用、用户界面友好、多任务控制等特点。特别是 Windows 95 版本和 Windows NT 版本的出现,使之很快地流行起来,成为微机的主流操作系统,并使 Windows 走向多用户操作系统。

(3) 多用户多任务操作系统

多用户多任务操作系统是指允许多个用户通过各自的终端,使用同一台主机,共享主机系统中的各类资源,而每个用户程序又可分为多个任务并发执行,从而提高资源的利用率和增加系统的吞吐量。它主要配置在大、中、小型计算机上,具有代表性的是 UNIX。

UNIX 是 Uniplexed Information and Computer Systems 的缩写,它是美国电报电话公司的 Bell 实验室于 1976 年推出的操作系统,可在微机、小型机和大型机上运行。

2. 多处理器操作系统

为了增加系统的吞吐量,节省投资,提高系统的可靠性,在 20 世纪 70 年代出现了多处理器系统(MPS: Multi-Processor System),试图从计算机体系结构上来改善系统的性能。

(1) 多处理器操作系统的概念

在多处理器系统上配置的操作系统称为多处理器操作系统。

(2) 多处理器操作系统的类型

根据多个处理器之间耦合的紧密程度,把多处理器系统分为紧密耦合 MPS 和松散耦合 MPS 两种类型。紧密耦合 MPS 是通过高速总线或高速交叉开关来实现多个处理器之间的互联,它们共享主存和 I/O 设备,系统中的所有资源都由操作系统实施统一的控制和管理。松散耦合 MPS 是通过通道或通信线路来实现多个计算机之间的互联,每台计算机都有各自的存储器和 I/O 设备,并配置了操作系统来管理本地资源和在本地运行的进程。多处理器操作系统可以分为非对称多处理器模式和对称多处理器模式两种。

①非对称多处理器模式。也称为主-从模式,在这种模式中,把处理器分为主处理器和从处理器两类。主处理器只有一个,其上配置了操作系统,用于管理整个系统的资源,并负责为各从处理器分配任务。从处理器有若干个,它们执行预先规定的任务及由主处理器所分配的任务。这种模式易于实现,但资源利用率低,在早期的特大型系统中,较多地采用了这种模式。

②对称多处理器模式。在这种模式中,所有处理器的地位都是相同的。在每个处理器上运行一个相同的操作系统副本,用它来管理本地资源,并控制进程的运行以及各计算机之间的通信。这种模式允许多个进程同时运行,但必须谨慎控制 I/O 设备,以保证能将数据送至适当的处理器,同时还必须使各处理器的负载平衡,以免有的处理器超载运行,而有的处理器空闲无事。

3. 网络操作系统

信息时代离不开计算机网络,特别是 Internet 的广泛应用正在改变着人们的观念和社会生活的方方面面。每天有成千上万人通过网络传递邮件、查阅资料、搜寻信息,以及网上订票、网上购物等。单台计算机的资源毕竟有限,为了实现计算机之间的数据通信和资源共享出现了计算机网络。

计算机网络是指通过通信线路和通信的控制设备,将相互独立的计算机系统连成一个整体,在网络软件的控制下,实现信息传递和资源共享的系统。所谓独立的计算机系统是指计算机具有独立处理能力;网络软件主要是指网络操作系统和网络应用软件。

(1) 网络操作系统的模式

网络操作系统的模式有:客户机/服务器模式(C/S)和对等模式两种。

①客户机/服务器模式。这种模式(C/S)是20世纪80年代发展起来的,是目前仍广为流行的网络工作模式。网络中有两种站点:服务器和客户机。服务器是网络的控制中心,它向客户机提供一种或多种服务。客户机是用于本地的处理和访问服务器的站点。C/S模式具有分布处理和集中控制的特征。

②对等模式。在对等模式中,各站点的关系是对等的,既可以作为客户机访问其他站点,又可以作为服务器向其他站点提供服务。该模式具有分布处理和分布控制的特征。

(2) 网络操作系统的功能

网络操作系统应具有下述5方面的功能:

①网络通信。这是网络最基本的功能,其任务是在源主机和目标主机之间实现无差错的数据传输。

②资源管理。对网络中的共享资源(硬件和软件)实施有效的管理、协调诸用户对共享资源的使用、保证数据的安全性和一致性。

③网络服务。这是在前两个功能的基础上,为了方便用户而又直接向用户提供的多种有效服务。主要的网络服务有:电子邮件服务;文件传输、存取和管理服务;共享硬盘服务;共享打印服务。

④网络管理。网络管理最基本的任务是安全管理。通过“存取控制”来确保存取数据的安全性;通过“容错技术”来保证系统出现故障时数据的安全性。

⑤互操作能力。所谓互操作,在客户机/服务器模式的LAN环境下,是指连接在服务器上的多种客户机和主机不仅能与服务器通信,而且还能以透明的方式访问服务器上的文件系统;而在互联网络环境下的互操作是指不同网络间的客户机不仅能通信,而且也能以透明方式访问其他网络中的文件服务器。

4. 分布式操作系统

(1) 分布式操作系统的概念

在以往的计算机系统中,其处理和 control 功能都高度地集中在一台主机上,所有的任务都由主机处理,这样的系统称为集中式处理系统。

分布式系统则是系统的处理和 control 功能,都分散在系统的各个处理单元上。系统中的

所有任务,也可动态地分配到各个处理单元上去,并使它们并行执行,实现分布处理。

所谓分布式处理系统是指由多个分散的处理单元经网络连接而形成的系统。在分布式系统上配置的操作系统称为分布式操作系统。

(2) 分布式操作系统的特点

分布式操作系统具有以下特点:

①分布性。分布式操作系统不是集中地驻留在某一个站点上的,而是均匀地分布在各个站点上,它的处理和控制在分布式的。

②并行性。分布式操作系统的任务分配程序将多个任务分配到多个处理单元上,使这些任务并行执行,从而提高了任务执行的速度。

③透明性。它可以很好地隐藏系统内部的实现细节,而对象的位置、并发控制、系统故障等对用户是透明的。

④共享性。分布在各个站点上的软、硬件资源,可供全系统中的所有用户共享,并以透明的方式访问它们。

⑤健壮性。任何站点上的故障都不会给系统造成太大的影响;当某一设备出现故障时,可通过容错技术实现系统重构,从而保证系统的正常运行。

(3) 分布式操作系统与网络操作系统的区别

分布式操作系统与网络操作系统的主要区别:

①能否适用不同的操作系统。网络操作系统可以构架于不同的操作系统之上,也就是说,它可以在不同的本机操作系统上,通过网络协议实现网络资源的统一配置,在大范围内构成网络操作系统;而分布式操作系统是由一种操作系统构架的。

②对资源的访问方式不同。网络操作系统在访问系统资源时,需要指明资源的位置和类型,对本地资源和异地资源的访问要区别对待;而分布式操作系统对所有资源,包括本地资源和异地资源,都用同一方式进行管理和访问,用户不必关心资源在哪里,或资源是怎样存储的。

5. 嵌入式操作系统

在机器人、掌上电脑、车载系统、智能家用电器、手机等设备上,通常会嵌入安装各种微处理器或微控制芯片。嵌入式操作系统就是运行在嵌入式智能芯片环境中,对整个智能芯片以及它所操作、控制的各种部件装置等资源进行统一协调、调度、指挥和控制的系统软件。

与一般操作系统相比,嵌入式操作系统具有微小、实时、专业、可靠、易裁剪、应用领域差别大的特点。代表性的嵌入式操作系统有 Symbian、WinCE、Linux、Palm OS、VxWorks 等。

1.3 操作系统的特征与功能

1.3.1 操作系统的特征

不同操作系统的特征各不相同。批处理操作系统主要突出成批处理的特点,分时操作

系统主要突出交互的特点,实时操作系统主要突出实时的特点。但这几种操作系统都具有以下基本特征:

1. 并发性

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,并发性是指宏观上在一段时间内有多道程序同时运行,但在单处理器系统中,每一个时刻仅能执行一道程序,故微观上这些程序是在交替执行的。

并行性和并发性是既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个时刻发生。并行的若干事件是并发的,反之,则不一定。

并发的目的是改善系统的利用率和提高系统的吞吐量。

2. 共享性

共享性是指系统中的资源可供多个并发执行的进程使用。根据资源的属性,把共享分为互斥共享和同时共享两种方式。

①互斥共享。是指系统中的资源,如打印机、扫描仪等,虽然它们可供多个进程使用,但在一段时间内只允许一个进程访问该资源。当这个资源正在被使用时,其他请求该资源的进程必须等待,仅当该进程访问完并释放该资源后,才允许另一进程对该资源进行访问。

②同时共享。是指系统中有些资源,如磁盘,允许在一段时间内有多个进程同时对它们进行访问。

这里提到的进程,是指程序的一次执行,是程序在一个数据集合上运行的过程,是系统进行资源分配和调度的一个独立单位。本书将在第2章详细介绍。

并发性和共享性是操作系统的两个最基本特征,它们互为存在条件。一方面,资源共享是以程序(进程)的并发执行为存在条件,若系统不允许并发执行,自然不存在资源共享问题。另一方面,若系统不能对资源共享实施有效管理,则将影响程序的并发执行,甚至无法执行。

3. 虚拟性

虚拟性是指通过某种技术把一个物理实体变成若干个逻辑实体。即物理上虽然只有一个实体,但用户使用时感觉有多个实体可供使用。

例如通过多道程序设计技术,可以实现处理器的虚拟;通过请求调进/调出技术,可以实现存储器的虚拟;通过 SPOOLing 技术,可以实现设备的虚拟。

在操作系统中虚拟的实现,主要是通过分使用的方法。显然,如果 n 是某一物理设备所对应的虚拟的逻辑设备数,则虚拟设备的速度必然是物理设备速度的 $1/n$ 。

4. 异步性

异步性也称为不确定性。是指在多道程序环境下,允许多个进程并发执行,由于资源的限制,进程的执行不是“一气呵成”的,是“走走停停”的,使得多个程序的运行顺序和每个程序的运行时间是不确定的。具体说,各个程序什么时候得以运行、在执行过程中是否被其他事情打断暂停执行、向前推进的速度是快还是慢等都是不可预知的,这由程序执行时的现场所决定。但只要环境相同,一个作业经过多次运行,都会得到相同的结果。

1.3.2 操作系统的功能

操作系统作为计算机系统的资源管理器,其主要任务是对系统中的硬件、软件实施有效的管理,以提高系统资源的利用率。计算机硬件资源主要是指处理器、存储器、和外围设备;软件资源主要是指信息(文件系统)。因此,操作系统的主要功能相应地就有处理器管理、存储器管理、设备管理和文件管理。此外,为了方便用户使用操作系统,还需向用户提供一个使用方便的用户接口。

1. 处理器管理

CPU 是计算机系统中最宝贵的硬件资源。在多道程序环境下,要组织多个作业同时运行,就要解决处理器的管理问题。

处理器管理的主要任务是对处理器进行分配,并对其运行进行有效地控制和管理。在多道程序环境下,处理器的分配和运行都是以进程为基本单位,因而对处理器的管理可归结为对进程的管理。

处理器主要功能包括进程控制、进程同步、进程通信、进程调度。

2. 存储器管理

存储器可分为内存和外存两类,存储器管理主要是指对内存的管理。

存储器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,提高存储器的利用率,并能从逻辑上扩充内存。

存储器管理的主要功能有内存分配、内存保护、地址映射和内存扩充。

3. 设备管理

设备管理是对除了 CPU 和内存以外的所有计算机硬件资源的管理。

设备管理的主要任务是完成用户提出的 I/O 请求,为用户分配 I/O 设备,提高 CPU 与 I/O 设备的利用率,提高 I/O 设备的运行速度,方便用户使用 I/O 设备。

设备管理的主要功能有缓冲管理、设备分配、设备处理、设备独立性和虚拟设备。

4. 文件管理

在现代计算机系统中,总是把程序和数据以文件的形式存储在磁盘上,供所有的或指定的用户使用。为此,在操作系统中必须配置文件管理机构。

文件管理的主要任务是对用户文件和系统文件进行管理,方便用户使用,并保证文件的安全性。

文件管理的主要功能有文件存储空间管理、目录管理、文件读/写管理和存取控制。

5. 用户接口

为了方便用户使用操作系统,操作系统又向用户提供了“用户与操作系统的接口”。任何软件都需要提供给用户易用和美观的使用界面,即与用户之间的接口,或称用户接口,操作系统也不例外。在以往的操作系统中,用户接口通常仅有命令和系统调用两种形式,前者供用户在终端键盘上使用,后者供用户在编写程序时使用。而现代操作系统除了向用户提

供上述两种接口外,还提供图形接口。

(1) 命令接口

为了便于用户直接或间接地控制自己的作业,操作系统向用户提供了命令接口。用户可以通过命令接口向系统发出字符命令,及时与自己的作业交互,控制作业的运行。该接口又可进一步分为联机命令接口和脱机命令接口两种。

①联机命令接口。这是最常用的一种用户接口。该接口是为联机用户提供的,它由一组键盘命令和命令解释程序组成。每当用户在终端或控制台键盘上输入一条命令以后,系统便立即转入相应的命令解释程序,对该命令进行解释并执行。命令完成后又返回到终端或控制台上,等待用户输入下一条命令。

②脱机命令接口。该接口是为批处理作业的用户提供的,故也称为批处理用户接口。它由一组作业控制语言 JCL 组成。批处理系统的用户在向系统提交作业时,必须用作业控制语言把对作业进行的控制和干预事先写在作业说明书上,然后将作业连同说明书一起提交给系统。

(2) 程序接口

程序接口由一组系统调用命令组成,用户通过在程序中使用这些系统调用命令来请求操作系统提供服务。

①系统调用的概念。系统调用是操作系统提供给用户程序使用的具有一定功能的程序段。具体地讲,系统调用就是通过系统调用命令中断现行程序,而转去执行相应的子程序,以完成特定的系统功能。完成后,控制又返回到发出系统调用命令之后的下一条指令,被中断的程序将继续执行下去。

不同操作系统的系统调用命令的条数、格式和执行功能也不相同。系统调用命令扩充了机器指令,增强了系统的功能,方便了用户的使用。系统调用命令也称为广义指令。

系统调用的类型按功能大致分为设备管理、文件管理、进程管理、进程通信、存储管理 5 大类。

②系统调用的实现。在操作系统的内核中设置了一组专门用于实现各种系统功能的子程序,并将它们提供给用户程序调用。当用户在程序中需要这些功能时,便可以用一条系统调用命令,去调用程序所需要的系统功能。所以,系统调用在本质上是一种过程调用。

(3) 图形接口

用户虽然可以通过联机用户接口来取得操作系统的服务,并控制本地应用程序运行,但要求用户能熟记各种命令的名字和格式,并严格按照规定的格式输入命令,这既不方便又花时间。于是,图形用户接口应运而生。

近十年来软件界面越来越讲究易用性和美观性,操作系统传统的字符形命令接口也逐渐换成了图形用户接口。该接口采用图形化的操作界面,用非常容易识别的图标将系统的各种命令直观、逼真地表示出来。用户可通过鼠标、菜单和对话框来完成对应用程序和文件的操作。这样,用户就不需要记忆那些操作系统命令以及调用它们的格式,达到了易用的效果,深受初学者的喜爱,但这是以牺牲系统资源和性能为代价的。

20 世纪 90 年代后推出的操作系统一般都采用图形用户接口。如我们熟知的 Windows 系列操作系统等。