

操作系统



了解我国自主研发的 operating 系统的整体情况,激发学生投身国产操作系统研发的爱国情怀,培养学生大国工匠精神以及为国奉献的优良品质。



- (1) 掌握操作系统的概念与主要功能;
- (2) 理解操作系统的目标和管理资源的方法;
- (3) 了解操作系统的形成和发展,分时系统和实时系统的特点,以及操作系统的安全和维护等。

操作系统(Operating System, OS)是配置在计算机硬件上的第一层软件,只在核心态下运行,受硬件保护,并密切地依赖于计算机硬件。它是整个计算机系统的控制管理中心,对内界面是管理和控制硬件资源(包括 CPU、内存和外设),对外界面是为用户提供方便服务的一组软件程序集合,在用户和计算机之间起着接口作用。计算机科学与技术发展到今天,所有计算机,从个人计算机到巨型计算机,乃至超级计算机,都无一例外地配置了一种或多种操作系统。那什么是操作系统? 操作系统经历了怎样的发展历程? 操作系统的功能、管理资源的方法等,将在本章进行简要叙述。

3.1 操作系统的定义和目标

计算机系统是一个复杂的系统,由硬件(子)系统和软件(子)系统两部分组成。其中,硬件系统是整个计算机系统的物质基础,而软件系统则是计算机系统的灵魂。通常,把没有配置软件系统的计算机称为裸机,而配置了软件的计算机则称为虚拟计算机。

正如人不能没有大脑一样,操作系统是现代计算机系统中不可缺少的关键部分,其他的诸如编译软件、数据库管理软件等系统软件以及大量的应用软件,都直接依赖于操作系统的支持,并取得它所提供的服务。因此,操作系统已成为现代计算机系统、多处理机系统、以及计算机网络等都必须配置的系统软件。

3.1.1 操作系统的定义

对于操作系统,至今尚无严格的定义,人们可以从不同的角度来看待它。

从功能角度,即从操作系统所具有的功能来看,操作系统是一个计算机资源管理系统,

负责对计算机的全部资源(包括硬件资源和软件资源)进行分配、控制、调度和回收。

从资源管理角度,即从机器管理控制来看,操作系统是计算机工作流程的自动而高效的组织者,计算机软硬件资源合理而协调的管理者。利用操作系统,可减少人工干预,提高计算机的利用率。

从用户角度,即从用户使用来看,操作系统是一台比裸机功能更强、服务质量更高,用户使用更方便、更灵活的虚拟机,即操作系统是用户和计算机之间的接口(或界面)。

从软件角度,即从软件范围静态地看,操作系统是一种系统软件,是由控制和管理系统运转的程序和数据结构等内容构成。

综上,这里给出操作系统的定义:操作系统是一组能有效地组织和管理计算机的硬件和软件资源,合理地组织计算机工作流程,控制程序的执行,以及方便用户使用的程序的集合。

其中:

“有效”主要指操作系统在资源管理方面要考虑到资源的利用率和系统运行效率,要尽可能提高处理机的利用率,让其尽可能少地空转,而其他的资源,如内存、硬盘等,则应该在保证访问效能的前提下尽可能减少空间的浪费。

“合理”主要指操作系统对于不同用户程序要“公平”,以保证系统不发生“死锁”或“饥饿”的现象。

“方便”主要指人机界面方面,包括用户使用界面和程序设计接口两方面的易用性、易学性和易维护性。

3.1.2 操作系统的目标

目前,存在多种类型的操作系统,不同类型的操作系统的目标侧重点不同。在计算机系统上配置的操作系统,其主要目标有以下几点:

(1) 方便性:配置了操作系统的计算机系统,能够给用户提供一个清晰、简洁、易于使用的用户界面,从而使计算机更易于使用。

(2) 有效性:有效性的第一层含义是提高系统资源的利用率。在未配置操作系统的计算机系统中,诸如处理机、I/O 设备等各类资源经常会处于空闲状态而得不到充分利用;内存及外存中存放的数据由于无序而浪费了存储空间。因此,提高系统资源利用率是当时操作系统设计的一个主要目标。有效性的第二层含义是提高系统的吞吐量。操作系统可以通过合理组织计算机工作流程,加速程序的运行,缩短程序的运行周期来提高系统的吞吐量。

(3) 可扩充性:随着超大规模集成(Very Large Scale Integration, VLSI)电路技术和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快速发展,也因此对操作系统提出了更高的功能和性能要求。为适应计算机硬件、体系结构以及应用发展的要求,操作系统必须具有良好的可扩充性。可扩充性与操作系统的结构有着十分紧密的联系,由此推动了操作系统结构的不断发展。

(4) 开放性:20 世纪 80 年代和 90 年代陆续出现了各种类型的计算机硬件系统。尤其是近年来,随着互联网的快速发展,计算机操作系统的应用环境也由单机环境转向了网络环境。为了使出自不同厂家的计算机及其设备通过网络加以集成化并能正确、有效地协同工作,实现应用程序的可移植性和互操作性,就要求系统具有统一、开放的环境,其中首先

要求操作系统具有开放性。

在上述的四个目标中,方便性和有效性是操作系统设计时最重要的两个目标。在过去很长的一段时间内,由于计算机十分昂贵,有效性显得非常重要。近十几年来,随着硬件越来越便宜,在设计配置在微机上的操作系统时,如何提高用户使用计算机的方便性则更加被重视。

3.2 操作系统的发展

了解操作系统的发展历史,有助于理解操作系统的关键性设计需求,也有助于理解现代操作系统基本特征的含义。

3.2.1 操作系统的形成与发展

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;20 世纪 60 年代中期开发出多道程序批处理系统;不久又推出分时系统;与此同时,用于工业和武器控制的实时操作系统也相继问世。20 世纪 70 年代到 90 年代,是超大规模集成电路和计算机体系结构大发展的年代,导致了微型机、多处理机和计算机网络的诞生和发展,与此相应地,也相继开发出了微机操作系统、多处理机操作系统和网络操作系统,并得到极为迅猛地发展。

可见,操作系统的发展演变是伴随计算机技术的进步而不断发展变化的,它的发展变化与计算机组成和体系结构密切相关。

1. 无操作系统的计算机系统

1) 人工操作方式

从 1946 年诞生的第一台计算机,到 20 世纪 50 年代中期的计算机,都属于第一代计算机。这时还未出现操作系统,对计算机的全部操作都是由用户(即程序员)采取人工操作方式进行。即由程序员将事先已穿孔(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),通过纸带输入机将程序和数据输入计算机,然后启动计算机运行。当程序运行完毕并取走计算结果后,才让下一个用户上机。这种人工操作方式存在以下两个缺点:

(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 和内存等资源是空闲的。

人工操作方式严重降低了计算机资源的利用率,即所谓的人机矛盾。随着处理机速度的提高、系统规模的扩大,人机矛盾也就变得日趋严重。与此同时,CPU 和 I/O 设备之间速度不匹配的矛盾(I/O 设备的速度提高缓慢)也更加突出。为此,先后出现了通道技术、缓冲技术,但都未很好地解决上述矛盾,直到后来引入了脱机 I/O 技术,才获得了相对满意的结果。

2) 脱机 I/O 方式

如上所述,为了解决人机矛盾以及快速 CPU 和慢速 I/O 设备之间的速度不匹配矛盾,20 世纪 50 年代末出现了脱机 I/O 技术。该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,然后在—台外围机的控制下把纸带(或卡片)上的数据(程序)输入磁带(盘)上;当 CPU 需要这些程序和数据时再从磁带(盘)上高速地调入内存。类似地,当 CPU 需要输出时,可先由 CPU 直接高速地把数据从内存送到磁带(盘)上,然后在另一台外围机的控制下,将磁带(盘)上的结果通过相应的输出设备输出。

图 3.1 给出了脱机 I/O 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进行,或者说它们是在脱离主机的情况下进行的,故称为脱机 I/O 方式。反之,在主机直接控制下进行的 I/O 方式则称为联机 I/O 方式。

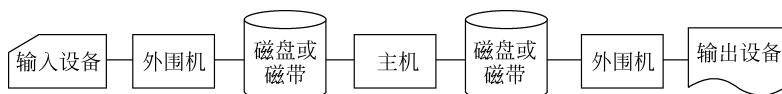


图 3.1 脱机 I/O 过程

这种脱机 I/O 方式的优点主要体现在以下两方面:

(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 设备送到高速的磁(盘)带上(或相反)的操作,都是在脱机情况下进行的,它们不占用主机时间,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(盘)带上将数据输入内存的,从而大大缓和了 CPU 和 I/O 设备速率不匹配的矛盾,并进一步减少了 CPU 的空闲时间。

2. 单道批处理系统

20 世纪 50 年代末期出现了第二代晶体管计算机,此时计算机虽有推广应用价值,但计算机系统仍非常昂贵。为了充分利用计算机系统,就应尽可能保持系统的连续运行,即在处理完一道作业后,紧接着处理下一道作业,以减少机器的空闲等待时间,以此便形成了早期的单道批处理系统。

1) 单道批处理系统的处理过程

为实现对作业的连续处理,需要先把一批作业以脱机方式输入磁带(盘)上,同时在系统中配置监督程序(Monitor),并在它的控制下,使这批作业一道接一道地连续处理。其处理过程如下:首先由监督程序将磁带(盘)上的第一道作业装入内存,并把运行控制权交给该作业;当该作业处理完成时,又把控制权交还给监督程序,再由监督程序把磁带(盘)上的第二道作业装入内存,以此类推。计算机系统就是这样自动地一道作业接一道作业地进行处理,直到磁带(盘)上的作业全部完成。

由于系统对作业的处理都是成批进行的,且在内存中任意时刻始终只保持一道作业,故称单道批处理系统。

2) 单道批处理系统的特征

单道批处理系统是最早出现的一种操作系统。但从严格意义上来讲,它只能算作操作系统的前身而并非现在人们所理解的操作系统。尽管如此,与人工操作方式相比该系统已是很大的进步。其主要特征包括:

(1) 自动性。顺利的情况下,磁带(盘)上的一批作业能自动地逐个依次运行,而无须人工干预。

(2) 顺序性。磁带(盘)上的作业按序进入内存,正常情况下各道作业完成的顺序与它们进入内存的顺序完全相同,即先进入内存的作业先运行。

(3) 单道性。任意时刻内存中仅有一道作业并使之运行。

可见,单道批处理系统是在解决人机矛盾和 CPU 与 I/O 设备速度不匹配矛盾的过程中形成的,即旨在提高系统资源的利用率和系统的吞吐量。但单道批处理系统仍没有很好地利用系统资源。主要原因在于 CPU 和 I/O 设备是串行运行的,即在程序进行 I/O 时,CPU 只能等待。也就是说,一个程序在进行 I/O 时,CPU 需要不断地探询 I/O 是否完成,因而不能执行其他程序,这是很大的浪费。一种有效的方法就是将 CPU 和 I/O 设备进行并发,即一个程序在 I/O 时,让另一个程序继续运行。换言之,将 CPU 运行和 I/O 设备的运行重叠起来以改善整个系统的效率。

3. 多道批处理系统

20 世纪 60 年代中期,IBM 公司生产了第一台小规模集成电路计算机 IBM 360(第三代计算机系统)。由于它较之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改善,因而获得了极大的成功。此时,为了进一步提高系统资源的利用率和系统吞吐量,引入了多道程序设计技术,即同时将几道作业放入内存,并使几道作业同时运行,且共享系统中的资源,由此形成了多道批处理系统。

1) 多道批处理系统的处理过程

在多道批处理系统中,用户(程序员)所提交的作业先存放在外存上,并排成一队列,称为“后备队列”。然后由作业调度程序按照一定的算法,从后备队列中选择若干作业调入内存,使它们共享 CPU 和系统中的各种资源,以达到提高资源利用率和系统吞吐量的目的。

2) 多道批处理系统的特征

(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,有效地提高了资源利用率,增加了系统吞吐量。

(2) 无序性。多道作业完成的先后顺序与它们进入内存的顺序之间并无严格的对应关系,即先进入内存的作业可能较后甚至最后完成,而后进入内存的作业可能先完成。

(3) 调度性。作业从提交给系统开始直至完成,需要经过作业调度和进程调度两级调度。

3) 多道批处理系统的优缺点

(1) 资源利用率高。由于内存中装入了多道程序,它们共享系统资源,从而使资源能保持处于忙碌状态,最终使各种资源得以充分利用。

(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任务总量。系统吞吐量大的原因可归结为两方面:一方面,CPU 和其他资源保持“忙碌”状态;另一方面,仅当作业完成或运行不下去时才进行 CPU 切换,系统开销小。

(3) 平均周转时间长。作业的周转时间是指从作业进入系统(外存)开始,直至作业完成并退出系统所经历的时间。在多道批处理系统中,由于作业要排队依次进行处理,因而作业的平均周转时间较长,通常需要几十小时甚至几天。

(4) 无交互能力。从用户把作业提交给系统直至作业完成,用户都不能与自己的作业进行交互,这对修改和调试程序极不方便。

在多道批处理系统时代,由于多道作业同时运行,系统不仅需要在多道作业之间进行切换,并且还要保护一道作业不被另一道作业所干扰,同时还要管理多个 I/O 设备。显然,与单道批处理系统相比,多道批处理系统是一种十分有效但又非常复杂的系统,它既要管理作业,又要管理内存和 I/O 设备,还要管理 CPU 的调度,同时还应提供用户和操作系统接口,以方便用户使用操作系统。

为此,应在计算机系统中增加一组软件以对上述问题进行妥善、有效地处理。这样的一组软件应包括:能有效地组织和管理四大资源的软件、合理地各类作业进行调度并控制它们运行的软件,以及能方便用户使用计算机的软件。正是这样一组软件构成了操作系统,故多道批处理操作系统才是真正意义上的操作系统。

4. 分时系统

1) 分时系统的引入

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满足用户的需求。具体地说,用户的需求主要表现在以下几方面。

(1) 人-机交互。对程序员来说,每当编写好一个程序时,都需要上机进行调试。由于新编写的程序难免存在一些错误或不当之处需要进行修改完善,因而用户希望能像早期使用计算机一样,独占全机并对它进行直接控制,以便能方便地修改错误,即用户(程序员)希望能进行人-机交互。

(2) 共享主机。20 世纪 60 年代,计算机仍然昂贵,一台计算机要同时供很多用户共享使用。用户在使用计算机时希望能够像自己独占计算机一样,不仅可以随时与计算机交互,而且还感觉不到其他用户的存在。

(3) 便于上机。用户希望能通过自己的终端直接将作业传送到计算机上进行处理,并能对自己的作业进行控制。

分时系统正是为了满足用户的上述需求而形成的一种新型操作系统。它是指在一台主机上连接了多个终端,每个联机用户通过自己的终端以交互的方式控制程序的运行,共享主机中的资源。此时,系统把 CPU 时间轮流分配给各联机作业,每个作业只运行极短的一个时间片,而每个用户都有一种“独占计算机”的感觉。

2) 分时系统的特征

(1) 多路性。又称同时性。允许在一台主机上同时连接多个终端,系统按分时的原则为每个用户服务。

(2) 独立性。每个用户在各自的终端上进行操作,彼此之间互不干扰。因此,用户会有一种一人独占主机的感觉。

(3) 及时性。用户的请求能在很短时间内获得响应。此时间间隔通常以人们所能接受的等待时间来确定,一般为 2~3 秒。

(4) 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各方面的服务,如文件编辑、数据处理、请求打印服务等。

5. 实时系统

1) 实时系统的引入

虽然多道批处理系统和分时系统已经获得了令人较为满意的资源利用率、系统吞吐

量,以及较好的人机交互,但随着计算机应用范围的扩大,它们已不能满足以下两个领域的需要。

(1) 实时过程控制领域。实时过程控制又可分为两类:一类是以计算机为控制中枢的生产过程自动化系统,如冶金、发电、化工、机械加工等自动控制。在这类系统中,要求计算机能及时采集现场数据,并对所采集的数据进行及时处理,进而自动地控制相应的执行装置,使得某些参数,如温度、压力、液位等能按一定规律变化,从而达到生产过程自动化的目的。另一类是飞行物体的自动控制,如飞机、导弹、人造卫星的制导等。这类系统要求反应速度快、可靠性高,通常要求系统的响应时间在毫秒甚至微秒级内。

(2) 实时信息处理领域。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主机通过通信线路连接成百上千个远程终端,主机接收从远程终端发来的服务请求,并根据用户提出的问题对信息进行检索和处理,同时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订票系统、情报检索系统、银行业务等。

可见,上述两个领域对计算机响应时间有着严格的要求。通常把对计算机响应时间有要求的系统称为临界系统或临界应用。为了满足这些应用对响应时间的要求,人们就开发了实时操作系统。所谓“实时”是表示“及时”“即时”。实时操作系统是指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运行的操作系统,即必须满足时序可预测性(Timing Predictability)。

实时系统需要保证响应时间的可预测性和稳定性,通常又分为软实时系统和硬实时系统。软实时系统要求任务在规定的时间内尽快完成,但可以容忍一定的延迟,即在规定时间内得不到响应所产生的后果是可以承受的。例如机械加工的流水装配线,即使流水线瘫痪,也仅是产生经济损失。而硬实时系统要求任务必须在规定的时间内完成,否则可能产生不能承受的灾难。如导弹防御系统,如果反应迟缓,导致的后果可能是以生命为代价。

需要注意的是,实时系统并不是反应很快的系统,而是反应具有时序可预测性的系统。当然,在实际中,实时系统通常是反应很快的系统,但这是实时系统的一个结果,而不是其定义。

2) 实时系统的特点

- (1) 对外部进入系统的信号或信息能做到及时响应。
- (2) 实时系统较一般的通用系统有规律,许多操作具有一定的可预计性。
- (3) 实时系统的终端一般作为执行和询问使用,不具有分时系统的较强会话能力。
- (4) 实时系统对可靠性和安全性要求较高,常采用双工工作方式。

可见,实时操作系统的最重要部分就是进程或工作调度。只有精确、合理和及时的进程调度才能保证响应时间。当然,对资源的有效管理也非常重要,没有精确复杂的资源管理,确保进程按时完成就成为一句空话。另外,基于其应用环境,实时操作系统对可靠性和可用性要求也非常高。如果在这些方面出了问题,时序可预测性将无法达到。

3) 实时系统与分时系统的比较

(1) 多路性。实时系统和分时系统一样具有多路性。对实时系统而言,多路性主要表现在周期性地对多路的现场信息进行采集以及对多个对象或多个执行机构进行控制。

(2) 独立性。实时系统和分时系统一样具有独立性。每个终端用户在向实时系统提出

服务请求时,彼此独立操作,互不干扰;同样在实时系统中信息的采集和对对象的控制,也都是彼此互不干扰。

(3) 及时性。分时系统的及时性通常以人们能接受的等待时间来确定,单位通常为秒;实时系统的及时性则是以控制对象所要求的开始截止时间或完成截止时间来确定,一般为百毫秒级、毫秒级,甚至为百微秒级。

(4) 交互性。实时系统虽然也具有交互性,但这里所指的人与系统的交互并不像分时系统那样能向终端用户提供数据处理、资源共享服务,而是仅限于访问系统中某些特定的专用服务程序,或由用户发送某些特定的命令,如开始、停止等,然后由系统立即响应。

(5) 可靠性。分时系统虽然也要求系统可靠,但相比之下,实时系统则要求系统高度可靠。因为任何差错都可能带来巨大的经济损失,甚至无法预料的灾难性后果。故实时系统一般都采用了多级容错技术,以保证系统的安全和数据的安全。

在操作系统发展过程中,多道批处理系统、分时系统和实时系统是操作系统的三种基本类型。但一个实际系统往往兼有它们三者或其中两者的功能,这就出现了通用操作系统。例如,批处理系统与分时系统相结合,当系统有分时用户时,系统及时响应他们的请求,而当系统没有分时用户或分时用户较少时,系统处理不太紧急的批处理作业,以便提高系统资源的利用率。在这种系统中,把分时作业称为前台作业,批处理作业称为后台作业。类似地,批处理系统和实时系统相结合,有实时请求则及时进行处理,无实时请求则进行批处理。运行时把实时作业称为前台作业,批处理作业则称为后台作业。

6. 现代操作系统

在 20 世纪 80 年代后期,计算机工业取得了井喷式发展,各种新型计算机和新型操作系统不断出现和发展,计算机和计算机操作系统领域均进入了一个百花齐放、百家争鸣的时代。

随着 VLSI 和计算机体系结构的发展,以及应用需求的不断扩大,操作系统仍在继续发展,先后形成了微机操作系统、多处理机操作系统、网络操作系统、分布式操作系统以及嵌入式操作系统。

微机操作系统可按微机的字长分成 8 位、16 位和 32 位的微机操作系统,也可按运行方式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。单用户单任务操作系统的含义是,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。最具代表性的单用户单任务操作系统是 CM/P 和 MS-DOS。单用户多任务的含义是,只允许一个用户上机,但允许将一个用户程序分成若干任务,使它们并发执行,从而有效地改善系统的性能。目前,32 位微机上配置的 32 位微机操作系统大多数是单用户多任务操作系统,其中最有代表性的是 OS/2、MS Windows 和 Linux。多用户多任务操作系统的含义是,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进一步分成几个任务,使它们并发执行,从而可进一步提高资源利用率和增加系统吞吐量。在大、中、小型机中配置的都是多用户多任务操作系统;在 32 位微机上,也有不少配置的是多用户多任务操作系统。其中,最有代表性的是 UNIX 操作系统。

多处理机系统(Multi-processor System, MPS)是从计算机体系结构上改善系统性能。

其引入的原因有三点：一是增加系统的吞吐量；二是节省投资；三是提高系统可靠性。在多处理机系统中所配置的多处理机操作系统，可以分为非对称多处理机模式（主-从模式）和对称多处理机模式两种模式。

网络操作系统(Network Operating System,NOS)是在计算机网络环境下,对网络资源进行管理和控制,实现数据通信以及对网络资源的共享,并为用户提供与网络资源之间接口的一组软件和规程的集合。网络操作系统旨在通过提供网络服务和协议的支持来使得多台计算机能够相互通信和共享资源。它不仅提供了常见操作系统的功能,如文件系统管理、进程调度、设备驱动程序等,还增加了网络相关的功能,例如网络协议栈、路由器功能、防火墙等。一般而言,网络操作系统具有硬件独立性、接口一致性、资源透明性、系统可靠性和执行并行性 5 个特征。

分布式操作系统(Distributed Operating System,DOS)是配置在分布式系统上的公用操作系统。它以全局的方式对分布式系统中的所有资源进行统一管理,可以直接对系统中地理位置分散的各种物理和逻辑资源进行动态地分配和调度,有效地协调和控制各个任务的并行执行,协调和保持系统内的各计算机间的信息传输及协作运行,并向用户提供一个统一、方便、透明的使用系统的界面和标准接口。一个典型的例子就是万维网,在万维网中,所有操作只通过一种界面(即网页)进行。与网络操作系统不同,分布式操作系统用户在使用系统内的资源时,不需要了解诸如网络中各计算机的功能与配置、操作系统的差异、软件资源、网络文件的结构、网络设备的地址、远程访问的方式等等情况,对用户屏蔽了系统内部实现的细节。分布式操作系统保持了网络操作系统所拥有的全部功能,同时又具有透明性、可扩展性、容错性以及并发性等特点。常见的分布式操作系统包括 Google 的分布式操作系统 GFS(Google File System)、Apache 的分布式文件系统 Hadoop、微软的 Azure Service Fabric 等。我国华为公司所研发的鸿蒙系统(HarmonyOS)是一个面向未来、面向全场景(包括移动办公、社交通信、媒体娱乐等)的分布式操作系统。

嵌入式操作系统(Embedded Operating System,EOS)是用于嵌入式系统的系统软件,负责嵌入系统的全部软硬件资源的分配、任务的调度工作、控制和协调并发活动。它必须体现其所在系统的特征,能够通过装卸某些模块来达到系统所要求的功能,在系统实时高效性、硬件的相关依赖性、软件固化性以及应用的专业性等方面具有较为突出的特点。EOS 过去主要应用于工业控制领域和国防系统领域,随着 Internet 技术的发展、信息家电的普及,以及 EOS 的微型化和专业化,EOS 开始从单一的弱功能向高专业化的强功能方向发展。嵌入式操作系统除了具备一般操作系统最基本的功能,如任务调度、同步机制、中断处理、文件等功能外,还具有可装卸性、强实时性、统一的接口、提供强大的网络功能、强稳定性、弱交互性、固化代码、良好的移植性以及操作方便、简单等特点。

3.2.2 我国操作系统的发展

国内操作系统的发展要追溯到 2001 年,彼时国家力量联合产业界与学界(国防科大、中软、联想、浪潮、民族恒星)共同推出了最早的商业闭源操作系统——麒麟操作系统(Kylin OS),打响了国产操作系统的第一枪。

此后,在开放平台生态不断成熟的背景下,国产操作系统凭借着开放平台生态和国家支持的东风,快速崛起。发展到今天,国内有基于 Linux 开源做的操作系统,如中科方德、

优麒麟、银河麒麟、中兴新支点、统信 UOS、Deepin(原名 Linux Deepin,中文通称深度操作系统)、一铭 Linux、思普等操作系统,以及基于微内核的华为鸿蒙操作系统(Harmony OS)。

深度操作系统由武汉深之度公司研发,是基于 Linux 内核,以桌面应用为主的开源操作系统,支持笔记本、台式机和一体机。它是中国第一个具备国际影响力的 Linux 发行版本,也是在全球开源操作系统排行榜上排名最高的中国操作系统产品。Deepin 注重桌面环境的美观和易用性,以提供直观、舒适和流畅的用户体验为目标,广受用户欢迎。它采用了自主开发的深度桌面环境(Deepin Desktop Environment),具有独特的设计风格和功能特点;适用于个人用户、学生、开发者以及对桌面环境有高要求的用户群体。同时,Deepin 也积极参与开源社区,与其他 Linux 发行版进行合作和贡献,推动了开源软件的发展。

统信 UOS 是由统信软件开发的一款操作系统,支持龙芯、飞腾、兆芯、海光、鲲鹏等国产品片平台的笔记本、台式机、一体机、工作站、服务器,以桌面应用场景为主,包含自主研发的桌面环境、多款原创应用,以及丰富的应用商店和互联网软件仓库,可满足用户的日常办公和娱乐需求。

华为鸿蒙操作系统是第一款基于微内核的全场景分布式 OS,由华为自主研发,是目前国内用户量最大,业内关注度最高的物联网操作系统。鸿蒙有三层架构,第一层是内核,第二层是基础服务,第三层是程序框架。作为面向全场景的物联网 OS,它拥有低延时、可扩展性强和安全性高的优点,可适配不同的硬件终端,打通了 PC、手机、物联网等之间的壁垒。

国产基于 Linux 开源的操作系统虽是基于 Linux 内核进行的二次开发,但在功能上却做了很多改进。目前,国产操作系统已经实现了从基础软件到应用软件的全链条覆盖,包括桌面操作系统、服务器操作系统、移动操作系统等各类操作系统。其中,麒麟操作系统、红旗操作系统、深度操作系统等国产操作系统已经在国内市场得到广泛应用。另外,随着人工智能和大数据等技术的发展,我国国产操作系统也在这些领域加速布局,如麒麟 OS 和昇腾 OS 等面向 AI 领域的操作系统。

国家安全是民族复兴的根基,网络安全是国家安全的重要组成部分,操作系统作为信息系统的基石,掌握着网络安全牢固的“开关键”。因此,我们应该不忘初心、牢记使命,以党的二十大精神为指引,奋力为国产操作系统事业贡献力量。坚信随着国家对于信息技术产业的重视和支持,我国国产操作系统的发展前景将十分广阔。

3.2.3 推动操作系统发展的主要动力

自 20 世纪 50 年代操作系统诞生以来,其经历了由简单到复杂、由低级到高级的发展。在这个发展历程中,操作系统在各方面都有了长足的进步,不仅能很好地适应计算机硬件和体系结构的快速发展,并且也能满足不断变化的应用需求。操作系统之所以能在短短的几十年中取得如此巨大的发展,其主要推动力可归纳为以下几方面。

1. 不断提高计算机系统资源的利用率

在计算机发展的初期,计算机系统非常昂贵,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,这就成为推动操作系统发展的最初动力。由此形成了能自动对一批作业进行处理的批处理系统,其通过减少计算机的空闲时间提高了系统中央处理器(Central

Processing Unit,CPU)和 I/O 设备的利用率。

20 世纪 60 年代,出现了可以支持多个用户使用同一台计算机的分时系统,使系统资源率得到了极大的提高,并推动了计算机的第一次大普及。与此同时,还推出了能改善 I/O 设备和 CPU 利用率的假脱机系统。

20 世纪 70 年代,提出了能够有效提高存储器系统利用率并能从逻辑上扩大内存的虚拟存储器技术。此后在网络环境下,通过在服务器上配置网络文件系统和数据库系统,将资源提供给全网用户使用,又进一步提高了资源的利用率。

2. 方便用户

当资源利用率不高的问题得到基本解决之后,用户(主要是程序员)在上机、调试程序时的不方便性便成为主要矛盾。于是人们又想方设法改善用户上机、调试程序的条件,这成为推动操作系统继续发展的主要因素,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。分时系统的出现,不仅提高了系统资源的利用率,还能使用户像早期使用计算机时一样,感觉自己独占全机资源,可以对其进行直接操控,同时也极大地方便了程序员对程序进行调试。20 世纪 90 年代初,图形用户界面的出现进一步方便了用户对计算机的使用,从而受到用户的广泛欢迎,这无疑又推动了计算机的迅速普及和广泛应用。

3. 器件的不断更新换代

信息技术(Information Technology,IT)的迅猛发展,计算机器件在不断地更新换代,由第一代的电子管发展到第二代的晶体管、第三代的集成电路,再到第四代的大规模和超大规模集成电路,使得计算机的性能不断提高,规模也在急剧扩大,从而推动着操作系统的功能和性能的不断发展。例如,当微机芯片由 8 位发展到 16 位,进而发展到 32 位甚至 64 位,相应的微机操作系统也就由 8 位发展到 16 位、32 位进而发展到 64 位,此时操作系统的功能和性能都有了显著的提高和增强。

与此同时,外部设备也在迅速发展,操作系统所能支持的外部设备的种类也越来越多。如现在的微机操作系统除了能支持传统外设外,还可以支持移动硬盘、闪存、数码相机、扫描仪等。

4. 计算机体系结构的不断发展

计算机体系结构的不断发展也是推动操作系统发展的一个因素,并催生了新的操作系统类型。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操作系统发展为多处理机操作系统。又如,当出现了计算机网络之后,配置在计算机网络上的网络操作系统也就应运而生,其不仅可以有效管理网络中的共享资源,还能向用户提供很多网络服务。

5. 不断提出新的应用需求

操作系统能如此迅速发展的另一个重要因素是,人们不断提出新的应用需求。例如,为了提高产品的数量和质量,需要将计算机应用到工业控制中,此时在计算机上就需要配置能进行实时控制的操作系统,由此产生了实时系统。再如,为了满足用户在计算机上播放视频、听音乐、玩游戏,以及从网络下载电影资源等需求,在操作系统中增加了多媒体功能。又如,超大规模集成电路的发展,CPU 体积越来越小,价格越来越便宜,大量智能设备

(特别是智能手机)应运而生,此时,嵌入式操作系统的产生和发展就成为必然。

另外,随着计算机技术应用的普及以及与信息技术的不断结合和发展,越来越多的重要信息都需要经过计算机系统进行加工、处理并存储,因此确保系统的安全性也就成为操作系统必须具备的功能。除此以外,操作系统本身的设计还应该遵循安全系统设计原则,从而提高整个计算机系统的安全性。

3.2.4 主流操作系统简介

1. Windows 操作系统

Microsoft Windows 是美国微软公司研发的一系列商业操作系统,用于支持台式机、笔记本电脑、平板电脑和服务器等设备。它采用了图形化用户界面,相比于 DOS 的字符界面更为人性化,是当今全球最广泛使用的个人计算机操作系统之一。

Windows 操作系统问世于 1985 年,随着计算机硬件和软件的不断升级,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 7、Windows 8、Windows 10。目前,Windows 操作系统的最新版本是 Windows 11。除了桌面版 Windows 之外,微软还开发了适用于服务器环境的 Windows Server 操作系统以及移动设备的 Windows 10 Mobile 操作系统(已停止更新)。

2. UNIX 操作系统

UNIX 操作系统最早是由肯·汤普森(Ken Thompson)和丹尼斯·里奇(Dennis Ritchie)于 1969 年在贝尔实验室开发。因其具有适应性和可变性,多年来 UNIX 不断发展,并伴随不同的需求以及新的计算机环境的变化而变化。它是一种通用、多用户、分时操作系统,可以出色地完成一般单用户多任务操作系统所能实现的功能,为用户提供较好的应用界面和软件开发环境。UNIX 是最早产生并商用化的操作系统之一,被认为是现代计算机操作系统的鼻祖,为后来的许多操作系统提供了基础和灵感。

UNIX 操作系统不需要和某一类型的计算机捆绑在一起,其可以轻松地适应各种类型的硬件而不损失标准特性。对工作站、微型计算机、大型机、甚至超级计算机等各种不同类型的计算机来讲是一种标准的操作系统,目前主要用于工程应用和科学计算等领域。

UNIX 操作系统诞生至今已经衍生出了许多不同的变种和分支,如 BSD、Linux 和 MacOS 等。这些变种继承了 UNIX 的设计原则和哲学,但也在特性和实现上有所差异。UNIX 的两个主流版本分别是著名的 System V 和 BSD UNIX。System V 是由 AT&T 贝尔实验室开发,而 BSD UNIX 是由加州伯克利分校研制。这两个版本的 UNIX 经过集成成为现在的系统 V 第四版,简称 SVR 4。

3. Linux 操作系统

Linux,全称 GNU/Linux,是一种自由和开源的操作系统内核,最早由芬兰计算机科学家林纳斯·托瓦兹(Linus Torvalds)于 1991 年开发。它是类 UNIX 操作系统的一个分支,由于其源码开放,在加入 FSF(the Free Software Foundation,自由软件基金会)后,经过互联网上所有开发人员的共同努力,已经成为能够支持各种体系结构、世界上使用最多的一种 UNIX 类操作系统。

Linux 操作系统的诞生、发展和成长过程始终依赖于 5 个重要支柱：UNIX 操作系统、MINIX 操作系统、GUN 计划、POSIX 标准和 Internet 网络。它不仅在高端工作站、服务器上表现出色，在网络环境下更具有优势。此外，它也同样适合作为桌面操作系统，有着良好的图形用户界面。

Linux 操作系统有多个发行版(Distribution)，如 Ubuntu、Debian、Fedora、CentOS 等，每个发行版都有自己的特点和目标用户群。

4. MacOS 操作系统

MacOS 是苹果公司为 Mac 系列产品开发的专属操作系统，是首个在商用领域成功的图形用户界面操作系统。MacOS 是全世界第一个基于 FreeBSD 系统采用“面向对象操作系统”的全面的操作系统，“面向对象操作系统”是史蒂夫·乔布斯于 1985 年被迫离开苹果公司后成立的 NeXT 公司所开发的。后来，苹果公司收购了 NeXT 公司，史蒂夫·乔布斯重新担任苹果公司 CEO，Mac 开始使用的 MacOS 系统得以整合到 NeXT 公司开发的 OpenStep 系统上。Mac OS 以简单易用和稳定可靠著称，全屏幕窗口是其最为重要的功能，一切应用程序均可以在全屏模式下运行。

5. Android 操作系统

Android 由 Google 开发的基于 Linux 内核的开源操作系统，主要用于移动设备，如智能手机、平板电脑和智能电视等。除了移动设备，Android 操作系统还扩展到了其他领域，如智能手表、智能家居和汽车娱乐系统等，已成为全球最广泛使用的移动操作系统之一。

Android 是一个开源项目，允许开发者自由地访问、修改和定制其代码；支持多任务处理，允许用户同时运行多个应用程序，并在屏幕上进行切换；拥有庞大的应用商店(Google Play Store)，其中包含数以百万计的应用程序，涵盖了各种类别，如社交媒体、游戏、办公工具和娱乐等。此外，Android 操作系统能在不同制造商的设备上运行，并具有广泛的硬件兼容性。

3.3 操作系统的功能和启动

3.3.1 操作系统的功能

操作系统是计算机系统软件的核心，它在计算机系统中担负着管理系统资源、控制输入输出处理，以及实现用户和计算机系统间通信的重要任务。若从资源管理和人机交互的角度来理解操作系统的功能，可把操作系统的功能分为资源管理功能和扩展的虚拟机功能。

1. 资源管理功能

操作系统的资源管理功能在计算机系统中扮演着非常重要的角色，其资源管理功能的实现使得整个计算机系统变得轻便简洁。计算机系统的资源包括硬件资源和软件资源。其中，硬件资源包括处理机、存储器、输入/输出设备等；软件资源则是指以文件形式存放在计算机中的信息，包括存放在内存的表、数据结构和系统例程等。系统的硬件资源和软件资源都由操作系统根据用户需求按一定的策略分配和调度。故从资源管理的角度来看，操

作系统的功能主要包括以下 4 方面。

1) 处理机管理

处理机管理主要是组织和协调用户对处理机的使用,管理和控制用户任务,并以最大限度提高处理机的利用率。在单道作业或单用户的环境下,处理机被一道作业或一个用户所独占,对处理机的管理十分简单。但在多道程序或多用户环境下,系统内通常会有多个任务竞争使用处理机,这就需要系统对处理机进行调度。

处理机作为计算机系统的关键资源,其管理策略和调度算法会直接影响整个系统的运行效率。在传统的多道程序环境下,处理机的分配和运行都是以进程为基本单位,因此对处理机的管理可归纳为对进程的管理。主要包括以下几方面。

(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。其任务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。为使多个进程能有条不紊地运行,系统必须设置相应的进程同步机制,以协调系统中各进程的运行。常用的协调方式有进程互斥方式和进程同步方式。

(4) 进程通信。系统中的各进程之间有时需要合作,往往要交换信息,为此需要进行通信。

2) 存储管理

存储管理主要指主存管理,其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。因此,存储管理不仅要为在系统运行的每个任务分配内存,还要完成内存回收、地址映射、内存保护,以及存储扩充的功能。存储管理方案的主要目的是解决多个用户使用主存的问题,其存储管理方案包括分区存储管理、分页存储管理、分段存储管理、段页式存储管理以及虚拟存储管理。

3) 设备管理

这里的设备是指除了 CPU 和内存以外的各种设备,如打印机、磁带、磁盘、终端等。这些设备种类繁多,物理性能各不相同,如有些设备可以共享,而有些设备则只能独占。因此,在多用户多任务环境下,为了提高这些设备的利用率,实现资源的有效共享,操作系统的设备管理往往很复杂。

设备管理的主要任务包括:①完成用户进程提出的 I/O 请求,为用户进程分配所需的 I/O 设备,并完成指定的 I/O 操作;②提高 CPU 和 I/O 设备的利用率,提高 I/O 速度,方便用户使用 I/O 设备。

为实现上述任务,设备管理应具有设备分配、设备处理、缓冲管理以及虚拟设备等功能。其中,缓冲管理是为了缓和 CPU 和 I/O 设备的速度差,提高 CPU 的利用率,进而提高系统吞吐量。在现代操作系统中,无一例外地在设备管理中建立 I/O 缓冲区,而且还可以通过增加缓冲区容量的方法来改善系统的性能。不同的系统可以采用不同的缓冲区机制。虚拟设备的功能是将低速的独占设备改造成高速的共享设备。

4) 文件管理

在计算机系统中,系统本身有很多程序,用户又有很多程序和数据,它们都以文件的形式进行组织,且大部分的文件平常都存放在外存上。

操作系统的文件管理功能主要是对文件进行逻辑组织,安排文件在外存的物理存储,并对外存空间进行管理。由于文件是用户频繁使用的信息,故文件管理还应解决用户操作文件的一系列问题,如文件的创建、文件打开、文件关闭、文件的读写操作等。当系统中文件数目较多时,文件管理还应该能按一定的方式将文件组织起来,如采用树形目录结构,并提供文件的高效存取访问功能。为了方便用户使用文件和有利于用户间的数据共享,文件管理系统一般都支持按名访问方式,用户无须了解文件的物理结构和确切的物理位置。

2. 扩展的虚拟机功能

操作系统作为虚拟机,屏蔽了计算机系统的硬件,隐藏了计算机硬件的操作细节,提供给用户简单使用系统或请求系统服务的接口。换言之,操作系统为用户提供了友好的人机交互以及程序级接口,用户通过使用这些接口达到了方便使用计算机的目的,而计算机看上去则像是一台功能得到了扩展的机器。

操作系统为用户提供的接口通常可分为如下两大类。

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。用户可以通过该接口向作业发出命令以控制作业的运行。该接口又进一步分为联机用户接口、脱机用户接口和图形用户接口三种。

其中,联机用户接口是为联机用户提供的,由一组键盘操作命令及命令解释程序组成。脱机用户接口是为批处理作业的用户提供的。图形用户接口采用了图形化的操作界面,用非常容易识别的各种图标将系统的各项功能、各种应用程序和文件直观地表示出来,用户可以通过鼠标、菜单和对话框来完成各种应用程序和文件的操作。

2) 程序接口

程序接口是为用户程序在执行中访问系统资源而设置的,是用户程序取得操作系统服务的唯一途径。它由一组系统调用组成,每个系统调用都是一个能完成特定功能的子程序。

早期的系统调用都是用汇编语言编写的,只有在用汇编语言书写的程序中才能直接使用系统调用。但在C语言等高级语言中,往往提供了与各系统调用一一对应的库函数,这样,应用程序便可通过调用相应的库函数使用系统调用。近几年提供的操作系统中,如UNIX、OS/2版本中,其系统调用本身就是用C语言编写的,并以函数形式提供,故在C语言编制的程序中,可直接使用系统调用。

随着计算机技术的进步和网络的普及,对于操作系统也提出了更多的功能要求,如要求操作系统要有面向安全、面向网络和面向多媒体等功能。因此,现代操作系统在传统操作系统功能基础上均增加了保障系统安全、支持用户通过联网获取服务和可处理多媒体信息等功能。

3.3.2 操作系统的启动

操作系统的启动是由引导程序来完成的。这个引导程序就是通常我们所说的BIOS(Basic Input Output System,基本输入/输出系统)。它是计算机启动后第一个执行的程序,被固化在ROM中,可以在每次机器加电后自动执行,主要完成上电自检、硬件初始化以及引导操作系统。

当系统开机时,CPU直接跳转到ROM的固定地址执行系统引导程序,引导程序主要

负责自检,并从硬盘、软盘或光盘中预先确定的位置将分区引导块读入内存,由引导块对操作系统进行引导;引导块将操作系统引导进入系统内存后,就跳转到操作系统中运行,由此将系统的控制权交给操作系统,接下来由操作系统控制机器的所有活动。

3.4 操作系统管理资源的方法

3.4.1 资源复用

由于计算机系统的物理资源是宝贵和稀有的,为解决物理资源数量不足的问题,操作系统让众多进程共享物理资源,这种资源共享称为资源复用。物理资源的复用共享有两种基本方法:时分复用共享和空分复用共享。

1. 时分复用共享

时分复用共享是指把资源的使用时间分割成更小的单位,并把时间片分给各进程,大家轮用使用。CPU 是时分复用的典型例子。

时分复用技术能提高资源利用率的根本原因在于,它利用某设备为一用户服务的空闲时间,又转去为其他用户服务,使设备得到最充分的利用。

2. 空分复用共享

顾名思义,空分就是按照空间进行划分。即将资源从“空间”上分割成更小的单位,把这些不同单位同时分给不同的进程使用。内存和外存就是空分复用的典型例子。

如果说多道程序技术(时分复用技术)是通过利用处理机的空闲时间运行其他程序,提高了处理机的利用率,那么,空分复用技术则是利用存储器的空闲空间分区域存放和运行其他的多道程序,以此来提高内存的利用率。

但是,单纯的空分复用存储器只能提高内存的利用率,并不能实现在逻辑上扩大存储器容量的功能,还必须引入虚拟存储技术才能达到此目的。虚拟存储技术在本质上是实现内存的分时复用,即它可以通过分时复用内存的方式,使一道程序能在远小于它的内存空间中运行。

3.4.2 资源虚拟化

虚拟化又称虚拟。虚拟化是一种资源管理技术,是指操作系统中实现对计算机软硬件资源进行有效管理的技术和手段,能进一步提高操作系统为用户服务的能力和水平。

虚拟化的本质是将计算机的各种实体资源,如 CPU、内存、磁盘空间、网络适配器等,予以抽象、转化、模拟和整合,把一个物理资源转变为多个逻辑上的对应物,也可以把多个物理资源变成单个逻辑上的对应物,也就是创建无须共享的多个独占资源的假象,或创建易用且多于实际物理资源数量的虚拟资源假象,以达到多用户共享一套计算机物理资源的目的。

通俗地说,所谓虚拟化,就是创造出一种虚拟的资源,然后将若干这种虚拟资源,对应于一种实际的物理资源,进程需要使用物理资源时,只需要使用这种虚拟资源即可,由操作系统来负责协调各个虚拟资源同时对物理资源的访问,进程无须关心竞争问题。

复用和虚拟化两者相比较,“复用”所分割的是实际存在的物理计算机资源,而“虚拟化”则实现假想的虚拟同类资源。采用虚拟化技术不仅可以解决物理资源数量不足的问题,而且能

够为应用程序提供易于使用的虚拟资源并创建更好的运行环境。

虚拟化技术可以用于外部设备、存储资源以及文件系统等。

3.4.3 资源抽象

复用和虚化的主要目标是解决物理资源数量不足的问题,抽象则用于处理系统的复杂性,重点解决资源的易用性。

资源抽象是指通过创建软件来屏蔽硬件资源的物理特性和实现细节,简化对硬件资源的操作、控制和使用的一类技术。即对内封装实现细节,对外提供应用接口。

操作系统中有以下三种基础的抽象。

(1) 进程抽象:对于运行的程序在 CPU 上状态的一种抽象,包括处理器状态(程序计数器、通用寄存器、堆栈指针寄存器等)和内存状态。

(2) 虚存抽象:物理内存被抽象成一种数组形式的虚拟主存,给进程造成独占整个主存的假象,由操作系统负责管理虚拟主存到真实物理内存的对应。

(3) 文件抽象:将磁盘、光盘等存储介质设备上存放的信息抽象为一个逻辑字节流,称为“文件”,用户通过创建、打开、读写、关闭等操作来控制文件,或者控制磁盘等的运行。

抽象技术也可以用于定义和构造多层软件抽象,每层软件都隐藏下一层的实现细节,从而形成多级资源抽象,数据 I/O 也依赖于多层抽象。

良好的抽象不但屏蔽使用上的复杂性,使用户容易理解和使用;还能防止程序员有意或无意地对资源的滥用;此外还能够为使用低层软件提供强有力的支持。

3.5 操作系统的安全和维护

3.5.1 操作系统的安全

在开放的网络环境中,操作系统安全在计算机系统整体安全中至关重要。操作系统的安全性可能会受到多种因素的影响,主要包括以下 6 种。

(1) 软件漏洞:操作系统中的软件漏洞是一个主要的安全威胁。这些漏洞可能包括缓冲区溢出、输入验证错误和错误的权限控制等问题,黑客可以利用这些漏洞来执行恶意代码、获取系统权限或篡改数据。

(2) 弱密码和身份验证机制:使用弱密码和不安全的身份验证机制,都会增加操作系统受到未经授权访问的风险。例如,管理员账户使用默认密码或共享密码,或者没有实施多因素身份验证。

(3) 不适当的权限管理:操作系统中不正确的权限分配和管理可能导致未经授权的访问或误用系统资源。例如,将过多的权限授予普通用户。

(4) 恶意软件和病毒感染:恶意软件(如病毒、间谍软件和勒索软件)可以破坏系统、窃取敏感信息或使系统无法正常工作。如果操作系统缺乏有效的防病毒和反恶意软件保护机制,就容易受到恶意软件的侵害。

(5) 未及时更新和补丁:操作系统厂商会定期发布安全更新和补丁,以修复已知漏洞

和增强系统安全性。如果系统没有及时安装这些更新和补丁,就会给攻击者留下可利用的漏洞。

(6) 不安全的网络连接:操作系统通过网络进行通信和数据传输。如果网络连接不安全,例如缺乏加密、未经身份验证的访问和不安全的协议使用,攻击者就可以窃听通信、篡改数据或进行中间人攻击。

实现操作系统安全的主要目标有四个:识别操作系统中的用户、依据系统设置的安全策略对用户的操作行为进行相应的访问控制、保证操作系统自身的安全性和可用性、监督并审计操作系统运行的安全性。常见的操作系统安全机制和方法主要包括:

(1) 访问控制:通过实施强密码策略、用户账户管理和访问权限控制,限制对系统资源和敏感数据的访问。包括使用多因素身份验证、账户锁定和权限分级等。

(2) 用户权限管理:操作系统通过用户权限机制,将特权分配给不同级别的用户,并限制用户在系统中的行为。这包括管理员特权、用户组管理和最小权限原则的应用。

(3) 文件系统权限:操作系统通过文件系统权限设置,控制对文件和目录的读、写和执行权,以确保用户只能访问其需要的文件,并防止未经授权的修改或删除。

(4) 加密和数据保护:使用加密算法和协议对敏感数据进行加密,以保护数据在存储和传输过程中的机密性和完整性。

(5) 更新和补丁:及时安装操作系统厂商提供的安全更新、补丁和漏洞修复程序,以修复已知的安全漏洞并防止潜在的攻击。

(6) 防火墙:配置和管理防火墙,以监控和控制进出系统的网络流量,并根据规则和策略阻止未经授权的访问和恶意行为。

(7) 定期备份和灾难恢复:制订定期备份策略和灾难恢复计划,确保操作系统和关键数据的备份,并测试恢复过程的可行性和有效性。

(8) 安全审计和日志管理:记录和监视系统活动,并定期审查系统日志,以便发现异常行为、入侵尝试和潜在的安全漏洞。

(9) 持续监测和漏洞管理:定期进行系统漏洞扫描和安全评估,及时修复发现的漏洞,并持续监测和评估系统的安全状况。

将以上安全机制和方法结合起来,可以有效提高操作系统的安全性,并保护系统免受各种威胁和攻击。然而,操作系统的安全是一个持续过程,需要不断更新和改进,以适应新型威胁和攻击技术的出现。

3.5.2 操作系统的维护

操作系统的维护是确保操作系统持续正常运行和保持良好性能的一系列活动。常见的操作系统维护任务包括:

(1) 更新和修补:及时安装操作系统厂商提供的更新、补丁和安全漏洞修复程序,以确保操作系统的安全和稳定性。

(2) 硬件驱动程序更新:根据硬件制造商发布的最新驱动程序,更新或升级操作系统中的硬件驱动程序,以提高设备的兼容性和性能。

(3) 安全管理:配置和管理操作系统的安全设置,包括访问控制、用户权限和防火墙等,以保护系统免受恶意攻击和未授权访问。

(4) 定期备份：制定和执行定期备份策略，确保操作系统和用户数据的备份，并验证备份的可用性和完整性。

(5) 监测和优化性能：使用系统监控工具来监视操作系统的性能指标，如 CPU 利用率、内存使用情况和网络流量等，并根据需要采取必要的优化措施。

(6) 日志管理：监视和管理操作系统生成的日志文件，以便追踪系统活动、故障排查和安全审计等。

(7) 定期更新系统文档：维护和更新有关操作系统的文档和记录，包括配置文件、安装说明和用户手册等，以便于管理员和用户参考和使用。

(8) 磁盘管理：定期进行磁盘清理和碎片整理，以释放磁盘空间并优化文件存储的连续性，从而提升系统响应速度和效率。

3.6 本章小结

操作系统是配置在计算机硬件上的第一层软件，是对硬件系统的首次扩充。操作系统用于管理计算机系统中的各种资源（包括硬件资源和软件资源），并为用户和应用程序提供一个良好的接口。故操作系统追求的目标主要有两点：一是提高资源的利用率，尽可能地使计算机系统中的各类资源得到充分利用；二是为用户提供良好的界面，一个好的操作系统应该给用户提供一个简单、清晰、易于使用的用户界面。操作系统性能的高低决定了计算机的潜在硬件性能能否得到发挥。

计算机中任何的一个概念，都是用户在解决某个具体问题的过程中逐渐形成的，操作系统这个概念也不例外。在软硬件技术的发展和应用需求的共同推动下，操作系统经历了从无到有、从简单到复杂的发展历程。目前存在着多种类型的操作系统，不同类型的操作系统，其目标各有所侧重。通常在计算机硬件上配置的操作系统都具备方便性、有效性、可扩充性和开放性。

操作系统具备 5 个功能，分别是处理机管理、存储管理、设备管理、文件管理以及用户接口。操作系统进行资源管理主要用到三大技术，分别是资源复用、资源虚化和资源抽象技术。

操作系统安全是网络安全的基石。操作系统的安全性是确保操作系统及其相关组件免受恶意攻击、未经授权访问和数据泄露等威胁的能力。操作系统的安全性会受到多种因素的影响，在维护和使用操作系统时，应重视这些因素并采取相应的安全措施提高操作系统的安全性，并保护系统免受各种威胁和攻击。

习题 3

简答题

1. 简述操作系统在计算机系统中的地位。
2. 什么是操作系统？其追求的主要目标是什么？
3. 何为脱机 I/O 和联机 I/O？



习题答案

4. 什么是多道程序设计技术？其特征是什么？
5. 试说明推动多道批处理系统形成和发展的主要动力是什么？
6. 试说明推动分时系统和实时系统形成和发展的主要动力是什么？
7. 简述分时系统和实时系统的主要差别。
8. 简述操作系统的功能。
9. 什么是时分复用技术？举例说明其能提高资源利用率的根本原因。
10. 为什么说操作系统实现了资源的抽象？