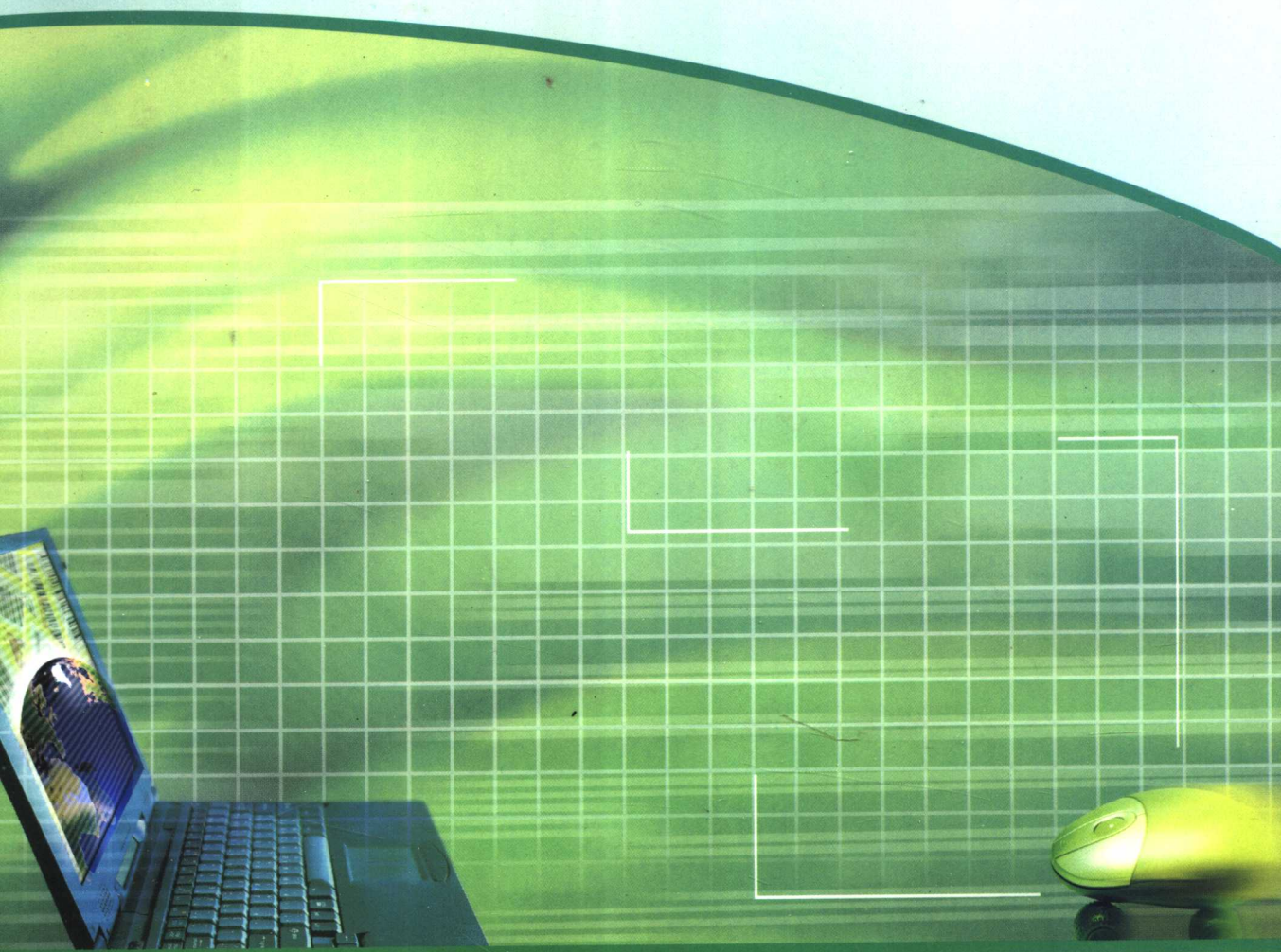


操作系统教程与实例分析

主编 / 李 芳 副主编 / 刘晓春 李晓莉



西安地图出版社

操作系统教程与实例分析

主 编 李 芳

副主编 刘晓春 李晓莉

西安地图出版社

图书在版编目(CIP)数据

操作系统教程与实例分析/李芳主编. —西安:西安地图出版社, 2006

ISBN 7-80670-903-7

I. 操... II. 李... III. 操作系统 IV. TP316

中国版本图书馆 CIP 数据核字(2006)第 001069 号

内 容 简 介

本书介绍了计算机系统中必不可少的一个重要系统软件——操作系统(OS)。全书共分为三个层次,第一个层次包括第一章至第七章,详细介绍操作系统的基本概念、基本原理与实现技术。第二个层次包括第八章至第九章,结合操作系统原理具体分析目前主流操作系统——Windows 2000/NT 和 Linux 的实现技术和方法。第三个层次第十章,介绍当前操作系统发展的最新趋势,网络操作系统、分布式操作系统和嵌入式操作系统的基本原理与应用环境。

本书内容全面,重点突出,在系统介绍经典内容的基础上,还介绍了近年来操作系统的先进技术及最新发展动态,如线程的概念及应用、实时调度的调度算法、多处理机调度技术、分布式操作系统和嵌入式操作系统等;书中列举了大量实例,力求将抽象的概念简单化,将复杂的理论与实际联系起来;书中还收集了大量习题,其中既有一般概念和基本原理测试题,还包括近年来操作系统等级考试与研究生入学考试试题。

本书既可作为计算机及相关专业的教科书,也可作为从事计算机工作的科技人员学习操作系统的参考书。

操作系统教程与实例分析

李 芳 主编

西安地图出版社出版发行

(西安市友谊东路 334 号 邮政编码:710054)

新华书店经销 长安大学雁塔印刷厂印刷

787 毫米×1092 毫米 1/16 开本 18.75 印张 490 千字

2006 年 1 月第 1 版 2006 年 1 月第 1 次印刷

印数: 0001~1000

ISBN 7-80670-903-7/TP·33

定价: 32.00 元

前 言

操作系统是用户与计算机之间的交互通道,管理和控制着系统资源。只有理解了操作系统,用户才能方便、灵活地使用计算机;只有掌握了操作系统提供的各种功能强大的系统服务,用户才能更好地利用系统资源,在操作系统的基础上建立自己的应用系统,开发自己的应用软件。

为做到理论与实践相结合,突出操作系统各主要部分关键环节的概念、功能、原理和方法,本书选择当前较为流行的有代表性的操作系统——Windows 2000/NT 和 Linux 为实例,深入分析和讲解了它们的部分关键环节的实现技术。此外,由于操作系统所介绍的原理与算法比较抽象,难以理解和掌握,笔者根据多年的教学经验,将一些典型实例引入本书,使读者通过实例的分析和应用,充分掌握操作系统的原理与算法思想,提高分析问题、解决问题的能力。考虑到近年来操作系统在技术与应用上都有了一些较新的发展,在本书中引入了线程、实时调度、多处理机、网络操作系统、分布式操作系统、嵌入式操作系统等新技术的介绍。

本书共分为十章。第一章为绪论,从操作系统的发展和作用引入操作系统的概念,简要介绍操作系统的功能,依据操作系统的不同设计目标对操作系统进行了分类介绍;第二章为操作系统接口,介绍操作系统的两种接口,重点介绍了 UNIX 常用操作命令和系统调用的实现原理;第三章为进程管理,从单道程序和多道程序执行的不同特征引入进程的概念,然后介绍进程的基本特征和运行状态及操作系统对进程的控制机构,通过实例分析进程的同步与互斥关系的解决方法以及进程通信的常用方式,最后介绍线程的概念和应用环境;第四章为处理机调度与死锁,介绍处理机调度级别与常用调度算法的基本思想,并通过实例对不同的算法进行优劣比较,包括进程调度与实时调度和多处理机调度;第五章为存储管理,从内存分配、地址变换、内存扩充与内存保护四个方面分别介绍了分区式管理、页式管理、段式管理与段页式管理四种常用的存储管理方案;第六章为设备管理,由低到高逐层介绍 I/O 系统的层次结构的一些关键技术;第七章为文件系统,主要讲述文件与文件系统、文件的组织和存取以及文件的保护;第八章为 Linux 内核结构,以目前十分流行的 Linux 系统为例,主要介绍 Linux 进程管理、Linux 存储管理、Linux 文件管理、Linux 设备管理的基本原理;第九章为 Windows 系统模型,主要介绍 Windows 2000/NT 的操作系统模型、进程管理、线程管理、文件管理、内存管理和设备管理。第十章为网络、分布式和嵌入式操作系统,本章以计算机的新应用模式为背景,着重讨论网络操作系统、分布式操作系统和嵌入式操作系统的特点及应用环境。

总之,本书具有如下特色:

(1)内容全面。本书将当前的主流操作系统融于一体,既有操作系统的常用原理介绍,又有具体实现技术的详细分析,从而使读者通过本书的学习即可较好地掌握各种常用操作系统的基本理论和实用技术。

(2)富有启发性。采用“实例引导,任务驱动”的编写方式,增加实例分析,使读者掌握操作系统实例解析题的解析方法,激发读者的学习兴趣。

(3)图文并茂。对于较深奥的理论知识,尽量以图表的形式来说明,便于读者理解和掌握。

(4)理论联系实际。具体分析当前流行的 Linux 操作系统和 Windows 操作系统的实现技术和方法,将抽象的原理和具体实例相结合,使读者能够在实际应用中更好地建立自己的应用系统,开发自己的应用软件。

本书的第二、三、四、五、六章由长安大学的李芳编写;第一、七、八章由长安大学的刘晓春编写,第九、十章由长安大学的李晓莉编写。

由于编者水平有限,在本书的编写中,仍难免会有错误和不当之处,恳请读者批评指正。

编 者
2005 年 12 月

目 录

第一章 绪论	(1)
1.1 操作系统的形成和发展	(1)
1.2 什么是操作系统	(5)
1.3 操作系统的功能和特性	(6)
1.4 操作系统的类型	(8)
1.5 操作系统硬件环境	(13)
第二章 用户接口	(17)
2.1 命令控制界面接口	(17)
2.2 UNIX 系统的命令控制界面	(19)
2.3 程序接口	(25)
第三章 进程管理	(32)
3.1 进程的概念	(32)
3.2 进程控制块和进程的状态	(37)
3.3 进程的控制	(41)
3.4 进程同步	(46)
3.5 经典的进程同步问题	(55)
3.6 进程通信	(58)
3.7 线程	(62)
第四章 处理机调度与死锁	(69)
4.1 调度的基本概念	(69)
4.2 调度算法	(74)
4.3 实时调度算法	(80)
4.4 多处理机调度	(83)
4.5 死锁	(85)
4.6 解决死锁问题的方法	(87)
第五章 存储管理	(96)
5.1 存储管理基本概念	(96)
5.2 分区式存储管理	(101)
5.3 覆盖与交换技术	(105)
5.4 页式存储管理	(107)
5.5 淘汰算法与抖动现象	(113)
5.6 段式存储管理	(119)
5.7 段页式存储管理	(128)
第六章 设备管理	(132)
6.1 设备管理概述	(132)
6.2 数据传送控制方式	(135)

6.3	中断处理与设备驱动程序	(141)
6.4	缓冲技术	(144)
6.5	设备分配	(148)
6.6	逻辑 I/O 系统	(153)
第七章	文件管理	(155)
7.1	文件和文件系统	(155)
7.2	文件的逻辑结构	(157)
7.3	文件的物理结构	(161)
7.4	文件存储空间的管理	(165)
7.5	文件目录管理	(167)
7.6	文件共享和保护	(172)
第八章	Linux 系统内核结构	(177)
8.1	Linux 概述	(177)
8.2	Linux 进程管理	(182)
8.3	Linux 进程通信	(188)
8.4	Linux 存储管理	(196)
8.5	Linux 文件管理	(208)
8.6	Linux 设备管理	(227)
第九章	Windows 系统模型	(232)
9.1	Windows 2000/NT 操作系统简介	(232)
9.2	Windows 2000/NT 进程管理	(243)
9.3	Windows 2000/NT 中线程的实现	(249)
9.4	Windows 2000/NT 进程互斥和同步的实现	(253)
9.5	Windows 2000/NT 中进程间通信	(254)
9.6	Windows 2000/NT 内存管理	(257)
9.7	Windows 2000/NT 设备管理	(269)
9.8	Windows 2000/NT 文件系统	(275)
第十章	网络、分布式和嵌入式操作系统	(283)
10.1	网络操作系统	(283)
10.2	分布式操作系统	(288)
10.3	嵌入式操作系统	(291)

第一章 绪 论

操作系统(Operating System,简称 OS)是每台计算机必须安装的最基本的系统软件,是操纵硬件并为程序执行建立一个更实用的平台的系统软件。本章首先介绍操作系统的发展过程,旨在使学生从操作系统的发展中对操作系统的一些基本概念有一个初步的了解,也能从历史中看到该领域的前景。接下来总结了操作系统的概念,最后介绍几种常见的操作系统。

1.1 操作系统的形成和发展

为了更好的理解操作系统的基本概念和功能,让我们首先回顾一下操作系统形成和发展的历史过程。

一个事物的产生总是有它的原因和其必然性的,操作系统也不例外。操作系统是伴随着计算机的产生而产生,并且随着计算机技术及其应用的发展而不断发展和完善的。在计算机硬件经历从电子管时代、晶体管时代、集成电路时代、大规模和超大规模集成电路时代的发展的同时,操作系统也完成了从无到有,从简单到复杂的演变过程。操作系统的发展经历了以下几个阶段:手工操作阶段、批处理系统、多道程序系统、现代操作系统等。

下面就从操作系统发展的基础谈起。

1.1.1 操作系统发展的基础

1. 计算机体系结构的发展

计算机体系结构经历了从简单到复杂、从单机到多机、从集中到分散式的发展,它所依赖的物质基础是微电子技术和超大规模集成技术,其代表是硬件资源的发展。

硬件体系结构的发展,给管理各种硬件资源提出了更高的要求,对系统各部分(I/O 系统、主机系统、存储子系统、网络子系统)要进行人工管理是不可能的,需要有一个超强功能的软件来组织和统一协调各个部件的工作,这就是操作系统。

2. 计算机软件资源的发展

软件的发展有两个重要的方面,一是支持用户进行系统开发,二是增强系统的管理和服务。计算机高级语言及其编译系统的发展打开了应用领域的大门,推动了应用系统的开发,随之而来的各种服务程序、实用程序、工具程序、娱乐程序使计算机系统资源概念发生了变化,资源不仅仅指硬件资源,还包括上述各类软件,以及由数据组成的数据库,文档信息等,这就是软件资源,对软件资源的管理和支持,操作系统同样责无旁贷。

3. 应用环境与需求的发展

计算机用户的需求随着应用领域的扩大和深入,对计算机系统的要求也在改变。一般

来说,希望系统的功能越强越好,系统的接口尽可能简单,速度尽可能快,最好能共享各种系统资源,有较高的可靠性和有效性,并且能按用户的需求重新组构或扩充等。而上述需求仍然在不断的变化,从而从另一方面促进了操作系统的发展。

1.1.2 操作系统的发展过程

由于在历史上操作系统与计算机体系结构存在非常密切的联系,我们将按照计算机的换代历程讲述操作系统的发展状况。

第一台真正的数字计算机是英国数学家 Charles Babbage (1792 - 1871 年)设计的, Babbage 投入了毕生的精力去构造他的“分析机”,但却没能让它成功地运行起来。因为它是纯机械式的,而当时的技术不可能使分析机的零部件达到它所需的精度。显然,分析机没有操作系统。但是 Babbage 认识到他的分析机需要软件,于是他雇用了英国著名诗人拜伦的女儿 Ada Lovelace 为他工作,由此 Ada 成了世界上第一位程序员, Ada 程序设计语言就是用她的名字命名的。

1. 手工操作阶段

从 Babbage 之后一直到二战期间,数字计算机几乎没有什么进展。在 20 世纪 40 年代中期,数学家冯·诺依曼(John Von Neuman)提出了程序的概念,先后设计制造了电子管数字计算机,其主要特点是:计算机的主要器件是电子管,以汞延迟线、磁芯、磁环作为主存储器,以磁鼓或磁带作为外部存储设备,计算机总体结构以运行器为中心。这种计算机体积庞大,耗能严重,速度极慢,且价格昂贵。这时没有编程语言,上机完全是手工操作,操作系统更是闻所未闻,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功能。计算机主要用于数值计算。

这种手工式处理方式存在着明显的缺点:

(1)上机用户独占全部资源,其他用户只能等待。

(2)手工操作的出错率比较高。

(3)随着计算机运行速度的提高,计算机的高速与慢速的手工操作之间形成矛盾。

直到 20 世纪 50 年代初期,出现了穿孔卡片,这时就可以不用插板,同时出现了脱机的输入/输出技术。该技术是指在进行程序的调试与执行时,程序员事先把写有程序的卡片装入卡片输入机,在一台外围机的控制下,把卡片上的数据输入到磁带上。当 CPU 需要时,再从磁带上把程序和数据送进计算机,然后启动计算机运行的。类似的,当需要输出时,可由 CPU 直接把数据送到磁带上,然后再在另一台外围机的控制下,将磁带机上的结果通过相应的输出设备输出。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,或者说它们是在脱离主机的情况下进行的,故称为脱机输入/输出方式;反之,在主机的直接控制下进行的输入/输出的方式称为联机输入/输出方式。

2. 单道批处理系统

20 世纪 50 年代后期,推出了晶体管计算机,计算机的运行速度得到了很大提高,外部存储设备除磁鼓、磁带外,引入了磁盘。这个时期被称为第二代计算机。在此期间,程序和编程语言得到较大的发展。这个时期,计算机安装在专门的机房,有专人操作。然而,慢速的手工操作和高速的计算机运行之间的矛盾越来越严重,变得让人无法忍受。另外计算机硬件仍然十分昂贵,需要最大限度的利用机器资源。人们为了提高利用率,减少机时的浪费,开发了批处理操作系统的概念。

单道批处理系统的中心思想是使用一个称作监督程序的软件,通过使用这个软件,用户不再直接访问机器,而是把程序和数据以及用户对程序的控制意图写在卡片或磁带中,以作业的形式提交给计算机操作员,由操作员把这些作业按顺序组织成一批,并将整批作业放在输入设备上,供监督程序使用。首先,监督程序将磁带上的第一个作业装入内存,该作业完成处理后返回到监督程序,然后,监督程序自动加载下一个程序。这样在监督程序的控制下,系统可以连续运行,一直到这批作业处理完毕。这里的监督程序也就是操作系统的雏形。由于系统对作业的处理是成批进行的,且在内存中只能保持一道作业运行,故称之为单道批处理系统。

这种监督程序可以使作业间自动转换,减少了时间的浪费,但是它还没有真正形成对作业的控制和管理。如果一个用户的计算机作业非常庞大,在运行期间,它将独占计算机系统所有资源,在运行完成之前,任何其他用户的作业,哪怕是很短的作业也只能等待。

第二代的计算机主要用于科学计算,程序大多用 FORTRAN 语言和汇编语言编写。典型的操作系统是 FMS(FORTRAN Monitor System, FORTRAN 监控系统)和 IBSYS(IBM 为 7094 机配备的操作系统)。

3. 多道程序系统

进入 20 世纪 60 年代初期,计算机的应用形成两个领域,即面向科学与工程复杂计算和面向字符处理的商务应用。为此,出现了通用计算机。这个时期被称为第三代计算机。其主要特点是:以中、小规模集成电路作为逻辑器件,主存储器除磁芯外,开始使用半导体存储器,外存储设备则以磁盘为主。著名的 IBM System360 计算机则是一个代表,它是第一台采用集成电路的计算机,比第二代计算机有更好的性能价格比,也是一台大型通用计算机。通用计算机要达到通用性,必须功能强大,能够满足环境应用要求,处理和适应各种设备环境,才能发挥极大效能。所以必须有一个强有力的、功能复杂的监控程序来监管和协调系统的所有操作,安排和调度用户所提交的作业,分配系统所共用的各种软件和硬件资源。初始的监督程序不能完成这些功能,继而进一步发展成为一个强功能的程序集合,即现在所说的操作系统。

在 IBM System 360 上运行的 OS/360 操作系统被认为是真正的操作系统。因为它真正实现了资源管理,建立了资源管理的机制,直到现在,许多操作系统中的技术和结构还多少留有它的影子。尽管 OS/360 存在较大的隐患和不足,但它引入了一种新技术,即多道程序技术。多道程序系统的思想是把用户提交的作业都存放在外部的存储器中,形成作业队列,等待运行。当需要调入作业时,将由操作系统中的作业调度程序对外存中的一批作业,根据其对资源的要求和一定的调度原则,调入几个作业进入内存,让它们交替运行。当某个作业完成,然后再调入一个或几个作业。在单处理机系统中,多道程序运行的特点是:

(1)多道:计算机内存中同时存放多道相互独立的程序。

(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,交替执行。

多道程序技术的引入,大大提高了 CPU 的利用率,提高了系统资源的利用率,但同时要求有专门的软硬件机构来支持多道程序,支持内存的分块及防止作业的相互干扰。因此,这种通用操作系统的设计是极其复杂的,隐患极多,难以排除,操作系统的设计者不得不以版

本更新的方式来清除原版本中的隐患并增加新的功能。

多道程序系统解决了资源利用率的问题,但随着计算机的普遍使用,早期的批处理暴露了它的缺陷,如一个作业从提交到取回运行结果往往要很长时间。对于程序员更不能忍受的是,一个误用的逗号就会导致编译失败,程序员必须重新提交,重新等待,浪费很长的时间。同时越来越多的用户希望在使用计算机的过程中能够进行非常方便的人机交互,也就是说可以在程序运行时直接对计算机进行控制。除此之外,很多用户还十分怀念手工操作阶段自己可以独占计算机的良好感觉。用什么办法能使得既能保证计算机效率,又能方便用户使用呢?在20世纪60年代中期,随着键盘、显示器和显示终端等交互设备的问世,以及计算机的小型化和微型化,使用户可以直接与机器打交道,也使得这种愿望成为可能。第一个分时系统CTSS出现了,它是由麻省理工学院开发的。在分时系统中,一台计算机可以同时连接多个带有显示器和键盘的用户终端,同时允许用户通过自己的终端,以交互方式使用计算机,共享主机中的资源。由于调试程序的用户常常只发出简短的命令,而很少执行费时长的命令,所以计算机能够为一些用户提供快速的交互式服务,同时在CPU空闲时还能运行后台的大作业。比较著名的多用户分时系统是UNIX系统。

分时系统的出现解决了用户方便实用的需求,然而用户的需求是不断变化的。随着计算机的应用范围的扩大,计算机越来越多地被用于生产过程的控制和武器系统的实时控制。这两种控制系统都要求能实时采集现场数据,并对所采集的数据进行及时处理,进而自动控制相应的执行机构,以保证正常执行。除此之外,人们还经常需要对外来的实时信息进行控制,比如说飞机订票系统和银行业务系统。无论是生产过程的实时控制还是外来信息的实时控制,都要求系统能够在允许时间范围内作出响应,并且对系统的可靠性要求也比较高,这样就导致了实时系统的出现。

随着多道程序技术的推广及应用,后期的操作系统具有了多种方式和类型。多道批处理系统和分时系统的不断改进,实时系统的出现致使操作系统日益完善。在此基础上出现了通用操作系统。也就是说在一种操作系统中同时具有了批处理、分时、处理和实时处理功能。至此,操作系统的基本概念、功能、基本结构和组成都已经形成并日趋完善。

4. 个人操作系统

随着微电子技术和VLSI技术的迅速发展,大规模和超大规模集成技术用于计算机,将运算器、控制器和相应接口集成在一块基片上,产生了微处理器。计算机硬件价格急速下降,按照计算机硬件分代的概念趋于模糊,计算机的体系结构趋于灵活、小型、多样化。小型、微型计算机在运算速度、存储容量、外存容量和I/O接口等方面有了很大的发展。许多原来只有在大型计算机上实现的技术,逐步下移到小、微型计算机上,出现了面向个人用户的计算机,简称PC机,并同时向便携式计算机发展。计算机直接与用户交互,系统操作界面更加友好、灵活方便,功能支持更加强大,可靠性高、体积小、价格低,得到了越来越广泛的应用。此时的软件系统(包括操作系统)要求面向用户,使用户操作更方便更灵活,无需了解计算机硬件及其内部操作。自1984年APPLE公司的Macintosh计算机系统引入图形界面(GUI)以来,视窗操作和视窗界面得以大大发展,从而形成了操作系统的用户界面管理功能模块。这个时期(198—1994年)被认为是第四代计算机系统发展过程,其配置的操作系统被称为现代操作系统。

目前,个人计算机操作系统层出不穷,它们有MS-DOS, UNIX, OS/2, WINDOWS,

MAC 等,其中有代表性的是 MS—DOS, WINDOWS 和 UNIX。个人操作系统使计算机的使用环境和开发平台越来越灵活和高效。多媒体技术的出现和多媒体数据与信息处理,给个人计算机操作系统提出了更高的要求。个人计算机操作系统要满足应用领域的各种要求,要满足信息社会中多种媒体信息的处理,要提供更灵活更友好的用户界面,它必须提供新的功能,支持各种不同类型的外部设备,这样的操作系统才会受到用户的欢迎,才会有强有力的生命力。

5. 操作系统的发展趋势

随着计算机技术的飞速发展,新型机器的层出不穷,按照硬件划分时代的概念已经模糊,分代已无多大意义。操作系统将进一步向以下几个方向发展:

- (1)个人操作系统;
- (2)嵌入式操作系统;
- (3)网络操作系统;
- (4)分布式操作系统;
- (5)智能型操作系统。

1.2 什么是操作系统

前面一节叙述了操作系统的发展过程,使我们对操作系统有了一定的认识。那么,究竟什么是操作系统,操作系统又具有哪些特点呢?

首先举一个简单的例子。当我们购买到一台计算机的时候,我们所作的第一件事就是装入系统,一般而言大多数人都会选择 Windows 系统。可以试想一下,如果不装入 Windows,我们要怎么样来使用计算机,怎样上网、编程序或处理图片呢?相信即使是顶尖的计算机好手也无法做到这一点。这就是操作系统的作用所在。我们把没有装入任何软件的计算机称为裸机,从上面的例子可以看出,裸机只是提供了计算机系统的物质基础,属于计算机硬件部分,是无法供用户使用的。庆幸的是,计算机系统除了硬件之外还有软件,而操作系统是覆盖在裸机上的第一层软件,为我们方便的使用其他软件起到了一个桥梁的作用。正是操作系统把一个裸机变成一个可“操作”的、方便灵活的机器,通常称之为虚拟机(virtual machine)或扩展机(extended machine)。而各种实用程序和应用程序运行在操作系统之上,它们以操作系统作为支撑环境,同时又向用户提供完成其作业所需要的各种服务。在一般情况下,实际呈现在用户面前的计算机系统已是经过若干层软件改造的计算机。也就是装入了各种应用软件的计算机。

如图 1-1 所示,可把整个计算机系统按功能划分为四个层次,即硬件、操作系统、系统实用软件和应用软件。这个层次表现为一种单向的服务关系,即外层可以使用内层提供的服务,反之,则不行。在这个层次结构中,包围着系统硬件的一层是操作系统。它是最基本的系统软件,控制和管理着系统硬件,向上层的实用程序和用户应用程序提供一个屏蔽硬件工作细节的良好使用环境,通过系统核心程序对系统中的资源进行管理,通过这些服务将对所有的硬件的复杂操作隐藏起来,为用户提供一个透明的操作环境。

当一台计算机(或网络)有多个用户时,因为用户间可能相互影响,所以管理和保护存储器、I/O 设备以及其他设备的需求随之增加。而且用户往往不仅需要共享硬件,还要共享信

息(文件、数据库等),此时操作系统的首要任务是跟踪资源的使用状况、满足资源请求、提高资源利用率,以及协调各程序和用户对资源的使用冲突。

因此,操作系统的作用可以归纳为三点:

1. 管理系统资源

众所周知,一个计算机系统资源包括两大类:硬件资源和软件资源。硬件资源通常包括 CPU、主存、输入/输出设备;现代计算机中软件资源包括计算机系统内的各类文件。相应地,操作系统的主要功能也正是针对这些资源进行有效地管理。

2. 为用户提供一个好的接口

操作系统是计算机与用户之间的一个接口,用户通过操作系统为用户提供良好的界面,而无需了解许多硬件和系统软件细节,就能方便灵活地使用计算机。

3. 应最大限度地发挥系统资源使用效率

由于多个用户共享系统资源,不可避免地会出现多个用户竞争资源,操作系统应合理地为用户分配资源,合理地组织计算机的工作流程,提高资源的利用率和系统的吞吐量。

综上所述,我们可以给操作系统下一个定义:

操作系统是直接控制和管理计算机软硬件资源的最基本的系统软件,它可以合理地组织计算机的工作流程,用以方便用户充分、有效地利用这些资源并增强整个计算机的处理能力。

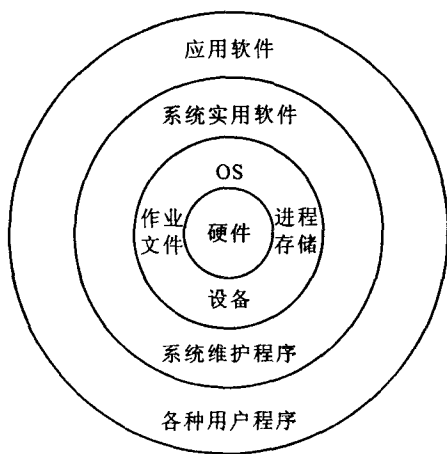


图 1-1 计算机系统层次结构

1.3 操作系统的功能和特性

从操作系统的定义可知,操作系统是直接控制和管理计算机软硬件资源的最基本的系统软件,那么操作系统具有哪些功能和与其他软件不同的特点呢?

1.3.1 操作系统的功能

操作系统是计算机系统资源的管理者,如何最大限度的发挥计算机系统资源的使用效率,如何为用户提供一个方便的使用环境是操作系统要解决的主要问题,也是使用计算机的目标。一般情况下我们将计算机资源分为硬件资源和软件资源。前者主要包括处理机、存储器(内存和外存)、输入输出设备。而后者主要包括各种程序和数据,它们都以文件形式存储在外存储器。因此,常规的操作系统通常具有如下五个方面的功能:

1. 进程管理

主要是对处理机的管理。为了提高 CPU 的利用率,采用多道程序技术,为了描述多道程序的并发执行,就引入进程的概念,通过进程管理协调多道程序之间的关系,以使多道程序可以以最高的效率执行,并且 CPU 资源得到最充分的利用。主要的功能包括如下:

(1)进程控制。在多道程序环境下,每个程序的运行都必须为之建立一个或几个对应的进程。进程控制主要包括进程的创建、进程的撤消、进程的阻塞与唤醒。

(2)进程的同步控制。引入多道程序后,由于进程之间的相互限制和相互依赖,导致进程之间出现了两种关系:互斥与同步,进程管理必须提供相应的同步机构,协调进程之间的运行,以保证进程执行结果的正确性。

(3)进程通信。多个进程之间具有相互合作的关系,因此它们之间必须交换一定的信息。进程通信的任务就是用来实现相互合作的进程之间的信息交换。

(4)调度。系统中同时存在多个进程,它们都要求得到 CPU 的处理,进程调度的主要任务就是按照一定的算法,从多个就绪队列中选中的一个进程,使之得到执行。

2. 存储管理

由于内存的需求量大,而内存价格又昂贵,且受 CPU 寻址能力的限制,内存容量有限。因此,当多个程序共享内存时,如何为它们分配内存空间,使它们彼此隔离、互不侵扰,并且可以在一定条件下互相调用。当内存不够用时,如何把当前运行的数据及程序即时调出内存,需要运行时再从外存调入内存等等,都是存储管理的任务。

3. 设备管理

指除 CPU 和内存外的所有 I/O 设备的管理。除了进行实际 I/O 操作的设备外,还包括诸如设备控制器、DMA(直接存储访问,direct memory access)控制器、通道等支持设备。设备管理的首要任务是为这些设备提供驱动程序或控制程序,以使用户不必详细了解设备及接口的技术细节,就可方便地对这些设备进行操作。其另一个任务就是利用中断技术、DMA 技术和通道技术,使外围设备尽可能与 CPU 并行工作,以提高设备的使用效率,并提高整个系统的运行速度。

4. 文件管理

主要指对程序和数据等软件资源进行的管理。程序和数据是以文件的形式存放在外部存储器的,当需要时再把它们装入内存(OS 本身也是文件)。文件是软件资源,有效地组织、存储、保护文件,以使用户方便、安全访问它们,如何对文件系统进行优化,这是文件管理的任务。

5. 用户接口

除对系统资源的管理外,操作系统还必须向用户提供直接使用 OS 的手段。目前的操作系统都提供三种方式的接口:一种是命令接口,用户可通过该接口向作业发出命令以控制作业运行,该接口又进一步分为联机用户接口和脱机用户接口。前者由一组键盘操作命令以及命令解释程序组成。后者由一组作业控制语言组成。第二种是程序接口,即提供一组广义指令(系统调用)供用户程序和其他系统程序调用。每当应用程序要求操作系统提供某种服务(功能)时,便调用具有相应功能的系统调用。第三种是图形接口,这是较晚出现的操作系统提供的,它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的标示出来,用户可用鼠标方便完成操作,而不用花费大量时间记忆各种命令的名字与格式,使得计算机操作变得更为方便。目前主流的 OS 都提供了图形用户的接口。

1.3.2 操作系统的基本特征

尽管每种类型的操作系统都具有自己的特征,但作为现代操作系统,都具有三个重要特

征:并发性、共享性和不确定性。这三个特点是互相联系和互相依赖的,他们是互相独立的用户如何使用有限的计算机资源的反映。

1. 并发性

并发性是指程序的并发执行,也就是在主存中同时存放多道程序,它们同时处于运行状态。并发性是操作系统的主要特征之一。这种并发性不仅体现在操作系统程序与用户程序一起并发运行,就连操作系统本身的各种程序之间也是并发运行的,它们为用户提供并发服务。

应当指出,通常的程序是静态实体,它们是不能并发执行的。为了使多个程序能够并发执行,在操作系统中引入了进程的概念,来描述程序的执行过程和共享资源的基本单位。关于进程在后面的部分会作详细阐述。

还有一点需要注意的是,并行性和并发性是两个既相似又有区别的概念。并行性指两个或多个事件在同一时刻发生,是针对多处理机的计算机系统来说的。而并发性指两个或多个事件在同一时间间隔内发生,也就是在单处理机的情况下,宏观上有多个程序在同时运行,但微观上每一时刻仅能有一道程序执行,也就是说这些程序只能分时交替执行。

OS 中程序的并发执行使系统变的更为复杂,以致使系统中必须增设若干新的功能模块,分别用于处理机、内存、外设以及文件系统等资源的管理。

2. 共享性

并发运行的程序可共享系统资源。这种共享分为两种方式,互斥共享方式和同时访问方式,对于系统中的一些资源,比如打印机,在一段时间里只允许一个进程使用,进程对此资源的共享是互斥的,只有一个进程使用完之后,另一个进程才能够使用。而对于另外一些资源,如磁盘设备,可以被进程同时访问。

并发和共享是互为存在的条件,只有程序并发执行,才能谈得上资源共享,反过来,如果不能对资源共享进行有效的管理协调,程序并发执行的效率也会降低,甚至无法执行。

3. 不确定性

不确定性,也称为 OS 的随机性,它是指 OS 面对的是各类随机事件。在多道程序环境下,允许多个程序并发执行。内存中每个进程在何时获得处理机运行,何时因某种原因,如打印机请求而暂停,以及进程以怎样的速度向前推进,每道程序总共需要多少时间完成,等等,都是不可预知的。

OS 的不确定性是并发与共享的必然结果。人们不能对于所运行的程序的状态以及硬件设备的情况做任何的假定,因而一般来说我们无法确切地知道系统正处于什么状态。为了随机响应和正确无误地处理各种随机事件,OS 必须事先安排好对各种可能事件的处理。

1.4 操作系统的类型

从前面的介绍中,我们已经知道了几种主要的操作系统。不同类型的操作系统可以满足不同的应用要求。根据硬件结构 and 应用环境,我们把操作系统分成以下七种类型。

1.4.1 批处理系统

批处理系统是一种早期的大型机用操作系统。不过,现代操作系统都具有批处理功能。批处理系统的特点是采用脱机服务方式,即用户将控制作业的意图、把数据以及程序利用系

统提供的作业控制命令书写成作业说明书提交给操作员,操作员将其输入外存,由 OS 控制、调度各作业的运行,最后输出结果。它是一种非人工的干预方式。有单道和多道批处理系统。

单道批处理系统提高了 CPU 的利用率,减少了操作员手工操作的出错率,比起人工操作有了很大的进步,单道批处理系统具有以下特征:

(1)自动性。如果情况顺利,磁带上的作业能够自动依次运行,无需人工干预。

(2)顺序性。磁带上的各道作业顺序进入内存,在正常情况下,先调入内存的作业先完成。

(3)单道性。监督程序每次只从磁带上调入一道程序进入内存运行,内存中仅有一道程序运行。

单道批处理系统克服了手工操作的缺点,实现了作业的自动过渡,但在单道批处理系统中,同一时刻内存中仅有一道程序运行。就有可能出现两种情况:对于以计算为主的作业,输入输出量少,外围设备空闲;然而对于以输入输出为主的作业,又会造成主机空闲。这样计算机资源的利用效率仍然不高。为了进一步提高资源的利用率,在 20 世纪 60 年代中期又引入了多道程序设计技术,由此形成了多道批处理系统。

多道批处理系统具有以下几个特征:

(1)多道性。计算机内存中同时存放几道相互独立的程序。

(2)无序性。多个作业完成的先后顺序与他们进入内存的顺序并不是严格对应的。后进入内存的程序有可能比先进入内存的程序先完成。

(3)调度性。作业从提交给系统开始到完成需要两级调度,作业调度和进程调度。

随着多道批处理系统的出现,操作系统的发展便趋于完善了,如今它仍是三大基本操作类型之一。这也说明它具有其他类型的操作系统所不具备的优点:

(1)资源利用率高。内存中多道程序可以共享资源。

(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总工作量。由于在批处理操作系统中,CPU 和其他资源能够保持忙碌状态,只有当作业完成或出错时才进行切换,从而减小了系统开销。

但是批处理系统中无交互能力。在批处理系统中,用户以脱机方式使用计算机,即用户在提交作业以后到作业完成这个过程中,无法进行干预。这样对于修改和调试程序存在很大障碍。

1.4.2 分时系统

分时系统采用了分时技术,就是把处理机的运行时间分成很短的时间片,按时间片轮流把处理机分配给各个联机作业使用。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,把处理机让给另一作业使用,等待下一轮时再继续运行。由于计算机速度很快,作业运行轮转的也很快,每个用户都感觉是自己独占了一台计算机,同时又可以通过终端向系统发出各种命令,来控制程序的运行。

分时系统与批处理系统相比,具有完全不同的特点:

(1)交互性。用户可以通过终端方便的进行人机对话,可以请求系统提供各种服务,特别对于远程终端用户,可以直接在自己的终端上提交、调试和运行程序。

(2)多用户同时性。允许在一台主机上连接多台终端,多个用户通过轮流运行自己的时

间片来共享计算机资源。

(3)独立性。每个用户通过自己的终端独立使用计算机,都感觉自己一个人独占计算机资源。

(4)及时性:系统能及时响应用户的请求,这里的及时是指用户的请求在用户能接受的时间范围内得到响应,这一范围通常为2~3秒。

分时和多道程序设计引发了操作系统中的许多问题。如果存储器中有多个作业,必须保护它们不互相干扰,例如不会修改别人的数据。对多个交互用户,必须对文件系统进行保护,只有授权用户才可以访问某个特定的文件。还必须处理资源(如打印机)争用的问题。这些问题和其他问题以及可能的解决方法,在本书的后面部分会一一阐述。

1.4.3 实时系统

实时就是“立即”或“及时”,具体的含义是指系统能够及时响应随机发生的外部事件,并以足够快的速度完成对事件的处理。实时系统的应用环境是需要对外部事件及时响应并处理的那些场合。按使用场合和作用,可以分为两类:

(1)实时控制系统:把计算机用于对飞行器、导弹发射等军事方面的自动控制,这时计算机要把测量系统测得的数据及时加工并及时输出结果,或对目标进行跟踪并向操纵人员显示结果。

它主要的特点是:与被控制过程的速度相比,其反应速度足够快,工作安全可靠,容错机制比较完善,即使系统中软硬件发生故障,系统也能安全运行,操作简便。

(2)实时信息处理系统:把计算机用于预订飞机票,查询航班、航线、票价等事宜时,或把计算机用于银行系统、情报检索系统时,都要求计算机能对终端发来的服务及时予以正确的回答。

实时系统的主要特点是即时响应和高可靠性。实时系统的响应时间一般在秒级、毫秒级甚至更小。因为在一些关键系统中(如飞机的自动驾驶系统),信息处理的延误往往会带来不堪设想的后果。而可靠性在实时系统中比在非实时系统中更重要。非实时系统中的暂时故障可以简单地通过重新启动系统来解决,但实时系统是实时地响应和控制事件,性能的损失或降低都可能产生巨大的灾难,从资金损失到毁坏主要设备甚至危及生命。因此,在实时系统中,往往都采用了多级容错措施来保障系统的安全性及数据的安全性。

另外,用户控制在实时操作系统中通常比在普通操作系统中更广泛。在典型的非实时操作系统中,用户或者对操作系统的调度功能没有任何控制,或者仅提供了概括性的指导,诸如把用户分成多个优先级组。但在实时系统中,允许用户细粒度地控制任务优先级是必不可少的。用户应该能够区分硬任务和软任务,并且在每一类中确定相对优先级。实时系统还允许用户指定一些特性,例如使用页面调度还是进程交换、哪一个进程必须常驻主存、使用何种磁盘传输算法、不同优先级的进程各有哪些权限等。

实时系统其他的例子还包括机器人、空中交通管制,下一代系统还将包括自动驾驶汽车、具有弹性关节的机器人控制器、智能化生产中的系统查找、空间站和海底勘探等。

通常把批处理系统、实时系统和分时系统称为操作系统的三种基本类型。一些计算机系统兼有它们三者或其中二者的功能。

表1-1从多路性、独立性、及时性、交互性和可靠性几个方面列出了批处理系统、实时系统和分时系统的不同之处。