

21世纪 高等学校本科系列教材

总主编 吴中福

操作系统原理

(13)

陈笑蓉 主编



重庆大学出版社

操作系统原理

陈芙蓉 主编

内 容 简 介

本书系统的介绍了操作系统的基本概念、工作原理以及实现技术,详细阐述了操作系统对处理机、内存、设备和文件等各种资源的管理方法,及其所支持的用户界面。全书共有9章,内容安排如下:

第1章介绍了计算机系统和操作系统的概念,通过学习可以了解操作系统在计算机系统中的地位、作用和功能;第2章叙述了作业管理的基本原理和用户与操作系统的接口;第3章介绍了进程管理与线程管理的有关概念和技术,讨论了进程的同步与互斥、进程间的通信等问题,通过学习可以了解操作系统是怎样组织进程的并发活动;第4章主要介绍处理机管理和调度策略,并讨论了有关算法的特性与评价指标;第5章介绍了存储管理,包括分区、分页以及分段和段页式管理虚拟存储管理等;第6章介绍了设备管理,阐述了操作系统对外围设备的启动和控制的工作原理和技术;第7章讲述了文件管理,主要介绍了文件的存取方法和文件的类型以及文件目录的查找方法;第8章介绍了死锁的基本概念,详细阐述了怎样预防和避免死锁的策略和实现方法;第9章简单介绍了UNIX操作系统,主要是对UNIX操作系统有一个整体的认识 and 了解。以上每章之后都有小结和习题,便于读者复习,训练读者从模仿走向创造的能力。

图书在版编目(CIP)数据

操作系统原理/陈芙蓉主编. —重庆:重庆大学出版社,2002.10

计算机专业本科系列教材

ISBN 7-5624-2346-6

I. 操... II. 陈... III. 操作系统(软件)—高等学校—教材 IV. TP316

中国版本图书馆CIP数据核字(2002)第058840号

操作系统原理

陈芙蓉 主编

责任编辑:梁涛 姚正坤 版式设计:梁涛

责任校对:任卓惠 责任印制:秦梅

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街174号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn (市场营销部)

全国新华书店经销

重庆璧山印务有限公司印刷

*

开本:787×1092 1/16 印张:12.5 字数:312千

2003年8月第1版 2003年8月第1次印刷

印数:1—6000

ISBN 7-5624-2346-6/TP·302 定价:18.00元

本书如有印刷、装订等质量问题,本社负责调换

版权所有 翻印必究

前言

操作系统是计算机系统的核心软件,《操作系统原理》是计算机专业的主干课程,也是理工科非计算机专业的软件技术基础的重要课程。对于从事计算机技术的工作者来说,必须了解操作系统的工作原理和主要技术,才能更好的发挥计算机系统的作用。

本书详细地阐述了操作系统的各种基本概念和原理以及在构造操作系统过程中可能面临的种种问题和解决方法。为了使读者能在较短的时间内掌握操作系统的工作原理和实现方法,本书力求表述准确、概念清楚、算法具体,通俗易懂。在介绍操作系统基本概念和原理的过程中,结合实用的例子,引导读者很自然的去思考和判断问题,并怎样去实现整体的构造与设计。当今,计算机系统以飞快的速度发展,不断有新的技术诞生,因此,在本书中我们还结合当今的操作系统设计原理,以及操作系统发展方向,重点讨论操作系统的基本原理,关键技术和整体结构。

本书内容丰富,深浅适度,安排合理,强调理论和实践的结合,例子可读性强,便于自学。既可作为大学计算机本科及相关专业教材,又适合计算机爱好者自学,还可作为有关工程技术人员的参考书。

全书由陈笑蓉担任主编,覃海生担任副主编。本书共有9章。第2章、第3章、第4章、第7章、第8章由陈笑蓉编写,第1章、第5章由覃海生编写,第6章由王翊编写,第9章由阿兹古莉编写。由于编者水平有限,书中难免出现错误和不妥之处,恳求读者批评指正,不胜感激。

主 编

2003年1月

目 录

第1章 计算机操作系统概论	(1)
1.1 什么是操作系统	(1)
1.2 操作系统的形成和发展	(7)
1.3 操作系统的类型	(10)
1.4 操作系统的功能	(14)
1.5 如何学习操作系统	(16)
小结	(18)
习题	(19)
第2章 作业管理	(20)
2.1 作业管理的基本概念	(20)
2.2 作业管理的基本功能	(21)
2.3 作业的状态及作业控制块	(23)
2.4 用户与操作系统的接口	(24)
小结	(26)
习题	(26)
第3章 进程管理	(27)
3.1 进程的引入和定义	(27)
3.2 进程状态及转换	(32)
3.3 进程的产生和终止	(34)
3.4 进程的描述	(35)
3.5 进程控制	(37)
3.6 进程互斥与同步	(43)
3.7 进程间通信	(53)
3.8 管道(pipe)	(59)
3.9 线程(thread)	(60)
小结	(62)
习题	(63)
第4章 处理机调度	(65)
4.1 调度的层次	(65)
4.2 调度的功能及时机	(66)

4.3	进程调度的方式	(67)
4.4	调度原则与评价指标	(68)
4.5	调度算法	(69)
	小结	(75)
	习题	(75)
第5章	存储管理	(77)
5.1	概述	(77)
5.2	分区存储管理	(80)
5.3	分页存储管理	(85)
5.4	分段存储管理	(90)
5.5	段页式存储管理	(92)
5.6	虚拟存储管理	(94)
	小结	(100)
	习题	(101)
第6章	设备管理	(102)
6.1	设备类型和设备管理子系统的功能	(102)
6.2	设备管理中的两种支撑技术	(104)
6.3	I/O 系统中数据的传输控制	(108)
6.4	设备分配	(113)
6.5	设备的启动及 I/O 中断处理	(118)
6.6	共享设备的驱动调度	(120)
6.7	虚拟设备	(123)
	小结	(126)
	习题	(127)
第7章	文件管理	(129)
7.1	文件与文件系统	(129)
7.2	文件的组织与存取	(131)
7.3	文件存储空间管理	(135)
7.4	文件目录管理	(137)
7.5	文件共享与安全	(140)
7.6	文件的使用	(145)
7.7	文件系统的层次结构	(146)
	小结	(150)
	习题	(150)
第8章	死锁处理	(152)
8.1	死锁的形成	(152)
8.2	资源分配图 RAG	(154)
8.3	死锁的排除方法	(155)

8.4 死锁的检测与解除	(161)
小结	(162)
习题	(162)
第9章 UNIX 操作系统	(165)
9.1 UNIX 操作系统的发展和特点	(165)
9.2 UNIX 的系统结构	(170)
9.3 UNIX 进程管理	(172)
9.4 进程调度	(177)
9.5 进程通信	(178)
9.6 UNIX 存储管理	(180)
9.7 UNIX 设备管理	(181)
9.8 UNIX 操作系统的文件管理	(186)
小结	(189)
习题	(190)
参考文献	(191)

第 1 章

计算机操作系统概论

1.1 什么是操作系统

1.1.1 计算机系统的层次结构

操作系统是现代计算机系统的重要组成部分,无论是巨型机、大型机、小型机、微型机以及计算机网络都配置有操作系统。那么,什么是操作系统呢?本章将从不同的角度介绍操作系统的基本概念,包括操作系统的地位与作用、操作系统的组成及其功能、操作系统的类型与服务接口、操作系统的特性、操作系统的硬件基础等,这些内容构成了操作系统的宏观面貌,是学习以后各章内容的基础。

一个完整的计算机系统是由硬件和软件两大部分组成的。硬件(即物理计算机)是系统的基本资源,其主要部件包括:中央处理机(CPU)、主存储器(简称主存或内存)、外部存储器(简称外存或辅存,包括磁盘和磁带)、终端(通常由键盘和显示器组成)、控制台以及字符打印机等。CPU和内存构成系统的主机,其他部件统称为外部设备(简称外设),或称为输入输出(I/O)设备。在硬件基础上配置的系统软件是对硬件功能的扩充和完善,一个实用的计算机系统必定配置有多种系统软件,它们可以分成控制管理类和应用服务类。应用服务类系统软件是各种系统应用程序,每一个都完成某种特定的功能以向用户提供某种例行服务或支持某种专门的应用需要,例如编译程序、汇编程序、编辑程序、诊断排错程序、数据库管理系统、计算机辅助设计系统等。不同的计算机系统可能配置或支持不同的系统应用程序,从系统资源的观点看,系统应用程序是计算机系统的软资源,它们以文件形式存储在外存中,当用户需要某种服务时,再把相应的系统应用程序调入主机执行。在所有的系统软件中,操作系统是惟一地控制和管理整个计算机系统的软件,它不仅管理所有的硬件资源,也管理所有的软件资源,软件资源不仅是指系统应用程序,也还包括以文件形式存放在外存的用户程序和数据。

如图 1.1 所示,如果把计算机系统结构按层次划分的话,硬件处于最底层,它只是“裸机”的形象,即不附加任何软件的物理计算机;操作系统是加在裸机上的第一层软件,它是对裸机功能的首次扩充,从而构成了一个比裸机功能更强、使用更方便的“虚拟计算机”,所谓虚拟是

指逻辑的而非物理的;所有系统应用程序以及更上层的用户程序都在操作系统虚拟机上运行,它们受操作系统的统一管理和控制,通过操作系统使用各种系统资源来完成特定的任务。从结构上看,操作系统是用户程序及系统应用程序与物理计算机之间的接口。用户程序是用户为解决某个特定应用问题而开发的程序系统。

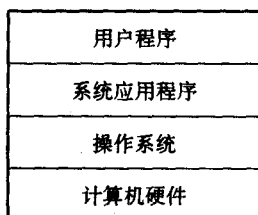


图 1.1 计算机系统的层次

可见,操作系统是计算机系统中的一个必不可少的关键组成部分,任何一个现代的计算机系统可以不配置某些系统应用程序,硬件的配置也各不相同,但决不可以没有操作系统。可以说,操作系统现在是计算机系统中的核心,将来也必然如此。

1.1.2 操作系统的定义

众所周知,在计算机系统中,中央处理器 CPU 是整个系统硬件的核心和基础,我们将它比喻为计算机的心脏。而在计算机软件系统中,操作系统具有同样的核心和基础作用,我们将它比拟为计算机的大脑。心脏和大脑的紧密配合,使计算机系统有了强大的生命力。而计算机硬件系统与软件系统的紧密配合,使计算机系统成为电子时代、信息社会无所不在、无处不在的基础设施。

操作系统(operating system,简称 OS)是计算机系统运行和工作必不可少的软件。无论是巨型机、大型机,还是中、小型机,也无论是台式个人计算机、便携式微型机,还是连接多台计算机的计算机网络,都离不开操作系统。

既然操作系统如此重要,那么它在计算机系统中处于一个什么位置?与硬件有什么关系呢?我们先看一幅示意图,它把计算机系统简单地分为4个层次,操作系统就是其中之一。

图 1.1 是操作系统位置的常规描述,从图中可以看到,操作系统介于计算机硬件和计算机用户之间,它与计算机硬件、软件、应用系统,以至计算机用户有千丝万缕的联系。这里,我们把一台没有任何软件配置和支持的计算机称为“裸机”,要让裸机接受用户发出的命令,执行相应的操作是非常困难的。操作系统在硬件之上建立了一个服务体系,为系统软件 and 用户应用软件提供了强大的支持,用户通过这个服务体系操作和使用计算机系统,面对一个非常友好、方便的环境界面。

计算机的硬件包括了组成计算机的集成芯片、电路板、电气连线、显示器、电源,以及各种物理部件,它们是看得见、摸得着的。但仅仅将这些部件连接起来并不能工作,还需要有直接控制部件和设备的微程序代码和基于硬件的机器级语言。这些机器级代码序列已经集成于处理器芯片中,或者固化于非易失性的存储器中,成为硬件系统的一部分,它们完成最低层的机器级操作,在机器中进行数据移动、数值计算、数据比较,并且将机器语言转换成一系列操作步骤和电子信号,去控制系统部件或者输入输出设备的操作。

在硬件之上就是操作系统,它是运行在计算机基本硬件系统上的最基本的系统软件。操作系统通过系统核心程序对计算机系统中主要的几类资源进行管理,如:处理器、存储器、输入

输出设备、数据与文档资源、用户作业等,并向用户提供若干服务。通过这些服务将所有对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。

计算机软件主要分为两类,一类是系统软件,另一类是应用软件。系统软件主要管理计算机系统本身的操作,而应用软件主要解决用户的问题。系统软件中最基本的就是操作系统,它为其他系统软件和应用软件提供运行支持,位于操作系统之上的是另外一些系统软件,如命令解释程序(或者称为外壳程序 shell)、编辑程序、编译程序、连接程序等,它们不是操作系统的一部分,但一般随着操作系统一起由计算机厂商提供,是系统开发中很重要和关键的一类软件。此外,系统实用程序、系统工具程序、系统调试程序也被认为是系统程序中的一部分,它们也常常采用套件的形式与操作系统一起提供。操作系统是以核心模式和超级管理模式运行的软件,它能防止用户对关键硬件和重要结构的有意无意的改动,而编译程序等运行于用户模式,可以随意替换和选择。

最上层是用户程序层。用户可直接通过系统应用程序层与计算机打交道,也可以建立各类应用软件和系统,通过它们来解决用户的具体实际问题。这些应用系统可以是用户自己开发的专用或者公用程序,例如事务处理、工程计算、娱乐游戏、教学训练等程序,也通过操作系统提供的支持和服务来使用系统资源,完成所进行的操作。

操作系统控制和管理着所有的系统硬件(如中央处理器、存储器、各种外部设备等),也控制和管理着所有的系统软件(如所有的系统程序和用户程序等)。操作系统对计算机使用者提供了一种良好的操作环境,也为其他各种应用系统提供了最基本的支撑环境。

(1) 对操作系统的认识

对操作系统的认识有两个方面,一是从用户的观点来看操作系统,二是从系统设计者(系统分析员)的观点来看操作系统。这两种看法反映了操作系统的抽象性和具体性,不同的出发点对操作系统的认识和理解有不同的程度与结果,反映了从外部还是从内部来看操作系统的不同的观点。认识的重点在于操作系统的功能、界面和结构。

1) 从外部来认识操作系统

从操作系统的外部来认识操作系统,长期以来有两种主要观点,它们表现了操作系统两种基本的服务功能,一种是虚拟机的观点,另一种是用户环境的观点。

①虚拟机的观点 虚拟机的观点是基于编程人员的观点,也称为扩展机器的观点,它是虚拟存储概念的自然扩充。从图 1.1 可知,装有操作系统的计算机极大地扩展了原计算机的功能,把用户面对的一个包含有各种硬件部件的计算机系统的操作和使用由复杂变得简单,从低级操作上升为高级操作,把基本功能扩展为多种功能。

众所周知,计算机系统的硬件结构和机器一级的操作包含了诸如指令集、存储器组织、总线结构和输入输出部件等的操作与控制,这些最基本的操作恰恰是最复杂和最难以由用户直接进行的。例如,用户要进行文件读写,而文件是以二进制代码的方式存放在磁盘、磁带等存储装置中的,需要有一种途径把用户的要求转换成对具体的硬件部件、电路信号、选择开关等的细微操作,用户自己不可能完成这些操作,但操作系统可以把用户的细微操作转换成一系列的低级操作,最终完成文件读写。所有的低级操作对用户来讲都是透明的,即无需用户关心的、看不见的操作系统把硬件全部隐藏起来,给用户提供了一个友好的、易于操作的界面。此外,操作系统还要进行大量的系统事务处理,如响应中断的发生、处理定时操作、管理存储器及其他低级操作,这样,操作系统对用户来说好像是一个扩展了的机器,即一台虚拟的机器(vir-

tual machine)。所以,有人把操作系统称为硬件系统的扩展,对用户而言,它比直接面对计算机硬件系统更容易操作了。

操作系统虚拟机在逻辑功能上与裸机不同,但物理实体是一样。即操作系统在裸机上运行,系统和应用程序在操作系统上运行,用户通过操作系统对机器进行操作。虚拟机的扩充包括了系统功能和数量上的扩充。在功能上,操作系统提供了一批功能很强的系统调用,用户通过命令和程序使用这些系统调用,命令在用户级使用,系统调用在程序级使用,它们构成了虚拟机的高级指令系统,相当于扩展了裸机的基本指令系统。在数量上的扩展,主要是对单机操作系统引入了多道程序技术,让多个用户同时使用一台机器,逻辑上每个用户都感觉有一台属于自己的机器。而对多任务系统,则使每个进程都独占一台虚拟机,每个用户不但拥有逻辑内存,而且还拥有逻辑 I/O 设备、逻辑文件以及其他逻辑资源。虚拟机给用户一个完全的硬件特性概念,一个本质上完全不同于实际机器的机器概念。虚拟机的硬件资源也同样在逻辑上从其他用户那里分离出来,每个用户似乎都拥有一台计算机,彼此间并不感到其他用户的存在,虚拟机与实际机器间的关系如图 1.2 所示。

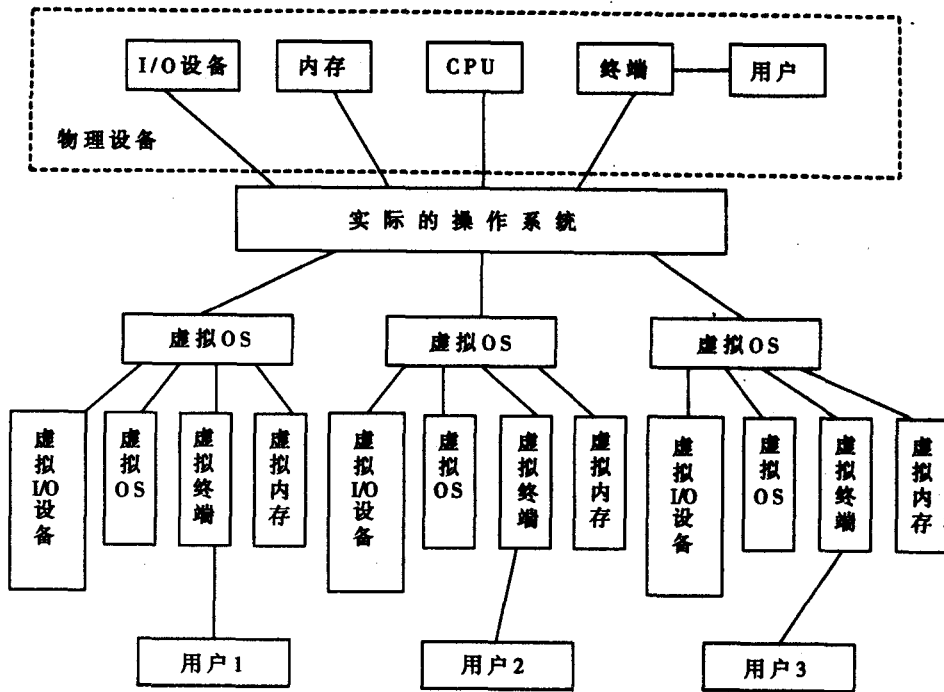


图 1.2 虚拟机与实际机器之间的关系

②用户环境的观点 从用户来看,操作系统是一个工作环境,是用户与计算机系统打交道的界面。用户使用操作系统为他们提供的各种服务、功能接口、系统调用等来使用计算机系统及其各类系统资源,用户不必知道各种系统资源的细节和控制过程。用户按“最适用的才是最好的”观点来选择和使用操作系统,而操作系统的结构及其发展也正是适应应用领域的需求而不断更新和发展的。

操作系统界面目前已成为系统的一个重要组成部分,用户和系统设计者都十分关注。这种界面已经从字符界面、菜单界面、窗口界面发展到目前的图形界面和多媒体界面,继而向 3D

界面进展。界面问题已经成为评价操作系统性能的一个重要因素,并且界面技术和支持工具也在不断更新,其原因是人们对操作系统的功能有了新的、更进一步的认识。早期由于硬件资源昂贵,操作系统的重点放在提高硬件资源的利用率上,而在硬件资源越来越丰富的今天,操作系统再一次考虑了如何有效地为用户服务。我们将在后面的章节中讨论用户界面。

2) 从内部来认识操作系统

从内部来看操作系统主要是看它的实体结构,即操作系统是如何组成、如何工作、如何对外提供服务的。它主要分为资源管理的观点和任务组织的观点。

①资源管理的观点 资源管理的观点是目前对操作系统描述的主要观点。资源是指计算机系统为了进行数值计算和数据处理所需要的各种物质基础,通常分为系统硬件资源和软件资源:硬件资源是组成计算机和计算机操作所需的物理实体,它是直接能为人们所感知的资源。如:中央处理器 CPU、主存储器、输入输出设备(显示器、磁盘驱动器、打印机)等;软件资源是依赖于一定的物理实体才能为人们所感知的一类资源。如:程序、数据、文档等,它们必须通过物理实体,如显示器或者打印机等设备展示出来,才能为人们所看到。上述资源中的很大一部分是可由计算机系统或者网络系统中的多个用户所共同享用的资源,如果多个用户都同时请求使用这些资源,就会发生冲突。而操作系统的工作就是当用户程序和其他应用程序争用这些资源的时候提供一种有序的和可控的分配,进而完成系统工作流程的控制。

在资源管理中,操作系统的任务是如何使系统资源得到充分合理的使用,解决用户作业因争夺资源而产生的矛盾。主要的功能是:

a. 监视资源 时刻保持系统资源分配的全局信息,了解系统资源的总数,已分配和未分配的资源情况,资源的增减和变动情况,通过系统中各类数据结构及表格记录资源的使用情况和环境状态。

b. 分配资源 处理对资源的使用请求,调解请求中的冲突,确定资源分配策略。当多个进程或者多个用户竞争某个资源时,进行仲裁判决。同时,根据资源分配的条件、原则和环境,确定是立即分配、暂缓分配还是不分配。对可以分配的资源,记录相应的分配情况,更新记录分配信息的数据结构和表格内容。

c. 回收资源 用户使用资源结束,提出释放请求,系统按照与分配过程相反的操作回收资源,更新相应的数据结构和表格内容。

d. 保护资源 对系统共享资源要进行保护,防止它们被有意或无意地破坏。在多用户系统中,对每一用户的数据及其程序运行进行保护,防止彼此间的干扰和冲突。对系统或者网络中重要的、敏感的(保密的)信息资源也进行保护。同时,防止非授权的操作,防止来自系统内部或者外部的人侵。

上述描述带有一种抽象性,实际上,从系统程序员的观点来看,它就是一个复杂的管理程序。它的组成是:

管理程序模块 + 算法 + 数据结构 + 事件驱动

操作系统程序是集中了多种功能的程序的集合,这些程序的运行基于各种数据结构,如表格、数组、记录、结构体、链表,也包括了对这些数据结构进行操作的算法实现。要完成资源管理,单一的程序是望尘莫及的。

操作系统的运行机制是事件驱动。上述各个功能模块要统一地控制系统工作流程,执行系统资源管理和分配,是通过系统的最小工作单位——进程来实现的,而代表进程存在、运行

和工作的实体,则是存在于操作系统模块内部的各种数据结构。实际上,操作系统在工作过程中,不断地接收来自系统内部和外部的各种事件引起的中断,根据不同的情况和请求,进行宏观的作业调度和微观的进程调度,完成系统资源的监测、分配和回收,协调系统工作流程。在这些过程中,它将不断地填写、改动、更新各个与任务和进程有关的表格,建立新的数据记录,使这些数据随时随地记录和代表进程的各种不同的状态。所以,有人说,操作系统的运行结构就是系统相关数据结构的反复地、不断地改动与更新。这些数据结构与表格的处理,是操作系统本身的开销,是操作系统设计过程中要考虑的一个重要问题。操作系统中所用的数据结构的合理选择,可以减少系统本身的开销,提高系统的吞吐率。

要了解资源管理,必须首先了解什么是系统的资源,读者也必须明白,今天的资源概念已经远远不是早期操作系统所涉及的资源的概念了。虽然本书仍然按照资源管理的观点来讨论操作系统的各个功能模块,但读者必须注意了解如下几个问题,即:管什么?管到什么程度?管理的方式如何?当管理的对象变了怎么办?因此,本书基于资源管理的讨论是一个基础,当今随着计算机网络的普及,系统资源(尤其是信息)逐渐从集中式向分布式、分散型转化,资源管理的模式也将随之变化。

②任务组织的观点 计算机系统要完成用户的各种各样的请求,这些请求通过作业或者任务的方式提交给系统,它们各自都是一次独立的计算过程,操作系统则是在这众多的计算过程中切换处理器。它是计算机系统工作流程的组织者,负责协调在系统中运行的各个任务的推进速度,提高系统的吞吐率。

由于系统中存在多道程序,要使程序间并发执行和资源共享,操作系统必须对用户作业和任务进行管理,要组织并控制系统工作的流程,因而设置了系统工作的基本单位,以这个基本单位为基础建立系统的数据结构、操作模式和管理程序。大多数操作系统引入了进程的概念,将其作为工作基本单位。随着操作系统的发展,根据系统工作的层次和并发控制程度,在操作系统中引入了作业、任务、进程、线程,以及对象等概念,对它们的组织和管理是操作系统的重点,我们将在后继章节中详细讨论。

操作系统的内部运行过程,也是进程状态不断变迁的过程。一个用户作业从外部提交给系统,就等待调度,等待被装入内存。一旦作业被装入,根据不同情况会创建若干任务或者进程,并等待分配处理器。一旦进程获得处理器就开始运行,运行中的进程要么正常运行结束而退出,要么被某个事件中断暂时停止,CPU 转去运行另一进程,如此不断,直至所有进程全部终止。操作系统也就不厌其烦地在这些状态之间切换,不断地更新表格中的内容,完成进程的管理和调度。

对操作系统运行的理解,读者必须与顺序程序执行予以区别。顺序程序模块的执行顺序是可知的,而操作系统的任何一个程序模块在什么时候运行是未知的,它需要由许多因素来决定。

综上所述,操作系统对计算机操作者和用户而言是一个用户环境,一个工作平台,是一个人与机器进行交互式操作的界面。对系统设计者来说是一种强功能的系统资源管理程序,一种包含有各种数据结构和算法的集成式软件系统。这两个方面、两种提法概括了操作系统的作用、实体和效果,即:

- a. 操作系统的作用是控制系统工作流程和资源分配管理;
- b. 操作系统的实体是程序和程序执行所需的数据及数据结构;
- c. 操作系统的效果是方便用户,提供一个良好的人机操作环境。

总之,操作系统是计算机系统中最基本、最核心和最重要的软件系统。在本书中,主要从资源管理的观点来学习和讨论操作系统。

(2) 操作系统的定义

随着计算机的发展,计算机系统的硬件和软件资源也愈来愈丰富。为了提高这些资源的利用率和增强系统的处理能力,最初出现的是监控程序。它是用户与计算机之间的接口,即用户通过监控程序来使用计算机。到了20世纪60年代中期,监控程序进一步发展,形成了操作系统。操作系统具有下述三方面的功能:①控制和管理计算机系统的硬件和软件资源,使之得到有效的利用;②合理地组织计算机系统的工作流程以增强系统的处理能力;③提供给用户一个好的操作界面,使用户能通过操作系统方便地使用计算机。因此,可以给出操作系统的定义:操作系统是直接控制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以及方便用户的程序的集合。

1.2 操作系统的形成和发展

1.2.1 无操作系统时期

计算机体系结构的发展是计算机系统分类划代的一种参照,操作系统与计算机体系结构的发展紧密相伴,虽然这种分类是粗略的和不够精确的,但是我们可以从中了解到操作系统的发展历史和功能变迁。这种参照是从电子数字计算机的诞生开始的。

20世纪40年代中期,数学家冯·诺依曼(John Von Neuman)提出了存储程序的概念,奠定了电子数字计算机的基础,成为电子数字计算机的历史开端。根据这种概念,先后设计建造了电子管数字计算机,其主要特点是:以电子管作为逻辑电路的主要器件,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算器为中心。这种计算机体积庞大无比,耗能严重,速度极慢,且价格昂贵,然而,它具有了程序的概念。用户在这些机器上的操作和编程,完全由手工进行,以机器语言形式(二进制代码形式)编程,采用接插板或开关板控制计算机操作,没有显示设备,由氖灯或数码管显示。在这个阶段,几乎没有程序设计语言,用户面对的是一个很不方便的操作环境。直到50年代初期,卡片穿孔成为程序编制和记录的方法,才形成一种可“阅读”的程序。这一时期(1946—1955年)被称为第一代计算机时期,此时的计算机主要用于科学计算,还仅仅是专家们的工具和宠物。在这一类早期的计算机系统中,有了程序,但没有操作系统。

1.2.2 早期的操作系统

晶体管的诞生使计算机产生了一次革命,20世纪50年代中期推出了晶体管计算机,它们被称为第二代计算机(1955—1965年)。其特点是:以晶体管及其他分立元件作为主要逻辑器件,主存储器以磁芯为主,外部存储设备除磁鼓、磁带外,引入了磁盘,因而重量有所降低,体积有所减少,耗电有所降低。然而,在计算机操作者和机器系统之间,仍然存在明显的分隔和困难的操作环境。在此期间,程序和编程语言得到较大的发展,由于有了计算机语言和相应的程序,就产生了对用户所提交的程序进行管理的程序,这就是监控程序(monitor)的雏形。虽然

此时的监控程序仅仅是处理用户的批量作业和简单的命令解释,但它毕竟建立了用程序来管理和控制其他的程序的方式。

在这个历史阶段,计算机的主要功能仍然是科学与工程计算,以及数据处理,由于设备价格昂贵,减少时间浪费是主要任务。于是,人们为计算机设计了批处理程序(batch program),并用这个程序来管理用户提交的其他应用程序。顾名思义,批处理就是集中处理一批用户提交的作业。例如,将一叠穿孔卡片,由一台专用机器顺序读入,并存入磁带卷,由机房操作员取下磁带卷,装入主计算机进行计算,磁带卷上的所有作业顺序地逐个执行,所有的计算结果也顺序存入另一磁带卷,再由操作员取下,送入另一台机器将结果逐一打印出来。可见,这本身就是一个顺序过程,一个烦琐而机械的过程。

这种批处理控制程序只是解决了作业间的自动转换,减少了时间浪费,尤其是主机 CPU 时间的浪费,但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作业非常庞大,它将独自一直占据 CPU,在它运行完毕之前,任何其他用户的作业,哪怕是短的作业也只有等待。所以,对批处理程序还需要进行改进,以适应越来越多的资源的管理,从而形成了系统资源管理的概念。对这一时期所采用批处理程序控制的计算机系统称为批处理系统,把早期的监控程序和这种批处理软件也称为早期操作系统。

1.2.3 完整的操作系统时期

(1) 中小规模集成电路,多道程序设计与完整的操作系统

计算机的发展进入 20 世纪 60 年代后,计算机的应用形成两个主要领域,即科学与工程计算和商务与数据处理。为解决面向这两个领域的问题,形成了通用计算机的概念。这个时期被称为第三代计算机(1965—1980 年)发展时期。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 System 360 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,才能发挥其强大的效能。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。初始的监控程序不能完成这些功能,于是产生了监管程序(supervisor),继而发展成为一个强功能的程序集合,即现在所说的操作系统。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。

在 IBM System 360 上运行的 OS/360 操作系统被认为是真正的(完整的)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来满足通用性,这就是多道程序设计技术,从而大大提高了 CPU 的利用率,但同时要求有专门的硬件机构来支持多道程序、支持存储器分块和防止作业间交叉混用。此外,磁盘存储装置和 CRT 显示装置的引入,改变了以往的脱机形式的批处理,成批作业的输入和输出都直接与磁盘打交道,磁盘成为作业的一个暂存缓冲区,批处理操作成为这个大型资源管理程序中的一项功能。这样,对系统不同资源的管理分工,使操作系统程序形成了相应的资源管理模块,建立了传统操作系统结构和层次的模型,这就是处理机管理、存储器管理、设备管理和文件管理。IBM 早

期的 SYS/38, AS/400 中仍然将操作系统称为系统控制程序 (CPF) 和系统扩展控制程序 (XPF), 可见, 它的功能仍是对系统工作流程及资源进行控制和管理。

随着电传打印机、键盘、显示器和显示终端等交互式计算机设备的问世, 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。面对交互式环境, 早期的批处理暴露了它的缺陷, 即批处理方式不适合交互式操作。为了改进系统响应时间和效能, 提供交互式操作环境, 导致了分时技术 (time sharing) 的出现。

在这一历史阶段中, 操作系统主要是基于多道程序的分时操作系统和多处理器操作系统, 也称为传统操作系统。后期的发展使操作系统具有了多种方式和类型, 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。

(2) 大规模集成电路, 现代操作系统

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、内存容量、外存容量、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, (personal computer 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 (1980—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, Linux, OS/2, WINDOWS, WINDOWS NT, WINDOWS XP, MAC 等, 其中有代表性的是 WINDOWS, UNIX 和 MS-DOS, 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。而且, 这些形形色色的操作系统在发展中也是彼此借鉴、相互补充、竞争发展的。

(3) 智能计算和网络计算, 新一代操作系统

目前, 虽然计算机技术仍在飞速发展, 新型机器层出不穷, 但一般认为它们仍属于第四代 (有人称为 4.5 代)。由于按照硬件分代的概念已经模糊, 分代已经无多大意义。所谓第五代计算机的特点, 一般认为是智能计算机。

智能计算机基于智能计算的概念, 力图使计算机具有与人类相当的智能, 并以人工智能的方式实现。然而, 要计算机真正完成人工智能方面的工作, 还有大量的工作要进行, 例如: 如何模拟人类智能的基本要素 (如感受、记忆、归纳、演绎、效应等), 如何模拟人脑神经元在存储和记忆方面的大容量、高可靠性、高随机性、多容错力、高协调力等功能。因此, 对并行分布、阵列存储、协作处理、容错组织、神经网络等方面的研究方兴未艾。其中, 网络计算 (network computing) 形成一大热潮。

网络计算的概念是基于多台计算机协作完成一项任务,共享系统所有资源(硬件、软件、计算、通信、信息等资源)。自20世纪80年代中期以来,计算机的互联成为高潮,形成了计算机网络,在网络互联和多机资源管理的基础上,网络上不同的体系结构,体现了不同的管理运作方式,出现了网络操作系统(network operating system,简称NOS)和分布式操作系统(distributed operating system)等新型操作系统。在这一历史阶段,计算机系统和操作系统的发展都异常迅速,操作系统技术逐渐成熟,新的技术不断引入,具有开放环境、高效数据处理、友好人机界面、强功能开发支持,以及网络互联与通信功能、多媒体处理功能等,形成了新一代的操作系统,人们对操作系统的认识也在不断深化、不断更新。

通过操作系统发展的历史,我们可以理解配置操作系统的目的,是为了有效地控制和管理计算机系统各种硬件和软件资源,使之得到更有效的使用。合理地组织计算机系统的工作流程,以改善系统的性能,为用户提供一个良好的工作平台,一个友好的人机交互环境。

1.3 操作系统的类型

1.3.1 分类的原则与观点

操作系统种类很多,也具有不同的特征,对操作系统的分类,可以基于不同的观点。

(1) 从计算机体系统结构的角度的分类

- 单机操作系统(单机单任务操作系统,单机多任务操作系统)
- 主从式多终端操作系统
- 多机多用户操作系统
- 客户/服务器结构操作系统
- 网络操作系统
- 分布式操作系统

(2) 从操作系统工作的角度的分类

- 单用户系统
- 批处理系统
- 分时系统
- 实时系统

(3) 从操作系统设计的角度的分类

- 面向进程的操作系统
- 面向对象的操作系统

需要特别注意,上述分类并非是一个很严格的界限,只是对操作系统进行描述、观察、分析的重点,也是对操作系统的结构和特点的重点描述。随着现代操作系统的发展,进程结构在不断地演化,在结构和功能上,各类操作系统形成了你中有我、我中有你的情况。例如,在一个操作系统中,既支持批处理,也支持分时或者实时操作。既具有进程结构,也具有对象结构,还可能具有支持多处理部件的任务结构。这是演化中的必然,初学者应当注意它的概念实质,而不仅仅是分类名称。