

操作系统

徐小青 主编



中国物资出版社

操作系统

徐小青 主编

中国物资出版社

图书在版编目(CIP)数据

操作系统/徐小青主编. —北京:中国物资出版社,19

98.2

ISBN 7-5047-1339-2

I. 操… I. 徐… III. 操作系统(软件) N. TP316

中国版本图书馆 CIP 数据核字(98)第 01837 号

责任编辑:印 丽

特约编辑:沙 金

装帧设计:木 贞

王 磊

责任校对:马永开

中国物资出版社出版发行

(北京市西城区月坛北街 25 号 邮编:100834)

全国新华书店经销

北京科发文化交流有限公司激光照排

安徽省蚌埠市红旗印刷厂

787×1092 毫米 1/16 印张:13.75 字数:343 千字

1998 年 1 月第 2 版 1998 年 2 月第 1 次印刷

印数:00001—10000 册

ISBN 7-5047-1339-2/TP·0012

定价:16.80 元

总 序

从1946年第一台计算机诞生至今已历经半个多世纪。计算机的出现和广泛应用,标志着人类社会的一次大飞跃,信息时代的一次大转折,生存方式的一次大变革,现代文化的一次大融汇。随着计算机技术的飞速发展和广泛普及,其应用已遍及社会生活的各个领域。由于计算机技术已进入到我们生活中的方方面面,人类社会的生活方式、思维方式以及时空观念等各个方面都已经发生了深刻的历史性的巨变。

随着信息化社会的发展,人们对信息交流的要求越来越高。世人已普遍公认:哪个国家的信息化程度高,其经济竞争力就越强,科技发展就越快,办事效率就越高,对下一代的教育条件就越好。信息化的进一步深入需要我们培养一大批高素质人才。当今社会,熟练掌握计算机应用技术已成为高素质人才的必备条件。因此,一个十分艰巨的任务,就是要使受教育者具有使用计算机的能力和与之相适应的计算机文化素质。如果我们的知识结构和文化修养准备不足,就不能适应时代和社会发展的需要。

一本好书,是人生旅途的一掬甘泉;一套好教材,是教学成功的必要条件。广大学子和读者殷殷所望,无非是博得一艺,学有所用。本着对读者负责的精神,我们组织北京电子科技学院、集美大学、合肥工业大学、安徽大学、安徽财贸学院、江苏广播电视大学等高等院校和部分大中专学校教学经验丰富的教师,以及一些具有较高理论基础和软件开发经验的计算机技术人员共同合作编写了这套计算机及应用专业教材。为保证教材的质量,我们还聘请了一批学术造诣较深的专家、教授作为本套教材的主审和顾问。本套教材具有以下几个方面的特点:

首先,作为一套计算机专业教材,必须保证整个计算机知识体系的完整性。本套教材包括必修课17种,选修课和配套教材3种,基本上涵盖了目前大中专院校计算机及应用专业所必修或选修的课程内容。各种教材在编排上,既注意到内容上的连贯性,又保证了教学上的相对独立性。

其次,在教材内容的组织上,注重介绍和汲取当今计算机领域的一些新技术和新知识,摒弃了传统教材中一些过时的内容,这些变化在各本教材中都得到程度不同的体现。本套教材编写时既参照了有关部委计算机及应用专业教学大纲,又参考了“程序员考试大纲”和“全国计算机水平等级考试大纲”的内容。因此本套教材既适合作为各级各类院校计算机及应用专业教材,亦可作为计算机水平等级考试学习用书。个别教材之间内容上的重复,是为了照顾部分读者单独选用单本教材的需要,敬请广大读者予以谅解。

再次,考虑到各校教学的特点,本着学以致用原则,在本套教材编写中我们始终贯彻“由浅入深,理论与实践相结合”的原则,以阐明要义为主,辅之以必要的例题、习题和上机实习,以便使读者尽快领悟和掌握。

在本套教材编写过程中,各位作者付出了艰辛的劳动,教材编委会的各位专家和教授对本教材的内容进行了认真的审定和悉心的指导。在本套教材出版过程中我们自始至终得到中国物资出版社领导和编辑以及印制单位的大力支持和帮助。本套教材承蒙中国科学院计算技

术研究所、国家智能计算机研究开发中心王川宝、高文、中国机械科学研究院江波等同志进行了较为细致的终审终校工作。正是由于各方面的通力配合,才使得本套教材得以顺利出版和发行。书中参考、借鉴了国内外同类教材和专著,在此一并表示感谢。

近年来,计算机技术发展日新月异,异彩纷呈,许多新的概念和内容都在不断扩展之中,囿于编者学识和水平,书中疏漏、错误之处还望广大读者不吝批评指正,以便对本套教材不断修订完善。

计算机及应用专业教材编委会

1998年1月

附:计算机及应用专业教材编委会名单

顾 问

(以姓氏笔划为序)

王仲文	北京电子科技学院院长、教授
韦 穗	安徽大学副校长、教授
张全寿	铁道部电子计算中心主任、北方交通大学教授
李文忠	全国计算机基础教育学会副理事长、东南大学教授
杨善林	合肥工业大学副校长、教授、博士生导师
辜建德	集美大学校长、教授
魏余芳	西南交通大学教授

编 委

鄂大伟	集美大学副教授
李树德	北京电子科技学院教授
刘 锋	安徽大学副教授
王川宝	中国科学院计算技术研究所、国家智能计算机研究开发中心 硕士研究生
高 文	中国科学院计算技术研究所、国家智能计算机研究开发中心 博士研究生
江 波	中国机械科学研究院硕士研究生
屈道良	上海铁路局蚌埠分局高级工程师
蒋翠清	上海铁路局蚌埠分局高级工程师

前 言

操作系统是现代计算机系统中不可缺少的基本系统软件,也是其它系统软件及应用软件的支撑系统和操作平台。因此,计算机操作系统不仅是大、中专院校计算机专业的必修课程,而且是一般从事计算机应用和开发人员所必不可少的基础知识。为此,我们编写了这本《操作系统》教材。

在内容安排上,我们将全书分为两大部分:第一部分主要介绍了操作系统的原理。第一章首先介绍了操作系统的基本概念、基本类型、基本特征和主要功能,以便使读者对其有一个初步的认识;第二章至第六章分别介绍了作业管理、进程管理、存贮管理、设备管理和文件管理。鉴于本书读者已具有使用计算机的经验,为此从使用的角度来讨论操作系统更容易入门,因而,我们先从作业管理入手来讲解操作系统,以激发学生的学习兴趣;第二部分主要介绍了操作系统的实例。考虑到当今领导操作系统潮流的是UNIX操作系统,以及在个人计算机上占有重要位置的 Windows NT 操作系统,因而我们将其选做本书实例加以介绍,并全面深入、系统地介绍了它们的原理和构成,以使读者对当今流行的操作系统有一全面的了解。

本书是编著者根据多年来教学经验和科研经验,依据有关部委教学大纲编写而成的。在编写过程中,作者参阅了国内外操作系统方面的著作、教材,以及有关操作系统发展的最新技术资料,并考虑到当代大、中专学生的学习特点,在编写中力求使内容由浅入深,文字通俗易懂,并且每章均配有习题,以帮助读者检验和加深对书中内容的理解,便于讲授和自学,这是本书的主要特点;此外本书较好地做到了基本原理与典型实例的结合。对不同类型的读者,在教学上可以作不同的处理:只讲原理、不讲实例,或原理和实例都讲。

本书第一、二、三、四、六章由北京电子科技学院徐小青同志编写,第五、七、八章由北京电子科技学院王景中编写。本书在编写过程中得到了安徽财贸学院马永开副教授的热情指导,并认真补充、修改了书稿,最后又不辞辛苦地审定了全书。书中参考引用了国内外有关文献,在此一并表示衷心地感谢。

由于编著者水平有限,编写时间仓促,书中难免存在不少错误和不妥之处,恳请广大读者不吝批评指正,以便不断修订完善。

编著者

1998年2月

于北京电子科技学院

目 录

第一章 操作系统概述	(1)
§ 1.1 什么是操作系统	(1)
§ 1.2 操作系统的形成	(2)
§ 1.3 操作系统的特征与功能	(4)
§ 1.4 操作系统的类型和发展	(6)
§ 1.5 操作系统的运行环境	(15)
本章小结	(17)
复习思考题	(18)
第二章 作业管理	(19)
§ 2.1 作业的基本概念	(19)
§ 2.2 作业的管理与调度	(24)
§ 2.3 用户与操作系统之间的接口	(28)
本章小结	(31)
复习思考题	(32)
第三章 进程管理	(33)
§ 3.1 进程的基本概念	(33)
§ 3.2 进程的状态	(39)
§ 3.3 进程的描述	(41)
§ 3.4 进程控制	(44)
§ 3.5 进程互斥	(52)
§ 3.6 进程同步	(58)
§ 3.7 进程通信	(62)
§ 3.8 进程调度	(65)
§ 3.9 死锁	(68)
本章小结	(71)
复习思考题	(72)
第四章 存储器管理	(74)
§ 4.1 概述	(74)
§ 4.2 覆盖与交换	(79)
§ 4.3 单道环境下的存贮管理——单一连续区分配	(80)
§ 4.4 分区存贮管理	(82)
§ 4.5 分页存贮管理	(88)
§ 4.6 分段存贮管理	(94)
§ 4.7 段页式存贮管理	(101)
本章小结	(105)
复习思考题	(106)
第五章 设备管理	(107)
§ 5.1 引言	(107)
§ 5.2 数据传送控制方式	(109)

§ 5.3 中断技术	(116)
§ 5.4 缓冲技术	(118)
§ 5.5 设备分配	(120)
本章小结	(125)
复习思考题	(126)
第六章 文件系统	(127)
§ 6.1 引言	(127)
§ 6.2 文件与文件系统	(128)
§ 6.3 文件的逻辑结构与存取方法	(130)
§ 6.4 文件的物理结构与存贮设备	(131)
§ 6.5 文件存贮器存贮空间的管理	(136)
§ 6.6 文件目录管理	(139)
§ 6.7 文件的存取控制	(145)
§ 6.8 打开文件表	(147)
§ 6.9 文件的使用	(148)
§ 6.10 文件系统的层次模型	(150)
本章小结	(152)
复习思考题	(153)
第七章 UNIX 操作系统	(154)
§ 7.1 UNIX 操作系统概念	(154)
§ 7.2 UNIX 操作系统的结构	(157)
§ 7.3 UNIX 系统内核的体系结构	(158)
§ 7.4 进程管理	(159)
§ 7.5 存贮器管理	(166)
§ 7.6 设备管理	(172)
§ 7.7 文件管理	(174)
§ 7.8 UNIX 系统中的 pipe 机构	(177)
§ 7.9 系统调用	(177)
§ 7.10 UNIX 操作系统中的 shell	(179)
§ 7.11 UNIX 系统的初步使用	(183)
第八章 Windows NT 操作系统	(187)
§ 8.1 Windows NT 操作系统简介	(187)
§ 8.2 NT 的系统结构	(188)
§ 8.3 Windows NT 基元成分	(193)
§ 8.4 微内核	(196)
§ 8.5 虚拟存储管理	(198)
§ 8.6 NT 的 I/O 系统	(201)
§ 8.7 NT 内装网络	(202)
主要参考文献	(209)

第一章

操作系统概述

计算机发展到今天,从个人计算机到巨型计算机系统,都毫无例外地需要配置一种或多种操作系统。那么何谓操作系统呢?它具有哪些功能呢?它在计算机系统中起到什么作用?它有哪些类型?我们将在本章中作一简要阐述。

§1.1

什么是操作系统

从1946年第一台ENIAC计算机诞生到今天,计算机经历了日新月异的飞速发展,不但有了个人计算机(微型机)、小型机、中型机、大型机和巨型机,而且在功能上也从单一的数值计算进入到数字、文字、图像和声音等信息的一体化处理,使得当今计算机技术已在各行各业各个领域得到广泛应用,并已进入家庭化。可以说计算机技术已成为现代社会发展水平的一个重要标志之一。

计算机系统,即使是最简单的系统,实际上也是十分复杂的。用户使用计算机解题不但要使用硬件,还要涉及软件系统,用户直接使用系统的硬、软件资源解决应用问题几乎是不可能的,因为需要掌握硬、软件系统的每个细节,即使这样,在多个用户共用一台计算机时还会互相干扰,发生资源使用上的冲突,有效地利用系统资源,也就更谈不上了。为此,计算机科学工作者为了使用户能从复杂的计算机资源使用方法中解脱出来,专心致力于解题方法本身的研究,并且提高考虑资源的使用效率和安全性而设计了这样一组程序模块:它们统一管理和控制系统资源,合理组织工作流程,有效地利用系统资源,为用户提供一个功能强而且使用方便的工作环境,在用户和计算机系统之间起着接口的作用,这些程序模块统称为操作系统。它是计算机系统中必不可少的重要组成部分。

没有任何软件支持的计算机称为裸机(Bare Machine),它仅仅构成了计算机系统的物质基础,而实际呈现在用户面前的计算机系统是安装有若干层软件的计算机。图1-1展示了这种情形。

由图可见,计算机的硬件和软件、部分软件之间是一种层次结构的关系。裸机在最里层,它的外面是操作系统,操作系统提供的资源管理功能和方便用户的各种服务功能把裸机改造成成为功能更强、使用更为方便的机器,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持环境,同时又向用户提供完成其作业所需的各种服务。

因此,引入操作系统的目的可以从以下两个方面来考察:

1. 从系统管理人员的观点来看,引入操作系统是为了合理地去组织计算机工作流程,以增强系统的处理能力;管理和分配计算机系统硬件及软件资源,使之得到充分有效的利用并能为

多个用户所共享。因此,操作系统是计算机资源的管理者。

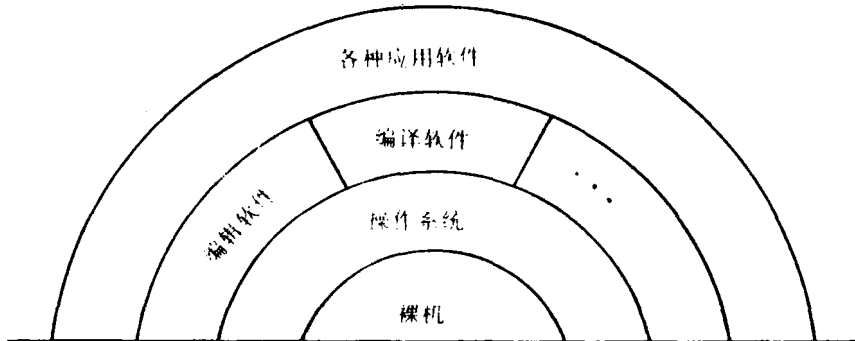


图 1-1 计算机系统层次结构

2. 从用户的观点来看,引入操作系统是为了给用户使用计算机提供一个良好的界面,以使用户无需了解许多硬件和系统软件的细节,就能方便灵活地使用计算机。

综上所述,我们可以把操作系统描述为:操作系统是计算机系统中的一个系统软件,它负责管理与控制计算机系统内的各种软、硬件资源,合理地组织计算机工作流程,以便有效地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一个友好的接口作用。

§ 1.2 操作系统的形成

为了更好地理解操作系统的基本概念、功能和特点,让我们首先回顾一下操作系统是如何形成的。

操作系统是在人们不断地改善计算机系统性能和提高资源利用率的过程中,逐步地形成和发展起来的。它的功能由弱到强,在计算机系统中的地位不断提高。至今,操作系统已成为计算机系统的核心,任何计算机系统都配置有操作系统。

1.2.1 手工操作阶段

在第一代计算机时期,计算机硬件大多数由电子管构成,计算机运算速度慢(只有几千次/秒)且内存容量小。没有操作系统,仅有汇编语言及少量的服务程序,基本上是采用人工操作方式。操作人员将纸带(或卡片)装进纸带输入机(或卡片输入机),以便把程序和数据输入计算机,然后通过控制台开关启动程序,仅当程序执行完毕,由用户取走纸带与计算结果之后,才能让下一个用户上机操作。这种人工操作的方式具有以下特点:

(一) 用户独占计算机

一台计算机为一个用户所独占,系统中的全部资源由他一人支配,此时用户可以方便地使用各种资源,而不会出现某资源已被其它用户占用而需要等待的现象,但此时资源的利用率也是非常低的。

(二) CPU 等待人工操作

用户仅在上机时才能将纸带或卡片装入相应的输入设备,此时 CPU 空闲等待。当计算完毕,进行卸带或取卡操作时,CPU 又空闲,可见,CPU 的利用率也是非常低的,这在运行计算量小而输入/输出量大的作业中显得更为突出。

50年代后期,计算机的规模不断扩大,其运算速度也有了很大的提高,从几千次/秒发展到几十万次/秒或上百万次/秒。这样手工操作的低速和计算机的高速之间形成了尖锐矛盾,手工操作与计算机有效运行时间之比逐渐地增大,唯一的解决办法是:摆脱手工操作,实现作业的自动过渡,这样就出现了批处理方式。

1.2.2 早期批处理技术

早期批处理系统是把若干作业合成一批,用一台或多台小型的卫星机把这批作业输入到磁带上,然后再把这盘磁带装到主机的磁带机上,由主机的监督程序——最早的操作系统雏型——把磁带上第一个作业调入内存中执行,该作业终止后(正常完成或非正常终止),再依次调下一个作业到内存中执行,依此下去,直到该批作业全部执行完毕为止。

到了60年代初期,计算机硬件取得了重大进展,出现了通道和中断技术。所谓通道是专门用来控制输入/输出设备的处理机,称为输入输出处理机(简称I/O处理机)。通道比起CPU来说,一般速度相对较慢,价格也较便宜。它可以与CPU并行工作。这样当需要传输数据时,CPU只需构造好通道程序,然后启动通道就可以了,由通道执行相应的通道程序来控制数据的传输,当通道完成数据传输工作后,用中断机构向CPU报告完成情况,这样就可以把原来由CPU直接控制的输入/输出工作转移给通道,使得宝贵的CPU主要用于数据处理工作,做到通道和CPU的并行操作,提高了系统效率。

但是,这时计算机系统运行的特征是单道顺序地处理作业,即用户作业仍然是顺序处理,此时的计算机系统是单道批处理系统。在单道批处理系统中可能会出现以下两种情况:对于以计算为主的作业,输入输出量少,通道和外设空闲;然而对于以输入/输出为主的作业,主机又会造成空闲。这样总的说来,计算机资源使用效率仍然不高。因此,为了充分发挥计算机各个部分并行工作的能力,操作系统进入到多道程序设计阶段,即多道程序合理搭配交替运行,充分利用系统资源,提高系统效率。

1.2.3 多道程序系统

所谓多道程序系统是指在内存中同时存放几个作业,并允许它们交替执行,共享系统中的各种软、硬件资源。当一道程序因I/O请求而暂停执行时,CPU便立即转去执行另一道程序,从而保证CPU以及系统中的其它设备得到尽可能充分的利用。

多道程序系统虽然提高了系统的吞吐量有效地改善了系统资源利用率,但需要解决以下一些技术:

(一)内存中的多道程序要共享计算机系统的各种软、硬件资源,既可能有资源竞争,又可能需要相互同步,因此互斥与同步机制成为操作系统设计中的重要问题。

(二)在内存中同时驻留了多道程序,应为每道程序分配好内存空间,使它们各得其所,不会因相互重叠而丢失信息;又应防止某道程序由于人为的因素或出现异常情况而破坏其它程序。

(三)在单处理机系统中,处理机只有一个,且它为各道程序所共享,这必将引起各道程序对处理机的争夺。系统应协调这些程序对处理机的使用,使各道程序最终都能获得处理机而运行。对于紧急事务,还应能使之优先获得处理机。

(四)通常系统把大量有意义的信息以文件的形式存放到各种存贮介质上。为提高存贮空

间的利用率,加速对信息的检索速度,系统应对它们进行组织和管理,并且为了方便用户,系统应提供存贮和检索文件信息的手段。

为了解决上述问题,在多道程序系统中,必须设置一组有机结合的软件,此外,还应提供方便用户使用计算机软件,这样便形成了操作系统。此时,操作系统的功能已变得十分丰富而完整。

综上所述,操作系统的发展是由手工操作阶段过渡到早期单道批处理阶段而初具雏型,而后发展到多道程序系统时才逐步完善的。

§ 1.3 操作系统的特征与功能

1.3.1 操作系统的特征

现代操作系统应具有以下特征:

(一)并发性

并发性是操作系统最基本的特征之一。

并发性与并行性是既相似而又有区别的两个概念,并发含有“同时进行”的概念,它与并行的区别在“时间”和“时刻”上。并发性是指两个或多个事件在一段时间间隔内同时发生;而并行性是指两个或多个事件在同一时刻发生。在多道程序环境下,并发性是指宏观上在一段时间间隔内有多道程序在同时运行,但在单处理机系统中,每一时刻仅能有一道程序在运行,故微观上这些程序是交替顺序执行的。能并发执行的程序称为并发程序,相应的系统称为并发系统。程序的并行执行是指在同一时刻这些程序都在真正地执行,显然,这需要有相应的系统硬件和软件的支持。例如,两道程序并行执行应该是两道程序分别在两个处理机(处理器)或两套处理部件中独立执行。

现代操作系统中普遍采用的是程序的并发执行,它有效地改善了系统资源的利用率,提高了系统的吞吐量,但它也使整个系统复杂化了。操作系统必须具有控制和管理各种并发活动的的能力。

(二)共享性

共享性也是操作系统最基本的特征之一。

共享性是指系统中所有各种软、硬件资源不再为某个程序所独占,而是被多个用户同时使用。

由此可见,并发性和共享性是互为存在条件的。一方面,资源共享是以程序的并发执行为条件的,若系统不允许并发执行,自然也就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,也必将影响到程序的并发执行,甚至根本无法并发执行。

(三)虚拟性

在操作系统中,所谓虚拟是指把一个物理上的实体,变为若干个逻辑上的对应物。前者是实的,即实际存在的;而后者是虚的,只是用户的一种感觉。例如,在分时系统中,虽然只有一个CPU,每次只能执行一道程序,但通过分时使用,在一段时间间隔内,宏观上这台处理机能同时运行多道程序。它给用户的感觉是每道程序都有一个CPU在为其服务。

(四)不确定性

在多道程序环境下,每个程序在何时执行、多个程序间的执行顺序以及完成每道程序所需的时间都是不确定的,因而也是不可预知的,换言之,多道程序环境下程序的执行是以异步方式进行的。然而,无论程序以何种顺序、速度向前推进,只要在相同的运行环境下给予相同的输入,其运行结果总是确定的。

1.3.2 操作系统的功能

现从资源管理和用户接口的观点出发,从以下五个方面来说明操作系统的基本功能。

(一)处理机管理

由于计算机系统的心脏是处理机,所有软硬件操作都必须由处理机分解执行。在单处理机的系统中,存在着用户作业争用处理机的情况。如何对使用处理机的请求做出适当的分配,这就是操作系统处理机管理功能模块所要解决的问题。在实际工作中,操作系统是以进程和作业的方式进行管理,完成作业和进程的调度,分配处理机时间,控制作业和进程的执行。

(二)存贮器管理

在计算机系统中,存贮器(一般称为主存或者内存)是程序运行、中间数据和系统数据存放的地方。存贮器管理的主要工作是对内存进行分配、保护和扩充。

1. 内存分配

在内存中除了操作系统、其它系统软件外,还要有一个或多个用户程序。如何分配内存,以保证系统及各用户程序的存储区互不冲突?这就是内存分配所要解决的问题。

2. 存贮保护

系统中有多个程序在运行,如何保证一道程序在运行过程中不会有意或无意地破坏另一道程序?如何保证用户程序不会破坏系统程序?这就是存贮保护问题。

3. 内存扩充

当用户作业所需要的内存量超过计算机系统所提供的内存容量时,如何把内存和外存结合起来管理,为用户提供一个容量比实际内存大得多的虚拟存贮器,让用户使用这个虚拟存贮器和使用内存一样方便,这就是内存扩充所要完成的任务。

(三)设备管理

设备管理的主要任务如下:

1. 按照一定的分配方式、分配策略,把通道、控制器和输入/输出设备分配给请求输入/输出操作的进程,并启动设备完成实际的输入/输出操作。为了尽可能发挥设备和主机、设备与设备的并行工作能力,还需要采用虚拟技术和缓冲技术。

2. 解决设备的无关性。输入/输出设备种类很多,使用方法也各不相同。因此,设备管理应为用户提供一个使用设备的良好界面,而不必去涉及具体的设备特性,以便使用户可以方便、灵活地使用这些设备。

(四)文件管理

上述三种管理都是针对计算机的硬件资源的管理。文件管理,则是对系统的软件资源的管理。

计算机要处理大量的数据,这些数据是以文件的方式存贮在外存贮器(如磁盘、磁带、光盘等)中,对这些文件如不能很好地管理,就会引起混乱,甚至遭受破坏。因此对文件的管理、对存

放文件的外存空间的管理就成为文件系统管理的问题。另外,信息的共享、保密和保护,也是文件系统所要解决的问题。

(五)用户接口

前述的四项功能是操作系统对资源的管理。除此之外,操作系统还必须为用户提供一个良好的人机交互界面,用户通过命令操作和程序操作与计算机交互,而交互的环境界面将对用户产生极大的影响,包括心理上和思想上的影响。用户管理模块则应极力解决用户操作问题,使计算机系统的使用更方便、更简洁、更实用。

一般说来,操作系统提供以下两种方式的接口来为用户服务。

一种用户接口是程序一级的接口,它通过一组系统调用(或称广义指令)供用户程序和其它系统程序调用。当用户程序要求进行数据传输、文件操作或使用其它资源时,通过这些系统调用向操作系统提出申请,由操作系统来完成。

另一种用户接口是用户(作业)一级的接口,它提供一组控制操作命令或称作业控制语言,供用户去组织和控制自己作业的运行。作业控制方式被典型地分为两大类:脱机控制和联机控制。相应地,操作系统提供脱机控制命令和联机控制命令。

此外,随着计算机硬件技术的发展和应用领域的拓宽,现代操作系统还对网络与通信资源、安全机制与设施资源、多媒体资源等的管理进行了新的功能扩充。

§ 1.4 操作系统的类型和发展

迄今为止,操作系统的诞生已有 30 多年。在此期间,随着计算机技术和软件技术的蓬勃发展,相应地也就出现了各式各样的操作系统,以满足不同的应用要求。对操作系统可采取多种方法进行分类,最常用的方法是按功能特征的不同而把操作系统分为以下三种基本类型:

1. 批处理操作系统 Batch Processing Operating System
2. 分时操作系统 Time Sharing Operating System
3. 实时操作系统 Real Time Operating System

随着计算机体系结构的发展,又出现了许多新型的操作系统,如下:

1. 微型机操作系统;
2. 多处理机操作系统;
3. 网络操作系统;
4. 分布式操作系统。

下面将对它们作一简要介绍。

1.4.1 批处理系统

批处理系统一般分为单道批处理系统、多道批处理系统和远程批处理系统。它们都是成批处理或者顺序共享式系统,它允许多个用户以高速、非人工干预的方式进行成组作业的执行。批处理系统将作业组成批地提交给系统,由计算机按顺序自动完成后再给出结果,从而减少了用户建立和打断的时间。

(一)单道批处理系统

这是一种在早期计算机系统中所配置的操作系统。其特征如下:

1. 无须作业调度和进程调度。只按作业在外存中排定的顺序,依次将它们调入内存,将处理机分配给调入的作业使之运行;

2. 内存中仅有一道作业在运行。仅当前一道作业运行完成,或出现不能处理的异常情况时,才新调其后的作业进入内存运算,而内存中始终保持一道作业;

3. 作业完成的顺序与作业进入内存的顺序直接相关,即先进入内存的先完成。

(二)多道批处理系统

这是在 60 年代中期以后经常配置的一种操作系统类型。它是基于多道程序设计技术的应用而配置的,也就是说,在内存中可以同时存放多个用户作业,允许这些作业在系统中交替地运行。在多道批处理系统中,用户可以通过系统提供的各种功能,如作业控制语言、命令、程序等,将用户程序、数据等一一提交给系统,在外部存贮设备上形成作业的后备等待队列。用户一旦将程序和数据提交给系统之后,便失去与该作业的交互能力,由系统根据一定的作业调度原则从后备队列中选取若干作业调入内存。在内存中的这些作业按照多道方式组织它们的运行,某一道作业运行完毕或暂停运行,系统又将调入新的作业,内存中始终存放多个作业,它们交替运行。这样,作业不断进入系统,又不断退出系统,形成源源不断的作业流,从而大大地提高了系统资源的利用率和系统的吞吐率。

从以上分析可以看出,多道批处理与单道批处理的主要区别在于以下几个方面:

1. 在作业管理中必须要有作业调度功能和进程调度功能。从宏观上看,作业调度从外存后备队列中选择一个或若干个作业,为其分配必要的资源(如内存),使其进入执行状态;从微观上看,进程调度从内存中的多个作业中选择一个,使其获得处理机的真正执行;

2. 多道批处理在内存中同时存放多道作业,宏观上这些作业同时都在运行,而每一时刻则只有一道作业在运行;

3. 作业完成的顺序与它们进入内存的顺序无严格的对应关系,后进入内存的作业可能先完成。

(三)远程批处理系统

这是配置在联机系统和计算机网络中的,能接收从远程系统送来的批量型作业,对它进行处理后,将结果送至指定系统的一种批处理系统。该系统是在通常的批处理系统基础上,再加上远程作业录入程序后形成的。远程作业录入程序具有接收远地作业,并将它排入本地处理作业队列中的功能,以后便像本地作业一样地处理。远程批处理的功能十分有用,它是构成分布式批处理系统的基本条件之一。

批处理系统的主要优点是:系统的吞吐量大,资源利用率高,系统开销较小。其缺点是:一是作业处理的平均周转时间长(所谓作业周转时间是指从作业进入系统开始到作业完成所经历的时间),在批处理系统中,一个作业进入系统,必然使系统内许多作业的周转时间增长;二是用户交互能力差。用户将作业提交给系统后,直到该作业运行结束,都无法与自己的作业进行交互,因此,用户必须提供一份作业说明书,由系统根据该作业说明书对作业加以控制,这必将会给程序的修改和调试带来不便。

1.4.2 分时操作系统

分时系统也是一类多道程序系统。它是用一台功能很强的计算机连接多个终端(几十台,上百台,甚至上千台终端),提供多个用户同时上机操作的系统。每一个用户占据一个终端,并

通过自己操作的终端,把程序和数据送入主机进行运算,主机采用时间分片的方式——即分时方式轮流为每个与之相连的终端用户服务,并及时对用户的服务予以响应,使得每一个用户都感觉到是一台计算机在专门为他服务,即独占计算机一样,这就是分时系统的概念。

分时系统是将主机 CPU 的运行时间分割成很短的时间片(时间片可以是固定的,也可以是可变的),按时间片依次轮流把 CPU 分配给与之相连的各个终端用户使用。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 CPU 让给另一作业使用,等待下一轮时再继续执行,直到该作业运行完毕。虽然在微观上用户程序的执行是不连续的,具有走走停停的特点,但是在客观上,用户的任何请求总是能够及时得到响应的。

分时系统具有以下特点:

(一)交互性

用户能与系统进行较为广泛的人机对话,即用户从键盘上输入命令,请求系统服务和控制程序的执行。系统能够及时地响应该命令,并在终端上显示响应结果。因此,用户可以在程序动态运行情况下对其加以控制,从而加快了调试过程,为软件开发提供了良好的环境。

(二)多路性

系统将若干个用户终端通过多路卡连接到一台主机上,宏观上,多个用户同时在自己的终端上工作,共享 CPU 和其它系统资源;微观上,各个终端作业是轮流执行一个时间片。多路性提高了资源利用率,节省了系统的开支,充分发挥了系统的效率,有些系统又把多路性称为同时性。

(三)独立性

由于采用时间片轮转方式使一台主机同时为多台终端服务,所以每一用户各占一台终端,彼此独立操作,互不干扰。从用户角度来看,每一用户并不感觉到有其它用户的存在,就如同自己独占计算机一样。

(四)及时性

系统保证对每一用户的输入请求做出及时的响应,使用户感觉到是他自己在使用和控制计算机。

分时系统也称为交互式系统,它与批处理系统有着明显的区别,而且他们追求的目标也不相同。在批处理系统中,主要考虑的是系统资源的利用率和吞吐量;而在分时系统中,必须注意用户的响应时间。因此,为了确保分时系统的整体功能和体现多用户分时操作的特征,分时系统还需考虑以下问题:

分时时间片的长短问题直接影响终端用户的及时性。由于系统中的用户终端数是有限的,只要时间片的长短选择适当,每一个终端用户从此次时间片使用完毕释放 CPU 到下次再拥有时间片(即获得 CPU)的等待时间就不会很长,不会影响用户程序的执行和终端操作。如果时间片过长,用户等待时间将会延长,使及时性得不到保证,严重影响用户操作、思维和心理。但如果时间片过短,CPU 就需要花费大量时间用于由于时间片的切换而引起的用户程序的切换、终端切换和内部管理上,增加系统的开销。

用作分时系统计算机的系统配置也将影响分时系统的性能,如主机运行速度太慢,内存容量太小,I/O 接口通道太窄等,都会直接影响到终端用户的及时响应,影响到交互操作。此时,用户发出请求命令后迟迟得不到系统的响应,会产生对系统操作的怀疑,甚至产生厌烦情绪。每开启一台终端,每增加一个用户,系统反应速度都会降低。所以,主计算机的配置要求一定要

能够满足分时操作系统的要求,只有硬件系统和软件系统的紧密配合,才能更好地发挥分时操作系统的功能。

UNIX 操作系统是一个典型的多用户、多任务的分时操作系统,我们将在第七章作详细介绍。

1.4.3 实时操作系统

实时系统是另一类特殊的多道程序系统,它主要应用于需要对外部事件进行及时响应并作出处理的领域。

所谓实时,是表示“立即”、“马上”、“及时”,而实时系统是指系统对特定输入做出反应的速度足以控制发出实时信号的对象。换句话说,计算机能及时响应外部事件的请求,在规定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,使之协调一致地运行。

实时系统可分为实时控制系统和实时信息系统,两者的主要区别是:一是服务对象;二是对响应时间的要求。

实时控制系统是指以计算机为中心的生产过程控制系统,也称为计算机控制系统。例如,钢铁冶炼和钢板轧制的自动控制、炼油、化工生产过程的自动控制等。在这类系统中,要求实时采集现场数据,并对它们进行及时处理,进而自动地控制相应的执行机构,使其参数(如密度、压力、流量等)能按预定规律变化或保持不变,以达到保证产品质量、提高产量的目的。同样,实时系统也可用于火箭的发射或武器的控制上,如火炮的射击控制、导弹的制导等等。

实时信息系统通常是指实时信息处理系统。它既可以是主机型多终端的联机系统,也可以是远程在线(online)式的信息服务系统,还可以是网络互联式的信息系统。作为信息处理的计算机接收终端用户或者远程终端用户发来的服务请求,系统将分门别类地进行数据与信息的检索、查找和处理,并及时反馈给用户。例如,航空机票预定系统、情报检索系统、信息查询系统等。

实时系统具有以下特征:

1. 及时性

实时系统的及时性是非常关键的,主要反映在对用户的响应时间要求上,对于实时信息系统,其对响应时间的要求类似于分时系统。分时系统对响应时间的要求是由人所能接受的等待时间来确定的,通常为秒级。而对于实时控制系统,则是以控制对象所能接受的延迟来确定的,它可以是秒级,也可以是毫秒级,甚至是微秒级。

2. 交互性

实时系统的交互性根据应用对象的不同和应用要求的不同,对交互操作的方便性和交互操作的权限性有着特殊的要求。由于实时系统绝大多数都属于专用系统,其交互性较差,对于实时控制系统来说,在某些情况下不允许用户干预,而对于实时信息系统,也仅允许用户在其授权范围内访问有限的专用服务性程序。

3. 多路性

与分时系统一样,它允许多个终端用户同时向实时信息系统提出服务请求,系统仍按分时原则为每个终端用户服务。对于实时控制信息,也经常具有多路采集现场信息和控制多个执行机构的功能。

4. 独立性