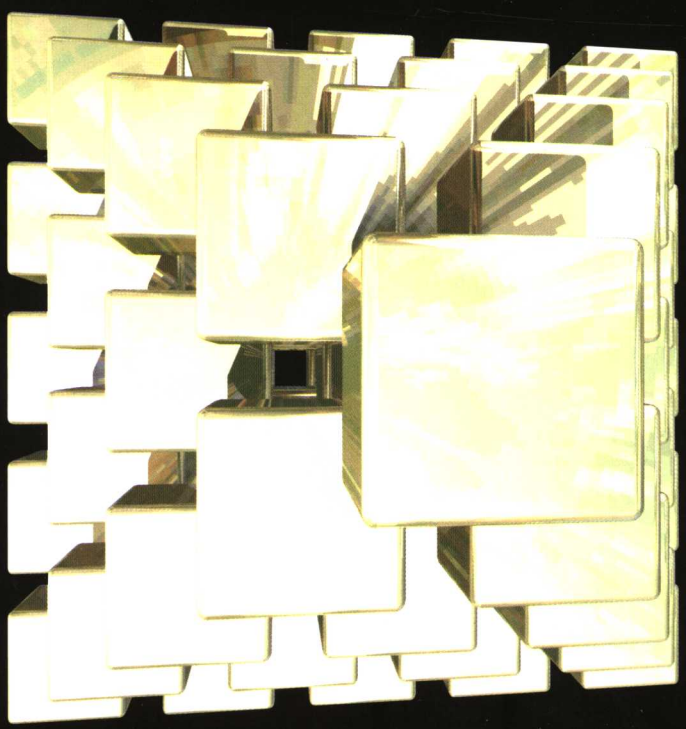
世界著名计算机教材精选

PEARSON Prentice Hall

操作系统原理

Lubomir F. Bic Alan C. Shaw 著
梁洪亮 等 译



OPERATING SYSTEMS PRINCIPLES

PEARSON Education

清华大学出版社

世界著名计算机教材精选

操作系统原理

Lubomir F. Bic
Alan C. Shaw 著
梁洪亮 等译

清华大学出版社
北京

EISBN:0-13-026611-6

北京市版权局著作权合同登记号　图字：01-2003-3537

图书在版编目(CIP)数据

操作系统原理/(美)伯尔(Bic, L. F.),(美)肖(Shaw, A. C.)著；梁洪亮等译. —北京：清华大学出版社，2005. 10
(世界著名计算机教材精选)
书名原文：Operating Systems Principles
ISBN 7-302-11602-4

Ⅰ. 操…　Ⅱ. ①伯…　②肖…　③梁…　Ⅲ. 操作系统－高等学校－教材　Ⅳ. TP316

中国版本图书馆 CIP 数据核字(2005)第 092539 号

出 版 者：清华大学出版社　　地　　址：北京清华大学学研大厦
　　http://www.tup.com.cn　　邮　　编：100084
　　社 总 机：010-62770175　　客户服务：010-62776969
组稿编辑：龙啟铭
文稿编辑：李　晔
印 刷 者：北京市世界知识印刷厂
装 订 者：三河市金元装订厂
发 行 者：新华书店总店北京发行所
开　　本：185×260　印张：28.75　字数：715 千字
版　　次：2005 年 10 月第 1 版　2005 年 10 月第1 次印刷
书　　号：ISBN 7-302-11602-4/TP·7580
印　　数：1～4000
定　　价：50.00 元

序

操作系统是连接计算机系统硬件和用户的桥梁。因此，它在很大程度上受到硬件技术和体系结构的影响。自从20世纪40年代首台计算机出现以来，硬件技术和体系结构始终在以惊人的速度发展。其中多数变化是可以量化的，例如，处理器、内存和外设的速度不断地提升，但是它们的体积、价格和能耗却一直在下降。除此之外，很多方面也发生了质的变化。例如，带有精密输入、输出和存储设备的个人计算机现已无处不在；并且大都已经与局域网或者因特网相联。这些进步给世界带来了激动人心的新局面，其中，操作系统的存在和协作成为必需。同时，要求现代操作系统能够管理高度并行的、分布式的和日渐增多的异构配置，而不只是管理单个处理器，控制本地内存和输入输出设备集合。

本书对操作系统原理进行了介绍，适合于计算机科学或计算机工程专业的低年级和高年级学生。从单处理器到分布式并行计算机系统的变化是一个主要的范例，本书的一个目标就是反映此转变，尤其是在目前无法严格区分面向集中环境的操作系统和面向分布环境的操作系统的情况下。尽管书中的大部分内容专注于传统主题，但是我们也扩充和融合进了分布计算的基本概念。

作者对如下人员表示真挚的感谢：Gary harkin（蒙大拿州立大学）、Mukkai Krisnimoorthy（仁斯利尔理工大学）、Scott Cannon（犹他州立大学）、John Hartman（亚利桑那大学）、Gopal Lakhani（得克萨斯理工大学）、Herb Mayer（波特兰州立大学）、Chung Kuang - Shene（密歇根理工大学），是他们对本书进行了认真仔细的审校。

内　　容

在“概述”一章之后，本书分成4部分：进程管理和同步、内存管理、文件和输入/输出管理以及保护和安全。在每章最后，列出了各章的关键概念、术语和缩略语。本书最后是全书的词汇表。

进程和线程

进程以及后来提出的线程是并发和并行的基础，且一直是操作系统研究中非常重要的部分。这个主题可以划分为两个组件：进程或线程的创建以及它们之间的协同。在第2章和第3章，我们从编程的角度分析了这个主题，描述了一系列表示并发和协调进程或线程并行执行的结构，其中包括在分布式环境中最终基于消息传递而非共享变量的进程协同。在第4章和第5章，我们从实现的角度展开讨论，描述了在操作系统层次实现和管理进程以及线程所必需的数据结构和操作，还讨论了进程和线程调度、中断处理和其他内核函数等。第6章重点论述了在集中式系统和分布式系统中都非常重要的死锁问题。

主存

主存一直以来都是一种稀缺资源，过去很多的操作系统研究集中于如何有效地使用主存。多数研究成果已经成为操作系统的经典论题，我们在第7、8、9章中对它们进行了论述。这些主题包括物理内存分配技术，使用分页或分段技术实现虚拟存储，以及数据和代码的静态共享和动态共享等。我们还介绍了分布式共享内存的原理，它可以被看作是对经由通信网络互联的多个计算机的虚拟存储的扩展。

文件系统和输入/输出

在早期计算年代，文件被设计为一种在辅助存储设备上组织和存储数据的方便方法。虽然设备发展很快，可是文件的基本原理没有变化。在第10章，我们讨论了文件类型及其在磁带和磁盘上的表示方法，表述了组织和实现文件目录的方法。近年来，网络的广泛普及促进了文件系统的重大发展。当前的许多系统都不在本地设备上维护自己的文件系统，相反地，更典型的配置是连成网络的许多机器，共同访问专门的文件服务器。更常见的是，文件系统散布于多个服务器或多个网络。第10章最后一节讨论了在这种分布环境下的文件系统问题。

操作系统的一个主要任务是通过支持高层次抽象来隐藏单个输入/输出设备的细节。现代系统必须继续提供这项基本的服务，并且要支持大量的更快更复杂的设备。在第11章中，我们描述了各种设备驱动所采用的轮询、中断和直接内存访问(DMA)的基本原理，还讨论了输入/输出处理中的与设备无关的问题，例如缓冲和高速缓存、错误处理和设备调度。

保护和安全

保护一个计算设备不受各种攻击需要许多类型的安全措施。第12章关注系统的保护和安全接口，它们将保护对系统的访问。这就要求对用户、远程服务和客户进行鉴别。尽管已经有了许多技术上的突破，用户鉴别仍然主要依赖于用户登录时提供的口令。而计算机网络的出现进一步刺激了保护和安全技术的极大发展：通信线路的脆弱性使得采用密钥或公钥的加密技术成为必需。我们对加密方法的应用进行了讨论，包括对计算机之间信息传输的保护以及验证其真实性。

用户一旦进入系统，系统就必须对用户可访问的资源进行控制。这需要结合指令层的硬件机制和软件层的访问列表或权能列表来完成。此外，系统还要提供防止不同用户之间的非法信息流的保护机制。第13章讨论了诸如此类的内部保护机制。

习题和编程项目

每章都以反映其主题的习题集结尾。这些练习经过了慎重的选择，以适应不同教学类型的需要。每个习题集既含有分析型题目又包括构建型题目，学生必须应用从该章中获得

的理论知识来解决具体问题。我们还提供了一些问题,留给学生自己讨论或探索。我们还为教师提供了一本解题指南,需要者可以从当地的 Prentice-Hall 代理处联系得到。

在本书正文最后,我们设计了 5 个大型编程项目和几个较小的编程练习,目的是通过实际的动手练习来补充对从本书所学概念的理解。可以选用它们作为学期项目,也可以作为单独的操作系统实验部分的基础。

方法和原则

我们在本书中深入描述了操作系统领域所有的标准主题。以前论述操作系统对分布式的基于网络的环境的支持时,一种常规的典型方法是在书后增加几章独立的内容。这种组织方法的问题是,在集中系统和分布系统之间人为地做了区分。实际上,两者之间并没有明显的界限,且两者具有许多共同之处。并发和并行一直是操作系统的主要内容。即使是 20 世纪 50、60 年代早期的大型机也试图通过重叠 CPU 执行和 I/O 处理以达到二者更好的利用率。20 世纪 80、90 年代的高级程序设计技术使得在用户层支持并发进程成为必需,这导致操作系统设计人员要提供新的进程同步和调度技术,这些技术同样可用于网络环境。最近 20 年,软件制造商开始认真考虑网络和物理的分布情况,并把必需的工具和技术集成到他们的操作系统产品中。

我们采用的方法是,通过在每一章中融合集中式操作系统和分布式操作系统的相关问题,保留两者间的本质联系和重叠部分。所讨论的分布式操作系统的主要内容包括:基于消息的同步、远程调用、分布式死锁、分布式共享内存、分布式文件系统和使用加密的安全通信。

按照上述指导原则,我们也没有将对现有操作系统的实例研究放在单独的章节里,相反,我们把对所有实例系统的研究,包括 UNIX、Linux、Windows 以及其他有影响的操作系统,分散和融合到各个章节中,这样在描述每个概念时也就阐明了其在各个实际系统中的特性。

Lubomir Bic

Alan Shaw

目　录

第2部分 内存管理

第 4 部分 保护和安全

第5部分 程序设计项目

第1章 概 述

在本章中,我们首先分析用户群对计算机系统的需求和期望与现有硬件的低层性能之间的差距。这个差距就是由操作系统(OS)和其他工具及支持程序所弥补的。然后我们勾画出操作系统的整体组织结构,包括对硬件的接口、对应用程序的接口和对用户的接口。最后我们描述了在技术日益更新和用户群日渐复杂多样的情况下,操作系统关键概念的进化。

1.1 操作系统的作用

1.1.1 弥补硬件和应用间的差距

当前大多数计算机系统都是以数学家约翰·冯·诺伊曼等在20世纪40年代末期提出的"存储程序计算机"的原理为基础的。图1-1以高级的方框图形式表示出了计算机的基本组件及其关系。计算机系统的核心是由中央处理器(CPU)和可执行的主存组成的计算引擎。这种存储器是可以直接寻址单元的线性序列,它存储程序(可执行的机器指令序列)和数据。CPU将连续地重复下述基本硬件循环:

- 获取一个称作"程序计数器"的专用寄存器所指向的指令。
- 程序计数器计数递增。
- 将存放在某个专用指令寄存器中的当前指令解码,以决定必须做什么。
- 获取这个指令引用的操作数。
- 执行指令。

这个循环是现今计算机上所有计算的基础。

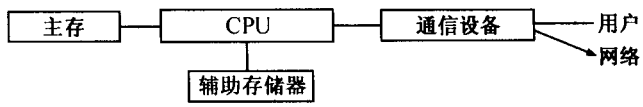


图1-1 计算机系统的主要组件

为使上述计算模式具有实用意义,还必须有两个基本组件。第一个组件是一组通信设备,使用它们可以在用户和计算机之间以及计算机和计算机之间进行数据和命令交换。这主要包括输入/输出(I/O)设备(例如键盘和显示终端)和网络接口设备。第二个组件是辅

助存储器(以下简称辅存),它用来保存当前没有载入主存的,或者部分或临时载入的程序和数据。这类存储器是必需的,因为,系统内存是不稳定的,因而断电时内容会丢失,并且它的容量远远小于辅存。

通信设备和存储设备间的差别有时很明显。例如,一个 CD-ROM 设备显然是一个输入设备,而一个硬盘明显是一个存储设备。但是,在很多情况下,这种区别并不明显。例如,一个可移动的软磁盘可以被认为是存储设备,但也可用作输入/输出设备在不同系统之间传递信息。从操作系统的角度来看,CD-ROM、硬盘、软磁盘和其他设备在本质上是相似的,并使用很多相同的技术为它们服务。我们后面将把辅助存储设备和通信设备统称为输入/输出设备。

当计算机系统包含多个 CPU 时,情况会较为复杂。它可以表现为几种形式,具体依赖于系统硬件组件的共享层次。图 1-2 表示了 3 种可能的体系结构,它们用不同的方法扩展了图 1-1 中基本的单 CPU 体系结构。在第一种情况下(图 1-2(a)),两个 CPU 共享一个公共的内存。辅助存储设备和通信设备是典型的共享设备。多 CPU 的出现给操作系统带来了新的挑战。其中之一是高速缓存技术(caching)。如果每个 CPU 都维护自己的本地内存缓存,那么系统必须保证,对于同一个存储单元,两个高速缓存不能含有不同的值。在共享内存时,这个问题称作高速缓存一致性,它是由硬件处理的,对操作系统是透明的。另一个重要问题是进程调度。对单 CPU 而言,调度就是控制进程执行的顺序。对多 CPU 而言,操作系统(或应用程序)还必须决定一个任务在哪个 CPU 上运行。通过共享内存来实现运行在不同 CPU 上的多个进程间的同步和通信,这个方法类似于在单 CPU 上协调多个进程。

在图 1-2(b)所示的结构中,每个 CPU 都有自己的主存。辅助存储器和其他设备仍然是被共享的。但是,由于没有共享主存可用,通信子系统必须含有一个互联网络,以使 CPU 之间可以相互联系。有很多种互联网络,简单的如共享总线,复杂的如以各种拓扑形式组成的专门连接。在没有共享内存的情况下,调度和进程协调变得更为复杂。调度不仅涉及到将进程分配到不同的 CPU,还包括数据到独立的本地内存模块的分配,这些数据可能是多个进程需要的。考虑到此类结构主要用于高性能科学计算,许多调度问题交由应用程序处理,而非由操作系统进行透明处理。但是,系统必须提供支持进程同步和通信所必需的原语。这些原语是基于通过互联网络在不同 CPU 间发送的消息的。根据网络的复杂程度,系统可能还必须解决高速缓存一致性问题。

图 1-2(a)和图 1-2(b)所示的结构一般被称为多处理器结构。它们提供多个 CPU,也可能提供多个内存模块,但是其他硬件是共享的。因此这类系统仍可以被看作单个的计算机。图 1-2(c)给出了一种不同的结构,一般称之为多计算机结构,其中,每个 CPU 不仅有自己的内存,还有自己的存储设备和通信设备。在通信设备中有专门的网络控制器,允许每台计算机通过网络与其他计算机发收数据。因此,一个多计算机结构可看作为多个完全自主的计算机的集合,它们通过网络相互联系。

有许多网络类型,它们在大小、拓扑、数据传输速率、可靠性和其他属性方面有所不同。两台计算机通信的一种常用方法是使用现有的电话线路。电话线路原本设计用来运载模拟的声音数据。为了在电话线上传输数字数据,每个通信计算机要使用一个调制解调器(modem),它是一个在发送端把数字数据转换为模拟数据,而在接收端把模拟数据转回数字数据的设备。同其他设备控制器类似,调制解调器也会连接到一条输入/输出总线上。

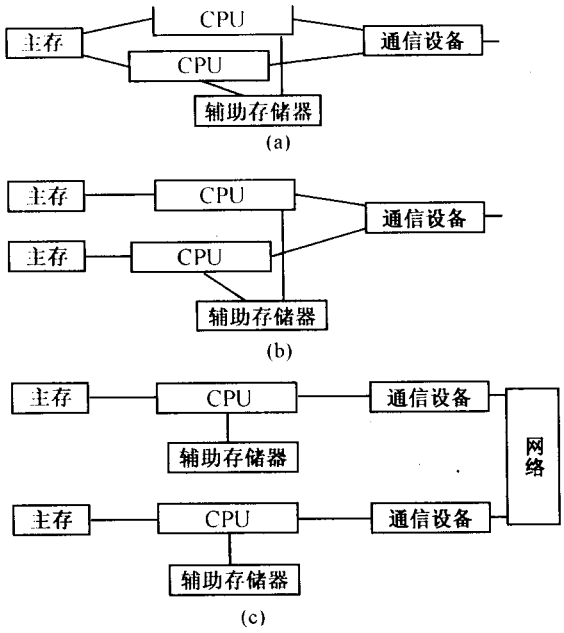


图1-2 使用多CPU的系统:(a)共享内存多处理器;(b)分布式内存多处理器;(c)多计算机

使用专门设计的连接两台以上计算机的通信线路,可以获得更快更可靠的连接。对于放在同一个房间或建筑物中的少量计算机而言,通常采用局域网(LAN)形式。图1-3说明了两种最常用的局域网的拓扑结构:环状网络和广播总线。在环状结构中,每台计算机与两台计算机相邻。数据沿着环路从发送者经过中间节点到达接收者。广播总线与内部系统总线相似。最常见的类型是以太网总线。每台计算机使用一个专用的网络控制器(以太网卡)连接到总线上。所有网络控制器"监听总线",即监视是否有数据放在总线上面。需要发送消息给其他计算机的机器把消息放在总线上,消息前面是接收者的地址。所有监听总线的节点将检查这个地址,与节点地址匹配的节点读取消息内容。网络控制器还要实现解决冲突的协议(冲突是指多台计算机同时在总线上放置数据)。

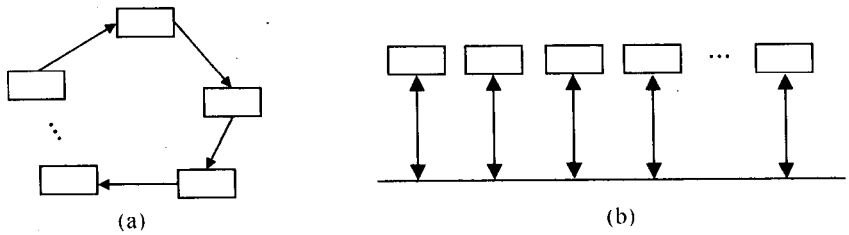


图1-3 局域网的网络拓扑:(a)环状网;(b)广播总线

局域网受到它们能够支持的计算机数目和计算机之间的物理距离的限制。要把大量相距很远的计算机互连起来,可以使用广域网(WAN)。这要使用多种硬件设备,当信号长距离传输时,这些设备可以重复和放大信号。对广域网而言,在所有通信计算机之间建立点到点的通信信道是不实际的。代替的作法是,消息被分割成多个数据包,每个数据包经过广域网的网关和节点,在目的地被重新组合在一起。广域网需要有复杂的通信协议以管理传输过程的所有方面,这样大量的异种计算机和局域子网就能够协作和竞争。超大型的异构网