并行与分布计算技术丛书

并行操作系统原理与技术

The Principle and Technique of Parallel Operating System

夏卫民　罗　宇　吴庆波
俞　晓　肖　侬　周良源　著

国防工业出版社

·北京·

图书在版编目(CIP)数据

并行操作系统原理与技术/夏卫民等著 .—北京:国防工业出版社,2002.2
(并行与分布计算技术丛书)
ISBN 7-118-02670-0

Ⅰ.并… Ⅱ.夏… Ⅲ.并行控制系统:操作系统(软件) Ⅳ.TP316

中国版本图书馆 CIP 数据核字(2001)第 073311 号

国防工业出版社出版发行
(北京市海淀区紫竹院南路 23 号)
(邮政编码 100044)
三河市腾飞胶印厂印刷
新华书店经售

开本 787×1092 1/16 印张 14¾ 320 千字
2002 年 2 月第 1 版 2002 年 2 月北京第 1 次印刷
印数:1-4000 册 定价:36.00 元

致 读 者

本书由国防科技图书出版基金资助出版。

国防科技图书出版工作是国防科技事业的一个重要方面。优秀的国防科技图书既是国防科技成果的一部分，又是国防科技水平的重要标志。为了促进国防科技和武器装备建设事业的发展，加强社会主义物质文明和精神文明建设，培养优秀科技人才，确保国防科技优秀图书的出版，原国防科工委于 1988 年初决定每年拨出专款，设立国防科技图书出版基金，成立评审委员会，扶持、审定出版国防科技优秀图书。

国防科技图书出版基金资助的对象是：

1. 在国防科学技术领域中，学术水平高，内容有创见，在学科上居领先地位的基础科学理论图书；在工程技术理论方面有突破的应用科学专著。

2. 学术思想新颖，内容具体、实用，对国防科技和武器装备发展具有较大推动作用的专著；密切结合国防现代化和武器装备现代化需要的高新技术内容的专著。

3. 有重要发展前景和有重大开拓使用价值，密切结合国防现代化和武器装备现代化需要的新工艺、新材料内容的专著。

4. 填补目前我国科技领域空白并具有军事应用前景的薄弱学科和边缘学科的科技图书。

国防科技图书出版基金评审委员会在总装备部的领导下开展工作，负责掌握出版基金的使用方向，评审受理的图书选题，决定资助的图书选题和资助金额，以及决定中断或取消资助等。经评审给予资助的图书，由总装备部国防工业出版社列选出版。

国防科技事业已经取得了举世瞩目的成就。国防科技图书承担着记载和弘扬这些成就，积累和传播科技知识的使命。在改革开放的新形势下，原国防科工委率先设立出版基金，扶持出版科技图书，这是一项具有深远意义的创举。此举势必促使国防科技图书的出版随着国防科技事业的发展更加兴旺。

设立出版基金是一件新生事物，是对出版工作的一项改革。因而，评审工作需要不断地摸索、认真地总结和及时地改进，这样，才能使有限的基金发挥出巨大的效能。评审工作更需要国防科技和武器装备建设战线广大科技工作者、专家、教授，以及社会各界朋友的热情支持。

让我们携起手来，为祖国昌盛、科技腾飞、出版繁荣而共同奋斗！

国防科技图书出版基金
评审委员会

国防科技图书出版基金
第三届评审委员会组成人员

名誉主任委员	怀国模
主任委员	黄　宁
副主任委员	殷鹤龄　高景德　陈芳允　曾　铎
秘书长	崔士义
委员 （以姓氏笔划为序）	于景元　王小谟　尤子平　冯允成 刘　仁　朱森元　朵英贤　宋家树 杨星豪　吴有生　何庆芝　何国伟 何新贵　张立同　张汝果　张均武 张涵信　陈火旺　范学虹　柯有安 侯正明　莫梧生　崔尔杰

并行与分布计算技术丛书编委会

总　　序

并行与分布计算技术是实现高性能计算的重要技术途径。高性能计算机技术是现代科学研究、工程技术开发和大规模数据处理的关键支撑技术。没有高性能计算机，大量复杂问题的计算和事务处理就无法在合理的时间内完成。利用高性能计算机还可以解决一些仅靠理论及实验方法无法解决的问题，分析处理靠传统技术无法应付的海量数据，例如核爆炸模拟、宇宙的形成及演变过程研究、中长期天气预报、石油地质勘探、数字地球和大规模事务处理等。人类对高性能计算能力的需求永无止境。计算速度、存储容量、通信带宽是衡量高性能计算能力的重要技术指标。一些推动人类文明和社会信息化的重大挑战性问题需要百万亿次每秒、千万亿次每秒以上的计算速度，需要万亿字节以上的存储容量和万亿位每秒以上的通信能力。人们已经认识到，高性能计算与网络通信技术是战略技术，是科技创新体系的基础技术，是反映一个国家综合实力的重要标志之一。

由于种种因素的制约及使用计算机方式的改变，在单台计算机系统的性能与功能难以满足应用需求的情况下，由多个计算节点经专门设计的互联网络紧密耦合而构成的大规模并行处理计算机系统，以及由多个自主的高性能计算节点经计算机网络连接组成的分布处理环境，将成为高性能计算机领域的两大重要研究方向。并行计算和分布计算是既有区别又联系密切的两个概念，前者重在发掘计算过程中的并行性，后者则重在有效组织管理各类异构资源，挖掘功能上的并发性。随着基于先进计算机网络的分布并行计算概念的发展，并行计算与分布计算在很多应用领域正面临共同的追求目标和技术挑战，如高效、实用的计算模型、计算方法、并行机制等。

20世纪90年代以来，随着高性能计算和网络计算技术的普及，并行与分布计算技术正渗透到现代社会的各个领域，用户通过高性能网络使用各类计算资源提供的服务已成为信息化社会中计算机应用的一种重要形式。事实上，各类基于并行与分布计算的应用系统正在工业、交通、金融、科研、政府、国防等部门支撑着现代社会的高效运行。

20多年来，我国科技人员依靠自己的力量，勇于开拓，奋勇拼搏，研制成功了多种型号序列的高性能计算机系统。国防科学技术大学计算机学院是我国研制高性能计算机系统的重要基地，多种巨型计算机的研制成功及推广应用，打破了国外对我国的技术封锁，缩短了我国同发达国家技术水平上的差距，为推动我国高性能计算机及并行与分布计算技术的发展作出了重要贡献。

为促进并行与分布计算技术领域的研究，国防工业出版社组织国防科技大学计算机学院有关专家、教授撰著了本套丛书。本丛书以并行与分布计算机系统组成为纲，结合作者多年科研与工程实践以及当前研究的热点问题，涵盖了计算机系统结构、计算机网络、系统软件、应用软件、计算方法等多方面内容。本丛书由以下9本专著组成：《并行计算机体系结构技术》论述并行计算机研究和设计的理论及工程问题；《先进计算机网络技术》论

述高性能网络计算的各种关键技术;《并行操作系统原理与技术》论述并行操作系统的机制、基本原理和主要实现技术;《并行编译方法》论述并行编译系统理论、编译器的设计方法;《分布计算——网络化软件新技术》论述分布计算技术的基本概念和关键技术;《分布式数据库技术》论述统一逻辑分布式数据库技术;《数字系统并行 CAD 技术》论述数字系统并行 CAD 的理论和技术;《可扩展并行算法的设计与分析》论述可扩展并行算法的设计与分析的理论和方法;《并行与分布式可视化技术及应用》论述并行与分布式可视化的各种技术及其应用。本丛书既介绍了当前国际上该领域的最新技术发展,又汇集了作者多年的研究成果和工程经验。丛书注重可读性,适合从事该领域工作和学习的科技人员、高等院校高年级学生及研究生作为工作和学习的参考书。

本丛书被列入"九五国家重点图书选题规划",并获得国防科技出版基金的资助。愿本丛书的出版能为并行与分布计算技术研究园地增添一朵花絮,为以后的研究工作提供有价值的参考。因时间和能力所限,书中不足之处,恳请读者指正。

并行与分布计算技术丛书编委会

前　言

并行处理是当今计算机发展的一个热点，自20世纪90年代以来，信息时代已进入了并行处理的年代，尽管并行处理技术还有待进一步发展和提高，但它已被广泛地应用于当今的科学计算、信息处理和移动通信，特别是高性能计算等领域。继“银河－Ⅰ”亿次向量巨型机于1983年诞生之后，我国已先后成功研制出“银河－Ⅱ”、“银河－Ⅲ”、“神州－Ⅰ”、“神州－Ⅱ”、“曙光1000”和“曙光2000”等多个并行计算机系列。并行操作系统是并行计算机的重要组成部分，它也随着并行处理技术和并行计算机体系结构的发展而发展，并且越来越受到人们的重视。现在我国已有大量的科研人员在从事并行处理方面的研究工作，大专院校的学生们也在进行或将要进行并行处理方面的学习和探索工作，但是目前在并行处理方面的中文书籍和参考资料却很缺乏，特别是并行操作系统方面的书籍还没见到。为此，我们在参阅了大量国内外有关并行处理资料的基础上，从分析传统操作系统与并行处理不相适应的地方入手，并结合我们自己的实践，撰写了本书，希望能起到抛砖引玉之功效。

本书共8章，论述了并行操作系统的并行机制、基本原理和主要技术。

第1章概论。简要介绍了并行操作系统的发展历程，展示了并行操作系统随现代科技进步和体系结构的发展而进化的艰苦历程，给出了现代并行操作系统的主要结构、特点和适应面。

第2章进程与线程。介绍了进程与线程的控制模型、运行机理以及它们的状态演变过程，特别介绍了在并行操作系统中广泛使用的线程的工作原理与实现方法。

第3章处理机调度。在分析一般进程调度的基础上，介绍了多机调度的基本概念以及线程调度和实时调度。

第4章存储管理。首先介绍了现代计算机存储体系结构，分析了存储体系中的层次关系及分布模式，针对不同的存储器模型介绍了用户空间的组织方法、数据访问模式，特别介绍了分布存储器系统上数据的迁移技术和数据一致性管理技术。

第5章同步与互斥。首先分析了同步机制在并行机系统中的重要地位，分析了单机系统中同步机制的不足，介绍了多机系统中常见的同步手段，特别对不同种类的锁机制进行了解剖和分析。

第6章机间通信技术。介绍了共享存储器和分布存储器环境下消息的发送与接收，特别介绍了现代流行的端口技术、主动消息及消息传递接口界面。

第7章并行I/O和并行文件系统。重点讨论了并行I/O、分布式文件系统、并行文件系统的技术和原理，并给出了相关的实现实例。

第8章Mach操作系统核心分析。在分析源代码的基础上，对Mach微内核的体系结构、进程管理、存储及虚存管理以及进程间通信进行了一定的剖析。

本书第1章由周良源编写，第2章、第3章由罗宇编写，第4章由俞晓和夏卫民编写，第5章由夏卫民和周良源编写，第6章由肖侬编写，第7章由吴庆波编写，第8章由刘忠编写。卢宇彤、陈松政和郑立刚等同志参加了本书的撰写工作。全书由夏卫民统稿。杨学军教授在百忙中仔细审阅了全书并提出了许多指导意见。在此，对关心和帮助过本书编著工作的所有同志表示由衷的感谢。

限于编者的水平，书中疏漏及错误之处恳请专家、读者批评指正。

目　录

Content