世界著名计算机教材精选

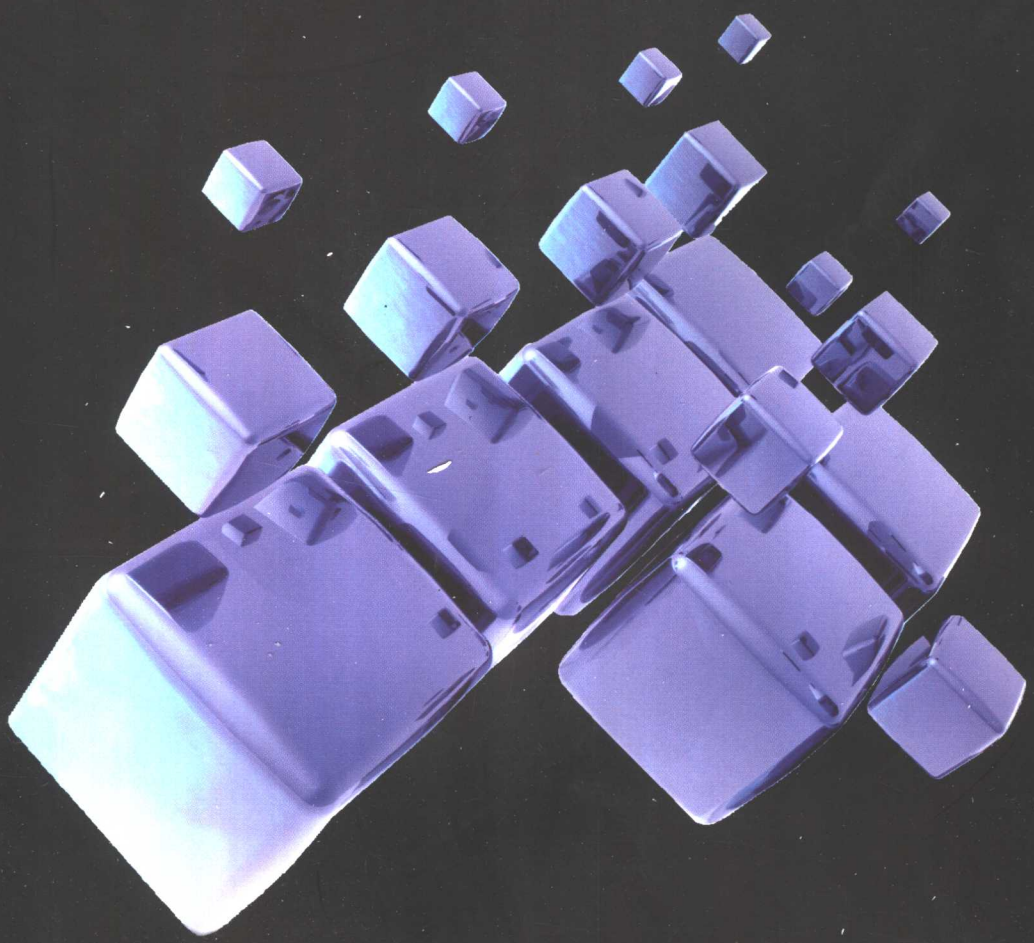
PEARSON Prentice Hall

分布式系统

原理与范型

Andrew S. Tanenbaum
Maarten van Steen 著

杨剑峰 常晓波 李敏 译



DISTRIBUTED SYSTEMS

Principles and Paradigms

清华大学出版社

世界著名计算机教材精选

分布式系统原理与范型

Andrew S. Tanenbaum，Maarten van Steen 著

杨剑峰　常晓波　李　敏　译

清华大学出版社

北　京

图书在版编目(CIP)数据

分布式系统原理与范型/特南鲍姆(Tanenbaum, A. S.),范施特恩(van Steen, M.)著;杨剑峰等译. —北京：清华大学出版社,2004.9

(世界著名计算机教材精选)

书名原文：Distributed Systems: Principles and Paradigms

ISBN 7-302-08961-2

Ⅰ. 分…　Ⅱ. ①特… ②范… ③杨…　Ⅲ. 分布式操作系统—教材　Ⅳ. TP316.4

中国版本图书馆 CIP 数据核字(2004)第 063337 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

责任编辑：袁勤勇

印 刷 者：北京四季青印刷厂

装 订 者：北京市密云县京文制本装订厂

发 行 者：新华书店总店北京发行所

开　　本：185×260　**印张**：39.75　**字数**：915 千字

版　　次：2004 年 9 月第 1 版　2004 年 9 月第 1 次印刷

书　　号：ISBN 7-302-08961-2/TP・6340

印　　数：1～4000

定　　价：68.00 元

译 者 序

随着计算机网络，特别是 Internet 的迅猛发展，传统的信息系统概念发生了巨大的变化，这些变化突出地表现在信息的存储、传递、发布以及获取方式所发生的革命性变革。与此同时，基于网络的分布式信息系统在各个领域得到了广泛的应用，在整个社会生活中正发挥着日益突出的作用。Internet 已经越来越多地成为构建信息系统的一个关键组成部分。如何在更为广域和异构的计算环境中有效地发布和获取信息，已成为亟待解决的问题。分布式系统正是解决了上述问题。现在分布式系统的研究、应用日益广泛深入，分布式系统的学习也成为计算机及相关专业必不可少的教学环节。

本书是 Tanenbaum 先生在所著的《分布式操作系统》的基础上，总结了分布式系统方面的最新进展，重新撰写的力作，是分布式系统的权威教材。本书循序渐进地、全面地、深入地讲解了分布式系统的原理，并列出了大量的范型。本书的结构分为两部分：原理和范型。第一部分（第 1～8 章）详细讨论了分布式系统的原理、概念和技术，其中包括通信、进程、命名、同步、一致性和复制、容错以及安全。第二部分（第 9～12 章）给出了一些实际的分布式系统，即基于对象的分布式系统、分布式文件系统、基于文档的分布式系统以及基于协作的分布式系统，介绍了一些实际系统的设计思想和实现技术。全书结构清晰，内容全面经典，系统性与先进性并茂。

本书的目标读者是计算机及相关专业的高年级学生或研究生。从事分布式计算研究和工程应用的科研人员和工程技术人员也会从本书中受益匪浅。

本书是多人共同努力的成果，参与本书翻译、审稿、录排的人员包括：杨剑峰、常晓波、梁金昆、张丽萍、汪青青、朱志博、李敏、李静、李娟、张颖、朱剑平、刘颖、吴东升、杨战伟、郭宁宁、李楠、聂晶、刘恒、刘敏、刘洋、吕喜熹、马睿倩等。全书由杨剑峰、常晓波和李敏负责统稿。

限于译者水平，难免有错误和疏漏之处，恳请读者不吝指正。希望这本书能成为您工作的好帮手。

杨剑峰　常晓波

2004 年 5 月

前　言

本书的出发点是对 Distributed Operating Systems 一书进行再版修订，但笔者很快就发现自 1995 年以来很多技术发生了改变，要完全体现出这些变化，仅仅对该书进行修订是不够的，而是需要写一本全新的书。因此，这本新书有了一个新的标题：《分布式系统原理和范型》。标题的改变体现了对重点的调整。虽然我们仍然讨论一些操作系统的问题，但现在这本书还从更广泛的意义上研究分布式系统。例如，WWW 作为已建立的最大的分布式系统，在 Distributed Operating Systems 一书中完全没有提到，因为它并不是一个操作系统。而在本书中，它几乎占去整整一章。

本书分为两部分：原理和范型。第 1 章是对主题的总体介绍。接下来的第 2～8 章分别讨论我们认为最重要的原理：通信、进程、命名、同步、一致性和复制、容错以及安全性。

实际的分布式系统通常是围绕一些范型来组织的，例如“所有事物都是文件”。接下来的第 9～12 章分别介绍一个不同的范型，并描述使用该范型的一些重要系统。涉及到的范型包括基于对象的系统、分布式文件系统、基于文档的系统以及基于协作的系统。

第 13 章包含一份附有说明的参考书目，可供该主题的进一步学习使用，还包含本书中引用的著作列表。

本书是作为计算机科学的大学高年级学生或研究生课程而编写的。因此，本书有一个 Web 站点，站点中以各种格式放置了本书中用到的 PowerPoint 表和图。要访问该站点，在 http://www.prenhall.com/tanenbaum 页面上点击本书标题即可。将本书作为教材使用的教授可以通过联系当地的 Prentice Hall 代理机构得到一本习题解答手册。当然，本书也十分适合希望更多地了解这一重要主题的社会人士。

许多人以多种方式对本书作出了贡献。我们尤其要感谢 Arno Bakker、Gerco Ballintijn、Brent Callaghan、Scott Cannon、Sandra Cornelissen、Mike Dahlin、Mark Darbyshire、Guy Eddon、Amr el Abbadi、Vincent Freeh、Chandana Gamage、Ben Gras、Bob Gray、Michael van Hartskamp、Philip Homburg、Andrew Kitchen、Ladislav Kohout、Bob Kutter、Jussipekka Leiwo、Leah McTaggert、Eli Messenger、Donald Miller、Shivakant Mishra、Jim Mooney、Matt Mutka、Rob Pike、Krithi Ramamritham、Shmuel Rotenstreich、Sol Shatz、Gurdip Singh、Aditya Shivram、Vladimir Sukonnik、Boleslaw Szymanski、Laurent Therond 和 Leendert van Doom，感谢他们阅读了部分书稿并提出了宝贵意见。

最后，我们还要感谢我们的家庭。Suzanne 已经经历过很多次这样的情况了。她从未说过“我受够了”，尽管这个念头肯定在她脑海里出现过。谢谢你！Barbara 和 Marvin 现在对教授们为谋生所做的工作有了更好的了解，并且认识到好教材和坏教材之间的差别。现在，对我来说他们是我努力创作出更多好教材的动力所在。

本书使用指南

我们使用本书中的材料已经很多年了，主要是用作大学高年级学生和研究生的教材。而且，这些材料还曾经作为为时 1～2 天的有关分布式系统和中间件的研讨会的基本资料，参加这些研讨会的人包括 ICT 专家(技术上的)。下面是我们根据经验对本书使用方式提出的一些建议。

大学高年级学生和研究生教材

如果作为大学高年级学生和研究生的教材，本书通常可以在 12～15 周内完成教学。我们发现，在大多数学生看来，分布式系统由很多似乎彼此紧密结合的主题所组成。在本书的组织上，我们按照不同的原理介绍这些主题，分别讲授各个原理，这对学生领会重点内容有很大帮助。这样安排的效果是当第一部分(第 1～8 章)结束时，即在讨论范型之前，学生已经对本书主题在整体上有了一个相当好的把握。

然而，分布式系统的领域涵盖许多不同的主题，其中一些主题在初次学习时很难理解。因此，我们强烈建议学生们随着课程的进展学习适当的章节。从 Web 站点(http://www.prenhall.com/tanenbaum)可以获得所有 PowerPoint 表，将它们预先分发下去，以便学生在课堂中能够积极参与讨论。这种方法非常成功，并得到了学生们的高度评价。

所有的材料都包括在一个为时 15 周的课程中。大多数时间花费在讲授分布式系统的原理，也就是前 8 章所包括的材料上。在讨论范型时，我们的经验是：只需要介绍要点。直接从书中学习每个案例的详细内容比在课堂上听授更加容易。例如，尽管书中有关基于对象的系统的内容达 80 页之多，但我们只用一周的时间讲授这类系统。下面是一个课程进度安排建议表(表 0.1)，其中包括每次讲座中包括的主题。

表 0.1　课程进度安排建议

周	主题	章	讲授内容
1	绪论	1	全部
2	通信	2	2.1～2.3
3	通信	2	2.4～2.5
4	进程	3	全部
5	命名	4	4.1～4.2
6	命名	4	4.3
6	同步	5	5.1～5.2

续表

周	主题	章	讲授内容
7	同步	5	5.3～5.6
8	一致性和复制	6	6.1～6.4
9	一致性和复制	6	6.5～6.6
9	容错	7	7.1～7.3
10	容错	7	7.4～7.6
11	安全性	8	8.1～8.2
12	安全性	8	8.3～8.7
13	基于对象的系统	9	全部
14	文件系统	10	全部
15	基于文档的系统	11	全部
15	基于协作的系统	12	全部

并不是所有材料都需要在课堂上讲授；我们希望学生能够自学特定的部分，尤其是细节部分。在讲授时间少于15周的情况下，我们建议跳过有关范型的章节，让感兴趣的学生自己学习这些部分。

如果用于低年级的课程，我们推荐将本书的学习延长至两个学期，并增加实验作业。例如，可以通过让学生修改一些组件，使这些组件具有容错性、处理多播RPC等功能来使学生理解简单的分布式系统。

行业的专业研讨会

在1～2天的研讨会上，通常将本书作为主要的背景材料使用。然而，如果跳过所有细节，仅将重点放在分布式系统的本质上，则有可能在两天内讲完整本书。此外，要使内容的表达更加生动实用，有必要重新安排章节的顺序，以提早说明原理是如何得到应用的。对于研究生来说，一般是在了解原理的应用之前(有时甚至根本不了解原理的具体应用)先对原理进行为期10周的学习，但专业人士如果能了解这些原理的实际应用，就会有更大的学习动力。下面是一个为期2天课程的试验性进度表(表0.2)，该表按照逻辑单元进行划分。

表0.2　按逻辑单元划分的课程进度

第1天				
单元	时间(分)	主　题	章	重　点
1	90	绪论	1	客户-服务器体系结构
2	60	通信	2	RPC/RMI和消息传递
3	60	基于协作的系统	12	消息传递问题
4	60	进程	3	移动代码和代理
5	30	命名	4	位置跟踪
6	90	基于对象的系统	9	CORBA

续表

第 2 天				
单元	时间(分)	主　题	章	重　点
1	90	一致性和复制	6	模型和协议
2	60	基于文档的系统	11	Web 缓存/复制
3	60	容错	7	进程组与 2PC
4	90	安全性	8	基本思想
5	60	分布式文件系统	10	NFS v3 和 v4

个人学习

本书同样也适用于个人学习。如果具有足够的时间和动力，建议读者仔细阅读整本书。

如果没有足够的时间仔细阅读所有材料，我们建议只集中学习最重要的主题。下面的表格中列举一些章节，我们认为这些章节涵盖了关于分布式系统的最重要的主题(表 0.3)。

表 0.3　自学内容

章	主　题	小　节
1	绪论	1.1、1.2、1.4.3、1.5
2	通信	2.2、2.3、2.4
3	进程	3.3、3.4、3.5
4	命名	4.1、4.2
5	同步	5.2、5.3、5.6
6	一致性和复制	6.1、6.2.2、6.2.5、6.4、6.5
7	容错	7.1、7.2.1、7.2.2、7.3、7.4.1、7.4.3、7.5.1
8	安全性	8.1、8.2.1、8.2.2、8.3、8.4
9	基于对象的系统	9.1、9.2、9.4
10	分布式文件系统	10.1、10.4
11	基于文档的系统	11.1
12	基于协作的系统	12.1、12.2 或 12.3

比较好的做法是对学习这些建议的材料需要花费的时间进行估算，但这在很大程度上取决于读者的背景知识，对各种背景的读者很难做一个一般性的估计。然而，如果一个具有全职工作的人抽出晚上的时间阅读本书，则可能至少花费几周时间。

目　　录