

# 分布式系统 — 分布式算法

苏运霖◎著

暨南大学出版社

# 分布式系统与分布式算法

FENBUSHIXITONG YU FENBUSHISUANFA

苏运霖 著

暨南大学出版社

**粤新登字 13 号**

责任编辑：张雨竹

封面设计：水 珠

---

**分布式系统与分布式算法**

著 者：苏运霖

出版发行：暨南大学出版社

经 销：广东省新华书店

排 版：广州市科新电脑技术服务中心

印 刷：广东省水电学校印刷厂

开 本：850×1168 1/32

印 张：18.25 43 万字

版 次：1995 年 5 月第 1 版

1995 年 5 月第 1 次印刷

印 数：1—1000 册

**ISBN 7-81029-381-8**

**T·5 定价：48.00 元**

---

## · 暨南学人丛书 ·

### 总前言

“朔南暨，声教讫于四海。”——语出《尚书》。

以传扬华夏“声教”为己任的暨南大学，是我中华第一间国立华侨大学。自清光绪三十二年建校始，经历沧桑已近一个世纪，泽育之暨南学人学子数以万计，有如千川百汇，流布海内外，传承华夏之文化灵慧，续我炎黄之魂魄命脉。暨南学人骄子，不负祖先厚望，不负母校育爱，在彼司职之岗位，勤奋耕耘，多有卓著奉献。有跻身于世界名人名流之行列者；有潜心于学术科技，摘得丹峰桂冠者；亦有投身于各行各业宏达，绘展祖先希冀之蓝图者。为人类之文明进步、华夏之昌盛振兴，献出了殷殷赤子之情。为母校暨南大学赢得了美好的社会声誉。母校为此而感到光荣、骄傲。

值此母校出版社创立三周年之际，我们推出“暨

南学人丛书”系列图书,旨在汇结暨南学人之有价值、有影响的学术研究成果,弘扬暨南校友的卓著声名、成功业绩,以增添祖国和母校的文明库藏和学术文化积累,沟通和系结海内外学人学子之心灵和情怀,铺设“龙的传人”的汇接桥梁。

《暨南学人丛书》乃为暨南大学历届学人自己的一套永续不断的书列。举凡功成名就、贡献突出者列入本丛书传记系列出版;在学术上有高水准建树,并有重大价值和影响的论著,则列入本丛书学术专著出版。

《暨南学人丛书》之出版计划,深得我校师生、校友及校董事会董事们的热情关注和真诚支持。或献智献策,或惠寄书稿,或解囊资助,令我们感到无限欣慰与鼓舞。谨愿五洲四海之暨南学子,以心系故园推动科技文化进步之情怀,支持《暨南学人丛书》的出版工作,俾使我们不负历史与时代之重托,不负历届学人之厚望。

暨南大学出版社

一九九三年三月

# 前言

在经历了两个酷暑与严寒之后，这本书终于完成了。现在，我以真挚的心情向计算机科学界的同行和读者献上本书，且愿它具有学习的价值和参考的价值，它的出版能推进我国计算机科学界在本领域的研究与应用。

这里，对于写作本书的起因和背景作些介绍，或许会对读者有所帮助。

那是在 1989 年，作者获得著名美籍物理学家、诺贝尔奖金获得者杨振宁教授的邀请，前往美国纽约州立大学石溪分校进行学术访问。该校计算机科学系的主任 Philip M. Lewis 教授在我抵达之后，即介绍我和该系教授 Arthur J. Bernstein 合作。Arthur J. Bernstein 教授立即向我介绍了他本人在分布式算法、并发性等方面的研究工作，并且向我提供了大量这方面的最新文献。事实上，Philip M. Lewis 教授也从事这个领域的研究工作，他们俩正合著《程序设计与数据库系统的并发性》一书。当时他就指出，这是一个很有前途、当前也很热门的研究领域。于是，我才开始系统接触这方面的论著，包括校审他们的合著新书稿。我发现，这个领域确实包含了许多非常有趣而困扰的问题。我在那里的时间愈长，就愈深刻感受到那

里对这个问题的研究热度，也愈发感到它的重要性。这不仅仅是学术界的兴趣爱好所致，也是实践向理论界提出的一个挑战。

回国后，我继续进行这个领域的研究，同时也准备本书的写作。当然，我的研究工作本身也是准备工作的一部分。我原计划用一年时间完成这件事，但是，繁重的教学和科研工作，加上同样繁重的行政工作，使我不得不一再推迟这个计划。有时，我简直忙得一天连写一行稿的时间都没有。然而，一旦有了可利用的时间，我就会争分夺秒地多写一些。正是在这种情况下，才终于把它完成了。我不能掩饰自己完成这个浩繁工作后的欣喜，如果我的这一工作能得到广大同行和读者的赞赏，那我就会更加倍地喜悦。

为了使读者对本书内容和结构有更多的了解，这里，先对一些问题作些交代，我想应该是必要的。

首先一个问题是阅读本书的预备知识。和学习任何较为深入的学科知识一样，我们对有意于阅读本书的读者在知识准备上做了假定：假定他们具有较好的高等数学和离散数学基础。自然，我们还假定他们具有程序设计——至少是一种语言的程序设计——的实践经验。除此之外，对于系统程序的了解，也会对学习课程有帮助，尽管对此并不特别要求。

其次，介绍本书的主要内容。本书共分为八章：

第一章，绪论。首先介绍什么是分布式系统。着重介绍其组成和特征，以及有关分布式系统研究的领域和主要问题。同时，介绍关于分布式算法的研究问题。

第二章，分布式系统。集中讨论关于分布式系统的一些问题，包括拓扑学、通讯系统的类型、文件系统、计算的方式、事件的顺序、同步、死锁处理、健全性、达到一致性及选举算法等等。本章所涉及的乃是分布式系统中人们较为关注的主要问题。

第三章，共享存储问题。这是人们所研究的、最初也是最深入研究的一种分布式算法。因此，这里列举了一系列共享存储的算法，包括艾森伯格——麦圭尔互斥算法、伯恩斯算法、兰勃特的“面包房”算法、彼得森——费希尔二进程互斥算法、彼得森——费希尔  $n$  进程算法等等，以及使用测试并建立机制的一些其它算法。在介绍这些算法过程中，有关彼得森——费希尔二进程互斥算法的断言式证明，既是让读者对于断言证明有所接触，也是作者为此书特意描绘的一笔——这曾是一个未解决的问题，而由作者一举解决了。

第四章，一致性问题。这里有关拜占庭问题的讨论，可以说在国内尚属罕见，这里，对它进行了解说。有关知识理论一节，则大部分是作者本人创立的。

第五章，静态网络算法。这一章所介绍的静态网络算法，包括路径算法、极大流量算法、极小代价流量以及应用等等，也反映了在基本静态网络算法中，人们最感兴趣的一些问题。

第六章，动态网络算法。动态通讯和 OSI 参考模型是本章的头一个论题。第二个论题多点对多点的通讯，也是本人给出的。第三个问题则是全局快照的问题。

第七章，关于分布式系统的模型。这里列出了三种较为流行的模型：I/O 自动机、Petri



网和通讯顺序进程 (CSP)。同时给出三种模型供读者来自行品味,也是本书的一个特色。本章中的许多例子,都是作者自己给出的。

第八章,形式化描述和验证。本章从一个例子谈起,逐步论述了有关形式化描述和验证所面对的各种问题。我们讨论了安全性、不变性、无干扰性,进而又讨论了消息传递的证明规则与证明活性性质,以及关于形式描述等问题。

本书主要参考书:

1. Nancy A Lynch & Kenneth J · Gold-  
man. Distributed Algorithms MIT Press. May  
1989.

2. A · J · Bernstein & P · M · Lewis.  
Concurrency in Programming and Database  
Systems. SUNY at Stony Brook, 1992.

3. M · W · Alford et al. Distributed Sys-  
tems. Methods and Tools for specification.  
Spring-verlag, 1985.

4. C · A · D Hoare. Communicating Se-  
quence Processes. Prentice-Hall, 1985.

5. J · L · Peterson. Petri Net Theory and  
The Modeling of Systems Prentice - Hall,  
1981.

本书还包括了作者的下列一些工作成果:

1. Peterson——Fischer 二进程互斥算法的断言式证明,《软件学报》1993年第4卷第3期;

2. 关于“知道”逻辑的组合逻辑算法及人类的逻辑推理过程,《全国人工智能会议文集》,西北工业大学出版社,1989年;

3. 关于时态逻辑的知识表示问题,《广东

省计算机学会第四届学术年会文集》，电子工业出版社，1992年；

4. 关于多点对多点的通讯；
5. 关于 I/O 自动机；
6. 关于 Petri 网。

这后三项虽然包含作者以往的工作，但是首次在这里发表。

为了帮助读者学习，我在每章之后附有一个小结，还有适量的习题供读者检查自己对该章的掌握程度。习题的选择是颇费心思的，我们鼓励读者开动脑筋来解决这些问题。这样做，必有好处。

从上述介绍可以看出，本书所涉及的问题以及算法，范围是相当广泛的。然而不能说分布式算法只限于这些类型。实际上，关于分布式操作系统和分布式数据库中的许多问题，我们并未涉及。即使是在所讨论的类型中，现存的算法也不仅限于此。所以，我们无意使本书成为有关分布式算法的一本百科全书。本书对于类型的选择及算法的介绍，都是经过了一番考虑的，它们既反映了作者的兴趣，也反映了作者对于这个领域的认识。这些类型的问题以及针对这些问题而提出的算法或者解决方法，应被认为是在分布式系统和分布式算法中具有较突出重要的地位。

著名计算机科学家 E·W·Dijkstra 曾经指出，当大约二十年前，计算机界面对并行性问题时，部分地由于它在技术上，不同于它所出现的环境，部分地则由于它同时引发的不确定性的历史事实，使得它导致了无穷无尽的困惑和混乱。想要摆脱这种困境，要求一个成熟的和具有献身精神的科学家进行艰难工作。

Dijkstra 在这里所指出的并发性，正是分布式系统的主要特征之一。今天，分布式系统比他说这话时又有了更大的进展。在这个领域里产生了更多的成果，同时也提出了更多的问题。这就更需要有他所说的那种科学家的艰难而有成效的工作。作者决不敢宣称是那种成熟的科学家，但仍然愿意并且自信有献身精神——既为了祖国的繁荣昌盛而献身，也为了整个科学技术的进步而献身。因此，如果本书的出版在这两方面都能起到一点点作用，也就心满意足了。

最后，在本书的写作过程中，作者除了得到许多同行的鼓励外，还得到我的学生们的许多帮助，尤其是审稿人对于本书提出了许多有益的建议和指正。我的家人——包括我远在印尼的母亲和弟弟妹妹们，和我的妻子、女儿、儿子给予我的理解与精神上的支持。没有这一切，我很难相信自己能有勇气和毅力来完成它。因此，在这里向他们所有人表示我由衷的谢意！

由于时间仓促，加之本人水平有限，书中的缺点错误在所难免，谨祈读者不吝指正。

苏运霖

1992 年于暨南大学

---

# 目 录

## 第一章

绪论 .....	(1)
1.1 分布式系统的兴起 .....	(1)
1.2 分布式系统的特征 .....	(7)
1.3 分布式系统的研究 .....	(7)
1.4 分布式算法的研究 .....	(10)
1.5 小结 .....	(12)
习题 .....	(13)

## 第二章

分布式系统 .....	(14)
2.1 拓朴学 .....	(14)
2.1.1 完全地连接的网络 .....	(15)
2.1.2 部分地连接的网络 .....	(16)
2.1.3 层次网络 .....	(17)

2.1.4	星形	(18)
2.1.5	环形网络	(19)
2.1.6	多路存取总线	(20)
2.2	通讯	(21)
2.2.1	路径策略	(22)
2.2.2	连接策略	(23)
2.2.3	竞争	(24)
2.2.4	安全性	(26)
2.2.5	设计策略	(29)
2.3	系统类型	(30)
2.3.1	计算机网络	(30)
2.3.2	局域网	(32)
2.4	文件系统	(34)
2.4.1	ARPA 网的 FTP	(34)
2.4.2	集中式的方法	(35)
2.4.3	分布式的方法	(35)
2.5	计算的方式	(36)
2.5.1	数据的移动	(36)
2.5.2	计算的移动	(37)
2.5.3	作业的移动	(38)
2.6	事件的顺序	(39)
2.6.1	在什么之前发生的关系	(39)
2.6.2	实现全序	(41)
2.7	同步	(43)
2.7.1	集中式的方法	(43)
2.7.2	完全分布的方法	(44)
2.7.3	标记传送的方法	(47)
2.8	死锁处理	(49)

2.8.1	时间标签排序方法	(49)
2.8.2	死锁检测	(51)
2.9	健全性	(58)
2.9.1	故障检测	(58)
2.9.2	重新布局	(59)
2.9.3	从故障中恢复	(60)
2.10	达到一致性	(61)
2.10.1	不可靠的通讯	(62)
2.10.2	出错进程	(63)
2.11	选举算法	(65)
2.11.1	霸道算法	(66)
2.11.2	环形算法	(68)
2.12	小结	(69)
	习题	(72)

### 第三章

#### 共享存储问题 (73)

3.1	使用共享存储的互斥及其实现	(74)
3.2	改进的互斥算法	(85)
3.2.1	艾森伯格—麦圭尔互斥算法	(86)
3.2.2	伯恩斯互斥算法	(89)
3.2.3	兰勃特的“面包房”算法	(98)
3.3	彼得森—费希尔二进程互斥算法	(103)
3.4	彼得森—费希尔 n 进程算法	(112)
3.5	测试并建立算法	(118)
3.5.1	实现互斥的一个简单的测试并建立算法	(118)
3.5.2	使用测试并建立算法的公平互斥	(120)

3.5.3 伯恩斯等的测试并建立算法 .....	(122)
3.6 小结 .....	(129)
习题 .....	(132)

## 第四章

### 一致性问题 .....

(134)

4.1 对于有关丢失消息时一致性的不可能性的一个 结果 .....	(135)
4.2 拜占庭一致性问题 .....	(138)
4.2.1 证实算法 .....	(149)
4.2.2 限制通讯费用 .....	(153)
4.2.3 关于进程的个数 .....	(155)
4.3 使用链自变量来建立不可能性结果 .....	(161)
4.3.1 链自变量 .....	(161)
4.3.2 费希尔—林奇下界定理 .....	(162)
4.3.3 停止的故障 .....	(171)
4.4 知识理论 .....	(179)
4.4.1 最优性 .....	(180)
4.4.2 知识的形式理论 .....	(183)
4.5 异步系统的一致性及随机一致性 .....	(190)
4.6 小结 .....	(197)
习题 .....	(200)

## 第五章

### 静态网络算法 .....

(204)

5.1 路径确定算法 .....	(205)
------------------	-------

5.1.1	极小跨越树算法 .....	(205)
5.1.2	加拉格尔—亨伯勒—斯皮拉算法 .....	(215)
5.2	领导者选举算法 .....	(227)
5.2.1	勒兰—张—罗伯特的领导者选举算法 .....	(229)
5.2.2	希尔伯格—辛克莱领导者选举算法 .....	(233)
5.2.3	彼得森领导者选举算法 .....	(235)
5.2.4	同步领导者选举算法 .....	(242)
5.3	极大流量算法 .....	(245)
5.3.1	求极大容量的算法 .....	(253)
5.3.2	求极大流量算法的分析 .....	(269)
5.4	多终端的极大流量算法 .....	(273)
5.4.1	可实现性 .....	(274)
5.4.2	分析 .....	(276)
5.4.3	综合 .....	(289)
5.4.4	多种商品的流量 .....	(298)
5.5	极小代价流程 .....	(299)
5.6	应用 .....	(303)
5.6.1	不同代表的系统 .....	(303)
5.6.2	PERT .....	(306)
5.6.3	最优通讯跨越树 .....	(312)
5.7	小结 .....	(319)
	习题 .....	(323)

## 第六章

### 动态网络算法 .....

(327)

6.1	网络通讯和 OSI 参考模型 .....	(329)
6.1.1	虚拟线路通讯 .....	(337)



6.1.2 半同步通讯 .....	(341)
6.1.3 数据网通讯 .....	(345)
6.2 多点对多点的通讯 .....	(348)
6.3 全局快照 .....	(369)
6.4 小结 .....	(378)
习题 .....	(380)

## 第七章

### 关于分布式系统的模型 .....

(382)

7.1 I/O 自动机 .....	(384)
7.1.1 定义和基本结果 .....	(386)
7.1.2 糖果机 .....	(397)
7.1.3 模拟多一多通讯的 I/O 自动机 .....	(404)
7.2 Petri 网 .....	(413)
7.3 通讯顺序进程 (CSP) .....	(424)
7.3.1 并发性 .....	(430)
7.3.2 图形 .....	(434)
7.3.3 例子: 就餐的哲学家 .....	(435)
7.3.4 符号的改变 .....	(441)
7.3.5 摘述 .....	(453)
7.3.6 确定进程的数学理论 .....	(454)
7.3.7 非正确性 .....	(462)
7.4 小结 .....	(464)
习题 .....	(466)

## 第八章

### 形式化描述和检验 .....

(468)

8.1 从一个例子谈起 .....	(469)
-------------------	-------