Distributed Computing

Principles, Algorithms, and Systems

Distributed computing deals with all forms of computing, information access, and information exchange across multiple processing platforms connected by computer networks. Design of distributed computing systems is a complex task. It requires a solid understanding of the design issues and an in-depth understanding of the theoretical and practical aspects of their solutions. This comprehensive textbook covers the fundamental principles and models underlying the theory, algorithms, and systems aspects of distributed computing.

Broad and detailed coverage of the theory is balanced with practical systems-related problems such as mutual exclusion, deadlock detection, authentication, and failure recovery. Algorithms are carefully selected, lucidly presented, and described without complex proofs. Simple explanations and illustrations are used to elucidate the algorithms. Emerging topics of significant impact, such as peer-to-peer networks and network security, are also covered.

With state-of-the-art algorithms, numerous illustrations, examples, and homework problems, this textbook is invaluable for advanced undergraduate and graduate students of electrical and computer engineering and computer science. Practitioners in data networking and sensor networks will also find this a valuable resource.

Ajay D. Kshemkalyani is a Professor in the Department of Computer Science, at the University of Illinois at Chicago. He was awarded his Ph.D. in Computer and Information Science in 1991 from The Ohio State University. Before moving to academia, he spent several years working on computer networks at IBM Research Triangle Park. In 1999, he received the National Science Foundation's CAREER Award. He is a Senior Member of the IEEE, and his principal areas of research include distributed computing, algorithms, computer networks, and concurrent systems. He currently serves on the editorial board of *Computer Networks*.

Mukesh Singhal is Full Professor and Gartner Group Endowed Chair in Network Engineering in the Department of Computer Science at the University of Kentucky. He was awarded his Ph.D. in Computer Science in 1986 from the University of Maryland, College Park. In 2003, he received the IEEE

Technical Achievement Award, and currently serves on the editorial boards for the *IEEE Transactions on Parallel and Distributed Systems* and the *IEEE Transactions on Computers*. He is a Fellow of the IEEE, and his principal areas of research include distributed systems, computer networks, wireless and mobile computing systems, performance evaluation, and computer security.

Distributed Computing

Principles, Algorithms, and Systems

Ajay D. Kshemkalyani
University of Illinois at Chicago, Chicago

and

Mukesh Singhal
University of Kentucky, Lexington

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521189842

© Cambridge University Press 2008

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2008
First paperback edition (with corrections) 2011

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-87634-6 Hardback
ISBN 978-0-521-18984-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

To my father Shri Digambar and
my mother Shrimati Vimala.
Ajay D. Kshemkalyani

To my mother Chandra Prabha Singhal,
my father Brij Mohan Singhal, and my
daughters Meenakshi, Malvika,
and Priyanka.
Mukesh Singhal

© in this web service Cambridge University Press

Contents

Preface *page* xv

1 Introduction 1
1.1 Definition 1
1.2 Relation to computer system components 2
1.3 Motivation 3
1.4 Relation to parallel multiprocessor/multicomputer systems 5
1.5 Message-passing systems versus shared memory systems 13
1.6 Primitives for distributed communication 14
1.7 Synchronous versus asynchronous executions 19
1.8 Design issues and challenges 22
1.9 Selection and coverage of topics 33
1.10 Chapter summary 34
1.11 Exercises 35
1.12 Notes on references 36
References 37

2 A model of distributed computations 39
2.1 A distributed program 39
2.2 A model of distributed executions 40
2.3 Models of communication networks 42
2.4 Global state of a distributed system 43
2.5 Cuts of a distributed computation 45
2.6 Past and future cones of an event 46
2.7 Models of process communications 47
2.8 Chapter summary 48
2.9 Exercises 48
2.10 Notes on references 48
References 49

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

3 Logical time 50
3.1 Introduction 50
3.2 A framework for a system of logical clocks 52
3.3 Scalar time 53
3.4 Vector time 55
3.5 Efficient implementations of vector clocks 59
3.6 Jard–Jourdan's adaptive technique 65
3.7 Matrix time 68
3.8 Virtual time 69
3.9 Physical clock synchronization: NTP 78
3.10 Chapter summary 81
3.11 Exercises 84
3.12 Notes on references 84
References 84

4 Global state and snapshot recording algorithms 87
4.1 Introduction 87
4.2 System model and definitions 90
4.3 Snapshot algorithms for FIFO channels 93
4.4 Variations of the Chandy–Lamport algorithm 97
4.5 Snapshot algorithms for non-FIFO channels 101
4.6 Snapshots in a causal delivery system 106
4.7 Monitoring global state 109
4.8 Necessary and sufficient conditions for consistent global snapshots 110
4.9 Finding consistent global snapshots in a distributed computation 114
4.10 Chapter summary 121
4.11 Exercises 122
4.12 Notes on references 122
References 123

5 Terminology and basic algorithms 126
5.1 Topology abstraction and overlays 126
5.2 Classifications and basic concepts 128
5.3 Complexity measures and metrics 135
5.4 Program structure 137
5.5 Elementary graph algorithms 138
5.6 Synchronizers 163
5.7 Maximal independent set (MIS) 169
5.8 Connected dominating set 171
5.9 Compact routing tables 172
5.10 Leader election 174

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

5.11 Challenges in designing distributed graph algorithms 175
5.12 Object replication problems 176
5.13 Chapter summary 182
5.14 Exercises 183
5.15 Notes on references 185
References 186

6 Message ordering and group communication 189
6.1 Message ordering paradigms 190
6.2 Asynchronous execution with synchronous communication 195
6.3 Synchronous program order on an asynchronous system 200
6.4 Group communication 205
6.5 Causal order (CO) 206
6.6 Total order 215
6.7 A nomenclature for multicast 220
6.8 Propagation trees for multicast 221
6.9 Classification of application-level multicast algorithms 225
6.10 Semantics of fault-tolerant group communication 228
6.11 Distributed multicast algorithms at the network layer 230
6.12 Chapter summary 236
6.13 Exercises 236
6.14 Notes on references 238
References 239

7 Termination detection 241
7.1 Introduction 241
7.2 System model of a distributed computation 242
7.3 Termination detection using distributed snapshots 243
7.4 Termination detection by weight throwing 245
7.5 A spanning-tree-based termination detection algorithm 247
7.6 Message-optimal termination detection 253
7.7 Termination detection in a very general distributed computing model 257
7.8 Termination detection in the atomic computation model 263
7.9 Termination detection in a faulty distributed system 272
7.10 Chapter summary 279
7.11 Exercises 279
7.12 Notes on references 280
References 280

8 Reasoning with knowledge 282
8.1 The muddy children puzzle 282
8.2 Logic of knowledge 283

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

8.3 Knowledge in synchronous systems 289
8.4 Knowledge in asynchronous systems 290
8.5 Knowledge transfer 298
8.6 Knowledge and clocks 300
8.7 Chapter summary 301
8.8 Exercises 302
8.9 Notes on references 303
References 303

9 Distributed mutual exclusion algorithms 305
9.1 Introduction 305
9.2 Preliminaries 306
9.3 Lamport's algorithm 309
9.4 Ricart–Agrawala algorithm 312
9.5 Singhal's dynamic information-structure algorithm 315
9.6 Lodha and Kshemkalyani's fair mutual exclusion algorithm 321
9.7 Quorum-based mutual exclusion algorithms 327
9.8 Maekawa's algorithm 328
9.9 Agarwal–El Abbadi quorum-based algorithm 331
9.10 Token-based algorithms 336
9.11 Suzuki–Kasami's broadcast algorithm 336
9.12 Raymond's tree-based algorithm 339
9.13 Chapter summary 348
9.14 Exercises 348
9.15 Notes on references 349
References 350

10 Deadlock detection in distributed systems 352
10.1 Introduction 352
10.2 System model 352
10.3 Preliminaries 353
10.4 Models of deadlocks 355
10.5 Knapp's classification of distributed deadlock detection algorithms 358
10.6 Mitchell and Merritt's algorithm for the single-resource model 360
10.7 Chandy–Misra–Haas algorithm for the AND model 362
10.8 Chandy–Misra–Haas algorithm for the OR model 364
10.9 Kshemkalyani–Singhal algorithm for the *P*-out-of-*Q* model 365
10.10 Chapter summary 374
10.11 Exercises 375
10.12 Notes on references 375
References 376

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

11 Global predicate detection 379
11.1 Stable and unstable predicates 379
11.2 Modalities on predicates 382
11.3 Centralized algorithm for relational predicates 384
11.4 Conjunctive predicates 388
11.5 Distributed algorithms for conjunctive predicates 395
11.6 Further classification of predicates 404
11.7 Chapter summary 405
11.8 Exercises 406
11.9 Notes on references 407
References 408

12 Distributed shared memory 410
12.1 Abstraction and advantages 410
12.2 Memory consistency models 413
12.3 Shared memory mutual exclusion 427
12.4 Wait-freedom 434
12.5 Register hierarchy and wait-free simulations 434
12.6 Wait-free atomic snapshots of shared objects 447
12.7 Chapter summary 451
12.8 Exercises 452
12.9 Notes on references 453
References 454

13 Checkpointing and rollback recovery 456
13.1 Introduction 456
13.2 Background and definitions 457
13.3 Issues in failure recovery 462
13.4 Checkpoint-based recovery 464
13.5 Log-based rollback recovery 470
13.6 Koo–Toueg coordinated checkpointing algorithm 476
13.7 Juang–Venkatesan algorithm for asynchronous checkpointing and recovery 478
13.8 Manivannan–Singhal quasi-synchronous checkpointing algorithm 483
13.9 Peterson–Kearns algorithm based on vector time 492
13.10 Helary–Mostefaoui–Netzer–Raynal communication-induced protocol 499
13.11 Chapter summary 505
13.12 Exercises 506
13.13 Notes on references 506
References 507

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

14 Consensus and agreement algorithms 510
14.1 Problem definition 510
14.2 Overview of results 514
14.3 Agreement in a failure-free system (synchronous or asynchronous) 515
14.4 Agreement in (message-passing) synchronous systems with failures 516
14.5 Agreement in asynchronous message-passing systems with failures 529
14.6 Wait-free shared memory consensus in asynchronous systems 544
14.7 Chapter summary 562
14.8 Exercises 563
14.9 Notes on references 564
References 565

15 Failure detectors 567
15.1 Introduction 567
15.2 Unreliable failure detectors 568
15.3 The consensus problem 577
15.4 Atomic broadcast 583
15.5 A solution to atomic broadcast 584
15.6 The weakest failure detectors to solve fundamental agreement problems 585
15.7 An implementation of a failure detector 589
15.8 An adaptive failure detection protocol 591
15.9 Exercises 596
15.10 Notes on references 596
References 596

16 Authentication in distributed systems 598
16.1 Introduction 598
16.2 Background and definitions 599
16.3 Protocols based on symmetric cryptosystems 602
16.4 Protocols based on asymmetric cryptosystems 615
16.5 Password-based authentication 622
16.6 Authentication protocol failures 625
16.7 Chapter summary 626
16.8 Exercises 627
16.9 Notes on references 627
References 628

17 Self-stabilization 631
17.1 Introduction 631
17.2 System model 632

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

17.3 Definition of self-stabilization 634
17.4 Issues in the design of self-stabilization algorithms 636
17.5 Methodologies for designing self-stabilizing systems 647
17.6 Communication protocols 649
17.7 Self-stabilizing distributed spanning trees 650
17.8 Self-stabilizing algorithms for spanning-tree construction 652
17.9 An anonymous self-stabilizing algorithm for 1-maximal independent set in trees 657
17.10 A probabilistic self-stabilizing leader election algorithm 660
17.11 The role of compilers in self-stabilization 662
17.12 Self-stabilization as a solution to fault tolerance 665
17.13 Factors preventing self-stabilization 667
17.14 Limitations of self-stabilization 668
17.15 Chapter summary 670
17.16 Exercises 670
17.17 Notes on references 671
References 671

18 Peer-to-peer computing and overlay graphs 677
18.1 Introduction 677
18.2 Data indexing and overlays 679
18.3 Unstructured overlays 681
18.4 Chord distributed hash table 688
18.5 Content addressible networks (CAN) 695
18.6 Tapestry 701
18.7 Some other challenges in P2P system design 708
18.8 Tradeoffs between table storage and route lengths 710
18.9 Graph structures of complex networks 712
18.10 Internet graphs 714
18.11 Generalized random graph networks 720
18.12 Small-world networks 720
18.13 Scale-free networks 721
18.14 Evolving networks 723
18.15 Chapter summary 727
18.16 Exercises 727
18.17 Notes on references 728
References 729

Index 731

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

Preface

Background

The field of distributed computing covers all aspects of computing and information access across multiple processing elements connected by any form of communication network, whether local or wide-area in the coverage. Since the advent of the Internet in the 1970s, there has been a steady growth of new applications requiring distributed processing. This has been enabled by advances in networking and hardware technology, the falling cost of hardware, and greater end-user awareness. These factors have contributed to making distributed computing a cost-effective, high-performance, and fault-tolerant reality. Around the turn of the millenium, there was an explosive growth in the expansion and efficiency of the Internet, which was matched by increased access to networked resources through the World Wide Web, all across the world. Coupled with an equally dramatic growth in the wireless and mobile networking areas, and the plummeting prices of bandwidth and storage devices, we are witnessing a rapid spurt in distributed applications and an accompanying interest in the field of distributed computing in universities, governments organizations, and private institutions.

Advances in hardware technology have suddenly made sensor networking a reality, and embedded and sensor networks are rapidly becoming an integral part of everyone's life – from the home network with the interconnected gadgets to the automobile communicating by GPS (global positioning system), to the fully networked office with RFID monitoring. In the emerging global village, distributed computing will be the centerpiece of all computing and information access sub-disciplines within computer science. Clearly, this is a very important field. Moreover, this evolving field is characterized by a diverse range of challenges for which the solutions need to have foundations on solid principles.

The field of distributed computing is very important, and there is a huge demand for a good comprehensive book. This book comprehensively covers all important topics in great depth, combining this with a clarity of explanation

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-18984-2 - Distributed Computing: Principles, Algorithms, and Systems
Ajay D. Kshemkalyani and Mukesh Singhal
Frontmatter
More information

and ease of understanding. The book will be particularly valuable to the academic community and the computer industry at large. Writing such a comprehensive book has been a Herculean task and there is a deep sense of satisfaction in knowing that we were able complete it and perform this service to the community.

Description, approach, and features

The book will focus on the fundamental principles and models underlying all aspects of distributed computing. It will address the principles underlying the theory, algorithms, and systems aspects of distributed computing. The manner of presentation of the algorithms is very clear, explaining the main ideas and the intuition with figures and simple explanations rather than getting entangled in intimidating notations and lengthy and hard-to-follow rigorous proofs of the algorithms. The selection of chapter themes is broad and comprehensive, and the book covers all important topics in depth. The selection of algorithms within each chapter has been done carefully to elucidate new and important techniques of algorithm design. Although the book focuses on foundational aspects and algorithms for distributed computing, it thoroughly addresses all practical systems-like problems (e.g., mutual exclusion, deadlock detection, termination detection, failure recovery, authentication, global state and time, etc.) by presenting the theory behind and algorithms for such problems. The book is written keeping in mind the impact of emerging topics such as *peer-to-peer computing* and *network security* on the foundational aspects of distributed computing.

Each chapter contains figures, examples, exercises, a summary, and references.

Readership

This book is aimed as a textbook for the following:

- Graduate students and Senior level undergraduate students in computer science and computer engineering.
- Graduate students in electrical engineering and mathematics. As wireless networks, peer-to-peer networks, and mobile computing continue to grow in importance, an increasing number of students from electrical engineering departments will also find this book necessary.
- Practitioners, systems designers/programmers, and consultants in industry and research laboratories will find the book a very useful reference because it contains state-of-the-art algorithms and principles to address various design issues in distributed systems, as well as the latest references.

© in this web service Cambridge University Press www.cambridge.org

Hard and soft prerequisites for the use of this book include the following:

- An undergraduate course in algorithms is required.
- Undergraduate courses in operating systems and computer networks would be useful.
- A reasonable familiarity with programming.

We have aimed for a very comprehensive book that will act as a single source for distributed computing models and algorithms. The book has both depth and breadth of coverage of topics, and is characterized by clear and easy explanations. None of the existing textbooks on distributed computing provides all of these features.

Acknowledgements

This book grew from the notes used in the graduate courses on distributed computing at the Ohio State University, the University of Illinois at Chicago, and at the University of Kentucky. We would like to thank the graduate students at these schools for their contributions to the book in many ways.

The book is based on the published research results of numerous researchers in the field. We have made all efforts to present the material in our own words and have given credit to the original sources of information. We would like to thank all the researchers whose work has been reported in this book. Finally, we would like to thank the staff of Cambridge University Press for providing us with excellent support in the publication of this book.

Access to resources

The following websites will be maintained for the book. Any errors and comments should be sent to ajayk@cs.uic.edu or singhal@cs.uky.edu. Further information about the book can be obtained from the authors' web pages:

- www.cs.uic.edu/∼ajayk/DCS-Book
- www.cs.uky.edu/∼singhal/DCS-Book.