DISTRIBUTED SYSTEMS

Concepts and Design

Fifth Edition

International Edition

George Coulouris
Jean Dollimore
Tim Kindberg
Gordon Blair

PEARSON

DISTRIBUTED SYSTEMS

Concepts and Design

Fifth Edition

This page intentionally left blank

DISTRIBUTED SYSTEMS

Concepts and Design

Fifth Edition

George Coulouris
Cambridge University

Jean Dollimore
formerly of Queen Mary, University of London

Tim Kindberg
matter 2 media

Gordon Blair
Lancaster University

International Edition contributions by:
Arup Kumar Bhattacharjee
RCC Institute of Information Technology, Kolkata

Soumen Mukherjee
RCC Institute of Information Technology, Kolkata

Addison-Wesley

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editorial Director: Marcia Horton
Editor-in-Chief: Michael Hirsch
Executive Editor: Matt Goldstein
Editorial Assistant: Chelsea Bell
Vice President, Marketing: Patrice Jones
Marketing Manager: Yezan Alayan
Marketing Coordinator: Kathryn Ferranti
Vice President, Production: Vince O'Brien
Managing Editor: Jeff Holcomb
Senior Production Project Manager: Marilyn Lloyd
Publisher, International Edition: Angshuman Chakraborty
Acquisitions Editor, International Edition: Arunabha Deb
Publishing Assistant, International Edition: Shokhi Shah
Senior Operations Supervisor: Alan Fischer
Manufacturing Buyer: Lisa McDowell
Art Director: Jayne Conte
Cover Designer: Suzanne Duda
Cover Image: Sky: © amygdala_imagery; Kite: © Alamy; Mobile phone: © yasinguneysu/iStock
Media Editor: Daniel Sandin
Media Project Manager: Wanda Rockwell
Cover Printer: Lehigh-Phoenix Color

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsoninternationaleditions.com

© Pearson Education Limited 2012

The rights of George Coulouris, Jean Dollimore, Tim Kindberg and Gordon Blair to be identified as authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Distributed Systems, Concepts and Design, 5th edition*, ISBN 978-0-13-214301-1 by George Coulouris, Jean Dollimore, Tim Kindberg and Gordon Blair published by Pearson Education © 2012.*

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a licence permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

Microsoft® and Windows® are registered trademarks of the Microsoft Corporation in the U.S.A. and other countries. Screen shots and icons reprinted with permission from the Microsoft Corporation. This book is not sponsored or endorsed by or affiliated with the Microsoft Corporation.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

10 9 8 7 6 5 4 3 2 1
14 13 12 11 10

Typeset in Times New Roman by authors using FrameMaker
Printed and bound by Edwards Brothers in The United States of America

The publisher's policy is to use paper manufactured from sustainable forests.

ISBN 10: 0-273-76059-9
ISBN 13: 978-0-273-76059-7

Addison-Wesley
is an imprint of

PEARSON

www.pearsonhighered.com

4 INTERPROCESS COMMUNICATION 161
4.1 Introduction 162
4.2 The API for the Internet protocols 163
4.3 External data representation and marshalling 174
4.4 Multicast communication 185
4.5 Network virtualization: Overlay networks 190
4.6 Case study: MPI 194
4.7 Summary 197

5 REMOTE INVOCATION 201
5.1 Introduction 202
5.2 Request-reply protocols 203
5.3 Remote procedure call 211
5.4 Remote method invocation 220
5.5 Case study: Java RMI 233
5.6 Summary 241

6 INDIRECT COMMUNICATION 245
6.1 Introduction 246
6.2 Group communication 248
6.3 Publish-subscribe systems 258
6.4 Message queues 270
6.5 Shared memory approaches 278
6.6 Summary 290

7 OPERATING SYSTEM SUPPORT 295
7.1 Introduction 296
7.2 The operating system layer 297
7.3 Protection 300
7.4 Processes and threads 302
7.5 Communication and invocation 319
7.6 Operating system architecture 330
7.7 Virtualization at the operating system level 334
7.8 Summary 347

8 DISTRIBUTED OBJECTS AND COMPONENTS 351
8.1 Introduction 352
8.2 Distributed objects 353
8.3 Case study: CORBA 356
8.4 From objects to components 374
8.5 Case studies: Enterprise JavaBeans and Fractal 380
8.6 Summary 394

9 WEB SERVICES 397
9.1 Introduction 398
9.2 Web services 400
9.3 Service descriptions and IDL for web services 416
9.4 A directory service for use with web services 420
9.5 XML security 422
9.6 Coordination of web services 427
9.7 Applications of web services 429
9.8 Summary 435

10 PEER-TO-PEER SYSTEMS 439
10.1 Introduction 440
10.2 Napster and its legacy 444
10.3 Peer-to-peer middleware 446
10.4 Routing overlays 449
10.5 Overlay case studies: Pastry, Tapestry 452
10.6 Application case studies: Squirrel, OceanStore, Ivy 465
10.7 Summary 474

11 SECURITY 479
11.1 Introduction 480
11.2 Overview of security techniques 488
11.3 Cryptographic algorithms 500
11.4 Digital signatures 509
11.5 Cryptography pragmatics 516
11.6 Case studies: Needham–Schroeder, Kerberos, TLS, 802.11 WiFi 519
11.7 Summary 534

12 DISTRIBUTED FILE SYSTEMS 537
12.1 Introduction 538
12.2 File service architecture 546
12.3 Case study: Sun Network File System 552
12.4 Case study: The Andrew File System 564
12.5 Enhancements and further developments 573
12.6 Summary 579

13 NAME SERVICES 581
13.1 Introduction 582
13.2 Name services and the Domain Name System 585
13.3 Directory services 600
13.4 Case study: The Global Name Service 601
13.5 Case study: The X.500 Directory Service 604
13.6 Summary 608

14 TIME AND GLOBAL STATES 611
14.1 Introduction 612
14.2 Clocks, events and process states 613
14.3 Synchronizing physical clocks 615
14.4 Logical time and logical clocks 623
14.5 Global states 626
14.6 Distributed debugging 635
14.7 Summary 642

15 COORDINATION AND AGREEMENT 645
15.1 Introduction 646
15.2 Distributed mutual exclusion 649
15.3 Elections 657
15.4 Coordination and agreement in group communication 662
15.5 Consensus and related problems 675
15.6 Summary 687

16 TRANSACTIONS AND CONCURRENCY CONTROL 691
16.1 Introduction 692
16.2 Transactions 695
16.3 Nested transactions 706
16.4 Locks 708
16.5 Optimistic concurrency control 723
16.6 Timestamp ordering 727
16.7 Comparison of methods for concurrency control 734
16.8 Summary 736

17 DISTRIBUTED TRANSACTIONS 743
17.1 Introduction 744
17.2 Flat and nested distributed transactions 744
17.3 Atomic commit protocols 747
17.4 Concurrency control in distributed transactions 756
17.5 Distributed deadlocks 759
17.6 Transaction recovery 767
17.7 Summary 777

18 REPLICATION 781
18.1 Introduction 782
18.2 System model and the role of group communication 784
18.3 Fault-tolerant services 791
18.4 Case studies of highly available services:
The gossip architecture, Bayou and Coda 798
18.5 Transactions with replicated data 818
18.6 Summary 830

19 MOBILE AND UBIQUITOUS COMPUTING 833
19.1 Introduction 834
19.2 Association 843
19.3 Interoperation 851
19.4 Sensing and context awareness 860
19.5 Security and privacy 873
19.6 Adaptation 882
19.7 Case study: Cooltown 887
19.8 Summary 894

20 DISTRIBUTED MULTIMEDIA SYSTEMS 897
20.1 Introduction 898
20.2 Characteristics of multimedia data 902
20.3 Quality of service management 903
20.4 Resource management 913
20.5 Stream adaptation 915
20.6 Case studies: Tiger, BitTorrent and End System Multicast 917
20.7 Summary 929

21 DESIGNING DISTRIBUTED SYSTEMS: GOOGLE CASE STUDY 931
21.1 Introduction 932
21.2 Introducing the case study: Google 933
21.3 Overall architecture and design philosophy 938
21.4 Underlying communication paradigms 944
21.5 Data storage and coordination services 951
21.6 Distributed computation services 972
21.7 Summary 980

REFERENCES 983

INDEX 1041