Texts in Theoretical Computer Science
An EATCS Series

Editors: W. Brauer G. Rozenberg A. Salomaa
On behalf of the European Association
for Theoretical Computer Science (EATCS)

Advisory Board: G. Ausiello M. Broy C.S. Calude
A. Condon D. Harel J. Hartmanis T. Henzinger
J. Hromkovič N. Jones T. Leighton M. Nivat
C. Papadimitriou D. Scott

J. Hromkovič R. Klasing A. Pelc
P. Ružička† W. Unger

Dissemination of Information in Communication Networks

Broadcasting, Gossiping, Leader Election, and Fault-Tolerance

With 73 Figures

Springer

Authors

Prof. Dr. Juraj Hromkovič
ETH Zentrum, RZ F2
Department of Computer Science
Swiss Federal Institute of Technology
8092 Zürich, Switzerland
juraj.hromkovic@inf.ethz.ch

Prof. Dr. Peter Ružička†

Dr. Ralf Klasing
CNRS - I3S & INRIA Sophia-Antipolis
2004 Route des Lucioles, BP 93
06902 Sophia Antipolis Cedex, France
Ralf.Klasing@sophia.inria.fr

Dr. Andrzej Pelc
Départment d'informatique
Université du Québec en Outaouais
C.P. 1250, Succursale Hull
Gatineau (Québec) J8X 3X7, Canada
andrzej.pelc@ugo.ca

Dr. Walter Unger
Lehrstuhl für Informatik I
Aachen University RWTH
Ahornstrasse 55
52074 Aachen, Germany
quax@i1.Informatik.RWTH-Aachen.de

Series Editors

Prof. Dr. Wilfried Brauer
Institut für Informatik der TUM
Boltzmannstrasse 3
85748 Garching, Germany
Brauer@informatik.tu-muenchen.de

Prof. Dr. Grzegorz Rozenberg
Leiden Institute of Advanced Computer Science
University of Leiden
Niels Bohrweg 1
2333 CA Leiden, The Netherlands
rozenber@liacs.nl

Prof. Dr. Arto Salomaa
Turku Centre for Computer Science
Lemminkäisenkatu 14 A
20520 Turku, Finland
asalomaa@utu.fi

Illustrations
Ingrid Zamečnikova

Library of Congress Control Number: 2004117337

ACM Computing Classification (1998): B.4.3, C.2.0-1, F.2.2, G.2.1-2, G.3

ISBN 3-540-00846-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

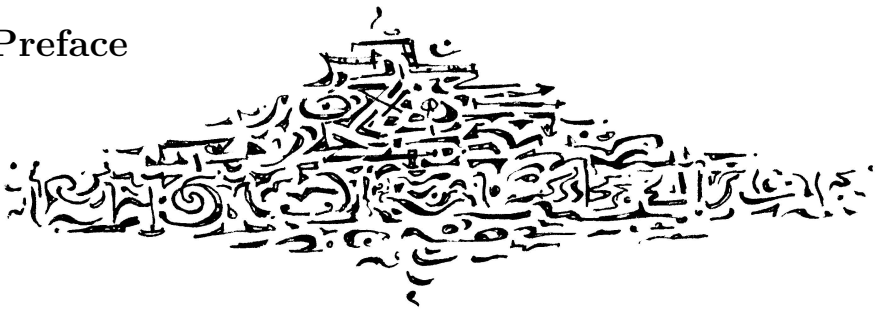
Cover design: KünkelLopka, Heidelberg
Typesetting: Camera ready by authors
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper SPIN: 10920657 45/3142/YL - 5 4 3 2 1 0

To our dear friend
and coauthor Peter Ružička,
in memoriam.

It is necessary to live in such a way
that death might seem
to be the greatest injustice done to us.
What is really important
is to give life a real sense:
to identify oneself with the yearnings of people,
to be responsible for the way things are
and to be present in all the struggles
which shape our existence.

Emmanuel Robles

Preface



Due to the development of hardware technologies (such as VLSI) in the early 1980s, the interest in parallel and distributive computing has been rapidly growing and in the late 1980s the study of parallel algorithms and architectures became one of the main topics in computer science. To bring the topic to educators and students, several books on parallel computing were written. The involved textbook "Introduction to Parallel Algorithms and Architectures" by F. Thomson Leighton in 1992 was one of the milestones in the development of parallel architectures and parallel algorithms. But in the last decade or so the main interest in parallel and distributive computing moved from the design of parallel algorithms and expensive parallel computers to the new distributive reality – the world of interconnected computers that cooperate (often asynchronously) in order to solve different tasks. **Communication** became one of the most frequently used terms of computer science because of the following reasons:

(i) Considering the high performance of current computers, the communication is often more time consuming than the computing time of processors. As a result, the capacity of communication channels is the bottleneck in the execution of many distributive algorithms.

(ii) Many tasks in the Internet are pure communication tasks. We do not want to compute anything, we only want to execute some information exchange or to extract some information as soon as possible and as cheaply as possible. Also, we do not have a central database involving all basic knowledge. Instead, we have a distributed memory where the basic knowledge is distributed among the local memories of a large number of different computers.

The growing importance of solving pure communication tasks in the interconnected world is the main motivation for writing this book. The main goals of this material are:

(i) to provide a monograph that surveys the main methods, results and research problems related to the design and analysis of communication algorithms (strategies) under different technological constraints; and
(ii) to provide an introductory textbook in the field of information dissemination in interconnection networks with a special emphasis on broadcast, information collection, gossip, leader election, and related tasks.

Our work is divided into two parts. This first textbook is devoted to the classical, direct communication between connected pairs of nodes of a communication network and to the related communication tasks such as broadcasting, gossiping, and leader election. The forthcoming part focuses on the fast communication via fixed paths between senders and receivers, which is based on new technologies such as optical networks, ATM networks, and wireless networks (for instance, mobile phones and radio networks).

This book aims to be a textbook accessible for students as well as a monograph that surveys the research on communication, presents the border between the known and the unknown, and can so be of interest to researchers and professionals, too.

We would like to thank Manuel Wahle for the successful embedding of our manuscripts into a common style, and for his technical help concerning the typical problems that occur when several people try to write something together. We are indebted to Ingrid Zámečniková for her original illustrations. The (as always) excellent cooperation with Alfred Hofmann, Ingeborg Mayer, and Ronan Nugent from Springer is gratefully acknowledged.

This textbook is devoted to our dear friend and coauthor Peter Ružička, who died during the work on this project. Slovakian computer science lost in him one of its greatest personalities, one of those who wrote the computer science history in Czechoslovakia. Peter was an excellent researcher, and a beloved teacher, who was able to inspire his students for the study and the investigation of the topic of his interest in a fascinating way. But, first of all, he was a man, and to express what we mean and feel by this, it is impossible to find anything better than the following words of Antoine de Saint-Exupéry:

Only he who carries inside a greater personality
than himself deserves to be called a man.

All quotations forthcoming in this textbook remember Peter and present his views and his way of living as we were able to understand them.

Juraj Hromkovič, Ralf Klasing,
December 2004 Andrzej Pelc, Walter Unger

Contents

1 Introduction 1
1.1 Motivation and Aims 1
1.2 Organization of the Book 2

Part I The Telegraph and Telephone Modes

2 Fundamentals 7
2.1 Introduction 7
2.2 Graphs 7
2.3 Combinatorics 20
2.4 Communication Tasks and Algorithms 25
2.5 General Properties of the Complexity of Communication Algorithms 33
2.6 Fundamental Interconnection Networks 41

3 Broadcasting 51
3.1 Introduction 51
3.2 Basic Facts and Techniques 52
3.3 Upper Bounds for Common Networks 59
3.4 Lower Bounds for Bounded-Degree Graphs 66
3.5 Overview on Broadcasting in Concrete Networks 78
3.6 Approximation Algorithms 79
3.6.1 General Observations 80
3.6.2 Two Approximation Algorithms 82
3.6.3 Bibliographical Remarks 90

4 Gossiping 93
4.1 Introduction 93
4.2 General Observations 94
4.3 Gossiping in Graphs with Small Bisection Width 100

4.4 Gossiping in Hypercube-like Networks of Constant Degree 123
4.5 Gossiping in Complete Graphs 126
4.6 Overview 134

5 Systolic Communication 137
5.1 Introduction 137
5.2 Definitions 139
5.3 Systolic Broadcast 143
5.4 General Observations for Systolic Gossip 145
5.5 Special Systolic Modes 147
5.6 Systolic Gossip in Concrete Networks 148
5.6.1 Systolic Gossip on a Path 149
5.6.2 Systolic Gossip on a Cycle 161
5.6.3 Systolic Gossip on Complete Trees 162
5.6.4 Systolic Gossip on Grids 172
5.6.5 Systolic Gossip on Butterfly and CCC Networks 175
5.7 Bibliographical Remarks 180

6 Fault-Tolerance 183
6.1 Introduction 183
6.1.1 Fault Distribution 184
6.1.2 Fault Classification 185
6.1.3 Flexibility of Broadcasting Algorithms 186
6.2 Preliminary Results 186
6.3 The Bounded-Fault Model 190
6.3.1 Permanent Link Faults 190
6.3.2 Permanent Node Faults 193
6.3.3 Transmission Faults 201
6.4 The Probabilistic Fault Model 211
6.4.1 Crash Faults 211
6.4.2 Byzantine Faults 216

Part II Distributed Networks

7 Broadcast on Distributed Networks 229
7.1 Broadcasting on General Networks 229
7.1.1 Distributed Depth-First Search 229
7.1.2 Distributed Breadth-First Search 235
7.1.3 Basic Lower Bound on Message Complexity of Broadcast 240
7.2 Broadcast on Tori 246
7.2.1 Upper Bound 246
7.2.2 Lower Bound 250
7.2.3 The Growth of the Core Graph 253
7.3 Broadcast on Hypercubes 257

7.3.1 Preliminaries . . . 257
7.3.2 Partial Broadcasting and Orientation . . . 259
7.3.3 Bit-Efficient Partial Broadcasting Algorithm . . . 261
7.3.4 Applications – Optimal Computable Masks for Special Target Sets . . . 263
7.3.5 The Broadcasting Algorithm . . . 265
7.3.6 Bit-Optimal Broadcasting in Time n . . . 266

8 Leader Election in Asynchronous Distributed Networks . . . 267
8.1 Distributed Point-to-Point Network Model . . . 267
8.2 Leader Election on Distributed Networks . . . 271
8.3 Leader Election on Unidirectional Rings with Optimal Average Performance . . . 271
8.4 Leader Election on Unidirectional Rings – Worst Case Analysis 276
8.5 Leader Election on Bidirectional Rings – Worst Case Analysis . 283
8.5.1 Election Algorithm Due to van Leeuwen and Tan . . . 285
8.5.2 Correctness and Message Complexity Analysis . . . 287
8.6 Leader Election on Complete Networks . . . 290
8.6.1 $O(N \log N)$ Algorithm . . . 290
8.6.2 $\Omega(N \log N)$ Lower Bound . . . 294
8.7 Leader Election on Hypercubes . . . 297
8.7.1 Preliminary Notions . . . 297
8.7.2 Leader Election . . . 298
8.7.3 Warrior Phase . . . 298
8.7.4 King Phase . . . 300
8.8 Leader Election on Synchronous Rings . . . 315

9 Fault-Tolerant Broadcast in Distributed Networks . . . 317
9.1 Consensus Problem . . . 317
9.1.1 Fault-Tolerant Broadcast with Unsigned Messages . . . 319
9.1.2 Fault-Tolerant Broadcast with Signed Messages . . . 323
9.2 Broadcasting in Synchronous Networks with Dynamic Faults . . 326
9.2.1 Broadcast in Hypercubes with Dynamic Faults . . . 330
9.2.2 Broadcast in Tori with Dynamic Faults . . . 333
9.2.3 Broadcast in Star Graphs with Dynamic Faults . . . 337

References . . . 341

Index . . . 357