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

Design and Analysis of Randomized Algorithms

Introduction to Design Paradigms

With 23 Figures

Springer

Author

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

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 Zámečniková

Library of Congress Control Number: 2005926236

ACM Computing Classification (1998): F.2, F.1, G.2

ISBN-10 3-540-23949-9 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-23949-9 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 author
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper 45/3142/YL - 5 4 3 2 1 0

To my teachers

O. Demáček

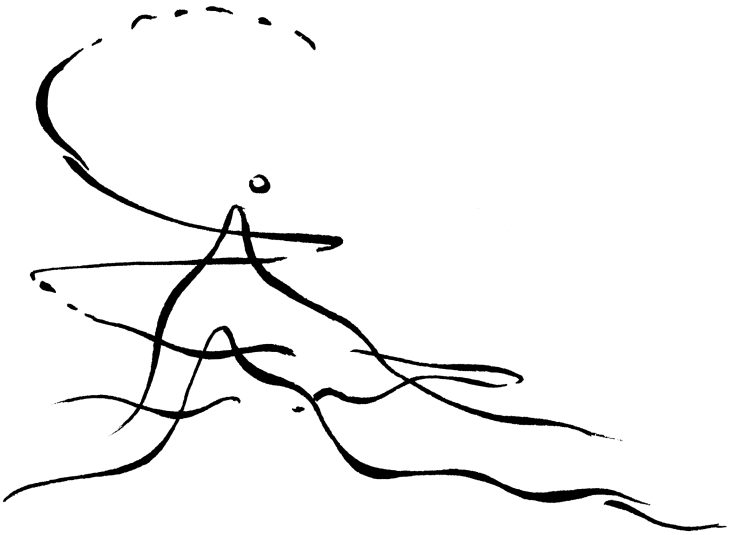
P. Ďuriš

R. Hammerová

B. Rovan

V. Šimčisková

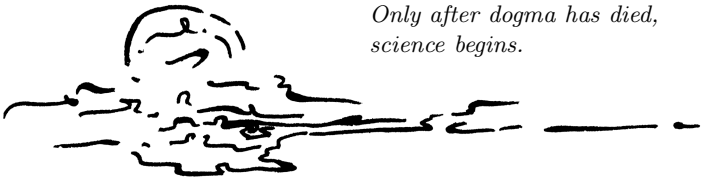
E. Toman



As soon as man is,
man has to look for
to be.
And as one is looking for
to be,
one has to be oneself
and not someone else,
as very often is the case.

Jan Werich

Preface



Only after dogma has died,
science begins.

Galileo Galilei

Randomization has become a standard approach in algorithm design. Efficiency and simplicity are the main features of randomized algorithms that often made randomization a miraculous springboard for solving complex problems in various applications. Especially in the areas of communication, cryptography, data management, and discrete optimization, randomization tends to be an indispensable part of the development of software systems. We know several situations and algorithmic tasks for which any deterministic approach leads to so much computer work that it is impossible to perform it in practice, but for which randomized methods can be successfully applied. This huge gain of going from an intractable amount of computer work[1] (computational complexity) to short computations of randomized algorithms is paid for by the risk of computing a wrong output. But the probability of executing an erroneous computation of a randomized algorithm can often be reduced to below the probability of winning the first prize in a lottery in the first attempt. This means that one pays for saving a huge amount of computing time with a very small loss in the degree of reliability. How to orchestrate such huge quantitative jumps in computational complexity by small relaxations of reliability (correctness) constraints and why such big gains for small "prices" are at all possible are the topics of this book.

Despite many spectacular and surprising results and successful applications of randomized algorithms, the basic concepts of randomization are not sufficiently broadcasted in academic education. One of the main reasons for this unpleasant circumstance is that simple explanations and transparent presentations of the discoveries in randomized computing are missing in textbooks for introductory courses that could be available even to non-scientists and beginners. This is usually the situation when principal contributions and concepts of a particular scientific discipline are not recognized as paradigms in the broad community, and are even considered to be too difficult to be presented in basic courses. The aim of this textbook is to close this gap. We focus

[1] Requiring, for instance, billions of years on the fastest computers.

on a transparent explanation of the paradigms of the design of randomized algorithms. This first book of our series on the design and analysis of randomized algorithms provides a readable introduction to this topic, giving an exhaustive, technical survey of the best and most important results achieved. Providing an accessible ticket to randomization we would like to encourage colleagues not working in the area of randomization to present some randomized concepts in their courses, or even give courses specialized on randomized algorithm design. In this way we would like to contribute to speeding up the inclusion of paradigmatic results on randomized computing in the educational mainstream of computer science.

The didactic method of this book is similar to that in our textbook *Theoretical Computer Science*. The main strategies are called "simplicity," "transparency," and "less is sometimes more." Especially in this first book in the series, clarity takes priority over the presentation of the current state of research and development. When a transparent argument of a weaker result can bring across the idea succinctly, then we will opt for it instead of presenting a strong but technically demanding and confusing argument of the best known result. Throughout this book, we work systematically, taking small steps to travel from the simple to the complicated, and so avoid any interruption in thoughts. We are far from trying to present too many results. Instead, we take the time and space to explain what we want to explain in detail, and also in the general context of our scientific discipline. We also dedicate time to the development of informal ideas, and ways of thinking in this area.

I would like to express my deepest thanks to Dirk Bongartz, Hans-Joachim Böckenhauer, and Manuel Wahle for carefully reading the whole manuscript and for their numerous comments and suggestions. Very special thanks go to Bagdat Aslan and Manuel Wahle for their technical help on this manuscript. The excellent cooperation with Alfred Hofmann and Ingeborg Mayer of Springer is gratefully acknowledged. Last but not least I would like to cordially thank Ingrid Zámečniková for her illustrations and Tanja for her collection of citations.

Aachen,
January 2005 *Juraj Hromkovič*

Contents

1 Introduction 1
1.1 What Is Randomness and Does There Exist True Randomness? 1
1.2 Randomness as a Source of Efficiency – an Exemplary Application 5
1.3 Concept of the Book 11
1.4 To the Student 14
1.5 To the Teacher 16

2 Fundamentals 19
2.1 Objectives 19
2.2 Elementary Probability Theory 20
2.3 Models of Randomized Algorithms 37
2.4 Classification of Randomized Algorithms 51
2.5 Classification of Randomized Algorithms for Optimization Problems 72
2.6 Paradigms of the Design of Randomized Algorithms 87
2.7 Summary 96

3 Foiling the Adversary 101
3.1 Objectives 101
3.2 Hashing 102
3.3 Universal Hashing 109
3.4 Online Algorithms 116
3.5 Randomized Online Algorithms 120
3.6 Summary 128

4 Fingerprinting 131
4.1 Objectives 131
4.2 Communication Protocols 133
4.3 The Substring Problem 139

4.4 Verification of Matrix Multiplication 141
4.5 Equivalence of Two Polynomials 144
4.6 Summary 149

5 Success Amplification and Random Sampling 153
5.1 Objectives 153
5.2 Efficient Amplification by Repeating Critical Computation Parts 154
5.3 Repeated Random Sampling and Satisfiability 166
5.4 Random Sampling and Generating Quadratic Nonresidues 174
5.5 Summary 181

6 Abundance of Witnesses 183
6.1 Objectives 183
6.2 Searching for Witnesses for Primality Testing 184
6.3 Solovay-Strassen Algorithm for Primality Testing 192
6.4 Generation of Random Primes 202
6.5 Summary 206

7 Optimization and Random Rounding 209
7.1 Objectives 209
7.2 Relaxation to Linear Programming 210
7.3 Random Rounding and MAX-SAT 216
7.4 Combining Random Sampling and Random Rounding 222
7.5 Summary 225

A Fundamentals of Mathematics 227
A.1 Objectives 227
A.2 Algebra and Number Theory 228
A.3 Combinatorics 256
A.4 Summary 264

References 267

Index 271