Stochastic Networks

Communication networks underpin our modern world, and provide fascinating and challenging examples of large-scale stochastic systems. Randomness arises in communication systems at many levels: for example, the initiation and termination times of calls in a telephone network, or the statistical structure of the arrival streams of packets at routers in the Internet. How can routing, flow control and connection acceptance algorithms be designed to work well in uncertain and random environments?

This compact introduction illustrates how stochastic models can be used to shed light on important issues in the design and control of communication networks. It will appeal to readers with a mathematical background wishing to understand this important area of application, and to those with an engineering background who want to grasp the underlying mathematical theory. Each chapter ends with exercises and suggestions for further reading.

FRANK KELLY is Professor of the Mathematics of Systems at the University of Cambridge. His main research interests are in random processes, networks and optimization. He is especially interested in applications to the design and control of networks and to the understanding of self-regulation in large-scale systems.

ELENA YUDOVINA is a postdoctoral research fellow at the University of Michigan. Her research interests are in applications of queueing theory. She received her PhD from the University of Cambridge, where her interest in the subject was sparked by a course on stochastic networks taught by Frank Kelly.

INSTITUTE OF MATHEMATICAL STATISTICS TEXTBOOKS

Editorial Board D. R. Cox (University of Oxford) A. Agresti (University of Florida) B. Hambly (University of Oxford) S. Holmes (Stanford University) X.-L. Meng (Harvard University)

IMS Textbooks give introductory accounts of topics of current concern suitable for advanced courses at master's level, for doctoral students and for individual study. They are typically shorter than a fully developed textbook, often arising from material created for a topical course. Lengths of 100–290 pages are envisaged. The books typically contain exercises.

Stochastic Networks

FRANK KELLY University of Cambridge

ELENA YUDOVINA University of Michigan, Ann Arbor



CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is a part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org Information on this title: www.cambridge.org/9781107035775

© Frank Kelly and Elena Yudovina 2014

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 2014

Printed in the United Kingdom by CPI Group Ltd, Croydon CR0 4YY

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

ISBN 978-1-107-03577-5 Hardback ISBN 978-1-107-69170-4 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.

Contents

Prefe	reface		
	Overview	1	
	Queueing and loss networks	2 4	
	Decentralized optimization		
	Random access networks	5	
	Broadband networks	6	
	Internet modelling	8	
	Part I	11	
1	Markov chains	13	
1.1	Definitions and notation	13	
1.2	Time reversal	16	
1.3	Erlang's formula	18	
1.4	Further reading	21	
2	Queueing networks	22	
2.1	An $M/M/1$ queue	22	
2.2	A series of $M/M/1$ queues	24	
2.3	Closed migration processes	26	
2.4	Open migration processes	30	
2.5	Little's law	36	
2.6	Linear migration processes	39	
2.7	Generalizations	44	
2.8	Further reading	48	
3	Loss networks	49	
3.1	Network model	49	
3.2	Approximation procedure	51	
3.3	Truncating reversible processes	52	

> vi 3.4 57 Maximum probability 3.5 A central limit theorem 61 3.6 Erlang fixed point 67 3.7 Diverse routing 71 3.8 Further reading 81 Part II 83 4 **Decentralized optimization** 85 4.1 An electrical network 86 4.2 Road traffic models 92 4.3 Optimization of queueing and loss networks 101 4.4 Further reading 107 5 **Random access networks** 108 5.1 The ALOHA protocol 109 5.2 Estimating backlog 115 5.3 Acknowledgement-based schemes 119 5.4 Distributed random access 125 5.5 Further reading 132 6 Effective bandwidth 133 6.1 Chernoff bound and Cramér's theorem 134 6.2 Effective bandwidth 138 6.3 Large deviations for a queue with many sources 143 6.4 Further reading 148 Part III 149 7 **Internet congestion control** 151 7.1 Control of elastic network flows 151 7.2 Notions of fairness 158 7.3 A primal algorithm 162 Modelling TCP 7.4 166 75 What is being optimized? 170

Contents

8	Flow level Internet models	186
7.9	Further reading	185
7.8	Modelling a switch	177
7.7	Time delays	173
7.6	A dual algorithm	172
1.5	what is being optimized.	170

		Contents	vii
8.1	Evolution of	of flows	186
8.2 α -fair rate allocations			187
8.3	Stability of	$f \alpha$ -fair rate allocations	189
8.4	What can g	go wrong?	192
8.5	Linear netw	work with proportional fairness	195
8.6	Further rea	ding	199
Appen	odix A	Continuous time Markov processes	201
Appendix B		Little's law	204
Appendix C		Lagrange multipliers	207
Appendix D		Foster–Lyapunov criteria	210
Refere Index	nces		217 221

Preface

Communication networks underpin our modern world, and provide fascinating and challenging examples of large-scale stochastic systems. Randomness arises in communication systems at many levels: for example, the initiation and termination times of calls in a telephone network, or the statistical structure of the arrival streams of packets at routers in the Internet. How can routing, flow control, and connection acceptance algorithms be designed to work well in uncertain and random environments? And can we design these algorithms using simple local rules so that they produce coherent and purposeful behaviour at the macroscopic level?

The first two parts of the book will describe a variety of classical models that can be used to help understand the behaviour of large-scale stochastic networks. Queueing and loss networks will be studied, as well as random access schemes and the concept of an effective bandwidth. Parallels will be drawn with models from physics, and with models of traffic in road networks.

The third part of the book will study more recently developed models of packet traffic and of congestion control algorithms in the Internet. This is an area of some practical importance, with network operators, content providers, hardware and software vendors, and regulators actively seeking ways of delivering new services reliably and effectively. The complex interplay between end-systems and the network has attracted the attention of economists as well as mathematicians and engineers.

We describe enough of the technological background to communication networks to motivate our models, but no more. Some of the ideas described in the book are finding application in financial, energy, and economic networks as computing and communication technologies transform these areas. But communication networks currently provide the richest and bestdeveloped area of application within which to present a connected account of the ideas.

The lecture notes that have become this book were used for a Mas-

Х

Preface

ters course ("Part III") in the Faculty of Mathematics at the University of Cambridge. This is a one-year postgraduate course which assumes a mathematically mature audience, albeit from a variety of different mathematical backgrounds. Familiarity with undergraduate courses on Optimization and Markov Chains is helpful, but not absolutely necessary. Appendices are provided on continuous time Markov processes, Little's law, Lagrange multipliers, and Foster–Lyapunov criteria, reviewing the material needed. At Cambridge, students may attend other courses where topics touched on in these notes, for example Poisson point processes, large deviations, or game theory, are treated in more depth, and this course can serve as additional motivation.

Suggestions on further reading are given at the end of each chapter; these include reviews where the historical development and attribution of results can be found – we have not attempted this here – as well as some recent papers which give current research directions.

The authors are grateful to students of the course at Cambridge (and, for one year, Stanford) for their questions and insights, and especially to Joe Hurd, Damon Wischik, Gaurav Raina, and Neil Walton, who produced earlier versions of these notes.

Frank Kelly and Elena Yudovina