Stochastic Network Calculus

Yuming Jiang · Yong Liu

Stochastic Network Calculus

Springer

Yuming Jiang, BSc, MEng, PhD
Department of Telematics
Norwegian University of Science and Technology
Trondheim
Norway

Yong Liu, BEng, MEng, PhD
Optical Networks Lab
Department of Electrical and Computer Engineering
National University of Singapore
Singapore

ISBN: 978-1-84800-126-8 e-ISBN: 978-1-84800-127-5
DOI: 10.1007/978-1-84800-127-5

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

Library of Congress Control Number: 2008934595

© Springer-Verlag London Limited 2008
Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.
The use of 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 laws and regulations and therefore free for general use.
The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed on acid-free paper

Springer Science+Business Media
springer.com

To our families

Preface

Network calculus is a theory dealing with queuing systems found in computer networks. Its focus is on performance guarantees. Central to the theory is the use of alternate algebras such as the min-plus algebra to transform complex network systems into analytically tractable systems. To simplify the analysis, another idea is to characterize traffic and service processes using various bounds. Since its introduction in the early 1990s, network calculus has developed along two tracks—deterministic and stochastic. This book is devoted to summarizing results for stochastic network calculus that can be employed in the design of computer networks to provide stochastic service guarantees.

Overview and Goal

Like conventional queuing theory, stochastic network calculus is based on properly defined traffic models and service models. However, while in conventional queuing theory an arrival process is typically characterized by the inter-arrival times of customers and a service process by the service times of customers, the arrival process and the service process are modeled in network calculus respectively by some arrival curve that (maybe probabilistically) upper-bounds the cumulative arrival and by some service curve that (maybe probabilistically) lower-bounds the cumulative service. The idea of using bounds to characterize traffic and service was initially introduced for deterministic network calculus. It has also been extended to stochastic network calculus by exploiting the stochastic nature of arrival and service processes. While stochastic network calculus can be considered as generalized from deterministic network calculus, this generalization is not straightforward. Deterministic network calculus is based on a worst-case analysis that usually does not need to worry about the stochastic behavior of traffic and service. On the other hand, stochastic network calculus must take into consideration in the analysis the stochastic characteristics of traffic and service processes to better make use of their statistical multiplexing gains. For example, while independent case analysis is not considered in deterministic network calculus, it is critical in stochastic network calculus since it can usually provide much better results.

The main goal of this book is to summarize results for stochastic service guarantee analysis using the theory of stochastic network calculus. Since many networks (such as wireless networks) provide only stochastic service guarantees, and many applications (such as multimedia applications) perform well with stochastic service guarantees, the results in this book will be useful for analysis and provision of service guarantees in such network scenarios.

Structure

This book is structured as follows. Chapter 1 gives an introduction to the basic properties required from a theory for tractable performance analysis of computer networks. Also in this chapter, the background knowledge useful for helping understand the rest of the book is presented. Chapter 2 reviews fundamental concepts and results of deterministic network calculus. They include the arrival curve traffic model, the service curve server model, and the basic properties supported by deterministic network calculus. From Chapter 3 to Chapter 9, we focus on introducing important concepts and results in the context of stochastic network calculus. Specifically, Chapter 3 introduces various stochastic traffic models and their relationships with each other as well as with some well-known traffic models such as the effective bandwidth model. Chapter 4 defines stochastic server models for stochastic network calculus and introduces their relationships with each other. Chapter 5 summarizes the basic properties of stochastic network calculus under different combinations of traffic and server models introduced in Chapters 3 and 4, that are obtained without considering the possible independence of arrival and service processes. Chapter 6 focuses on independent case analysis, under which the basic properties of stochastic network calculus under different combinations of traffic and server models are presented. In Chapter 7, the analysis is on stochastic service guarantees under different scheduling disciplines. Also in this chapter, an application using the analysis results in admission control is presented. In Chapter 8, stochastic network calculus is extended and applied to study to the extent to which a flow becomes non-conformant with respect to its initial traffic characterization after it passes through a network. Finally, in Chapter 9, the theory of stochastic network calculus is applied to study generalized processor sharing systems with long-range dependent traffic inputs. In order to provide a more complete picture of the field, an appendix is presented that summarizes the book and discusses open research challenges in the area of stochastic network calculus.

Use of the Book

This book presents an overall picture of the state of the art of stochastic service guarantee analysis and a comprehensive treatment of this active research area. The content of the book can be divided into two parts. The first part, consisting of Chapters 1 to 6, provides the main set of results. The second part, Chapters 7 to 9, extends the first part with more results and example applications of stochastic network calculus. In the appendix, a comprehensive discussion of open research challenges in the area of stochastic network calculus is provided. The book can be used in advanced undergraduate or graduate

courses on performance evaluation of computer networks. Such a course should cover the full first part of the book, which provides a first course on stochastic network calculus. The second part can be used flexibly, where each chapter is self-contained. In addition, researchers in the area of performance evaluation of computer networks and provision and analysis of service guarantees in computer networks can also benefit from the comprehensive discussion of stochastic service guarantee issues in this book.

Notes and Acknowledgments

Part of content of this book has been lectured by the first author in two graduate courses related to traffic analysis of communication networks at the Norwegian University of Science and Technology since 2006. Many results in the book are based on works authored or co-authored by the two authors. We would like to thank other people who have contributed to these works. They include Dr. Peder J. Emstad, Dr. Markus Fidler, Dr. Shengming Jiang, Ms. Anne Nevin, Dr. Victor F. Nicola, Dr. Chunming Qiao, Dr. Chen-Khong Tham, Dr. Ian L.-J. Thng, Dr. Qinghe Yin, and Dr. Xiang Yu. In addition, we would like to thank the publisher and Ms. Catherine Brett and Mr. Wayne Wheeler there for their patience, support, and help.

Trondheim, April 2008 *Yuming Jiang*
Singapore, April 2008 *Yong Liu*

Contents

Preface vii

List of Figures xvii

List of Tables xix

1 Introduction 1
- 1.1 Quality of Service Guarantees 2
- 1.2 Basic Properties for Network Analysis 3
- 1.3 Notation and Mathematical Background 6
  - 1.3.1 Notation 6
  - 1.3.2 Min-Plus Algebra Basics 8
  - 1.3.3 Maximum Horizontal Distance and Maximum Vertical Distance 10
- 1.4 Random Variable and Stochastic Process Basics 10
  - 1.4.1 Random Variables 10
  - 1.4.2 Stochastic Processes 13
  - 1.4.3 Stochastic Ordering 13
- 1.5 Min-Plus Linearity of Queuing Systems 15
- 1.6 Summary and Bibliographic Comments 16
- Problems 17

2 Deterministic Network Calculus 19
- 2.1 Traffic Models 19
  - 2.1.1 (σ, ρ) Traffic Characterization 19
  - 2.1.2 Arrival Curve 20
  - 2.1.3 Envelope Process 22
- 2.2 Server Models 23
  - 2.2.1 Service Curve 23
  - 2.2.2 Network Elements Offering Service Curves 24
  - 2.2.3 Strict Service Curve 26

2.2.4 Service Envelope Process 26
2.3 Basic Results 27
2.3.1 Service Guarantees 28
2.3.2 Output Characterization 30
2.3.3 Concatenation Property 31
2.3.4 Leftover Service 33
2.3.5 Superposition 34
2.3.6 Example: Analysis of a Network of Arbitrary Topology 34
2.4 Summary and Bibliographic Comments 36
Problems 37

3 Traffic Models for Stochastic Network Calculus 41
3.1 $(\sigma(\theta), \rho(\theta))$ Traffic Characterization 41
3.2 t.a.c. Stochastic Arrival Curve 42
3.2.1 Difficulties and Additional Constraints 45
3.3 v.b.c. Stochastic Arrival Curve 48
3.4 m.b.c. Stochastic Arrival Curve 53
3.4.1 θ-m.b.c. Stochastic Arrival Curve 56
3.5 Stochastic Envelope Process 57
3.6 Summary and Bibliographic Comments 60
Problems 63

4 Server Models for Stochastic Network Calculus 65
4.1 Weak Stochastic Service Curve 65
4.2 Stochastic Service Curve 66
4.2.1 θ-Stochastic Service Curve 68
4.3 Stochastic Strict Service Curve 69
4.3.1 Stochastic Strict Server due to Impairment 71
4.4 Service Envelope Process 74
4.4.1 Latency Rate Characterization 76
4.4.2 Relationship with $(\sigma(\theta), \rho(\theta))$ Traffic Characterization . 77
4.5 Summary and Bibliographic Comments 78
Problems 79

5 Basic Properties of Stochastic Network Calculus 83
5.1 Service Guarantees 83
5.1.1 Backlog Bound 83
5.1.2 Delay Bound 85
5.2 Output Characterization 88
5.2.1 Output t.a.c Stochastic Arrival Curve 88
5.2.2 Output v.b.c. Stochastic Arrival Curve 90
5.2.3 Output m.b.c Stochastic Arrival Curve 93
5.2.4 Output θ-m.b.c. Stochastic Arrival Curve 96

5.3 Concatenation Property 97
5.4 Leftover Service Characterization 101
5.4.1 Leftover Weak Stochastic Service Curve 101
5.4.2 Leftover Stochastic Service Curve 102
5.4.3 Leftover θ-Stochastic Service Curve 105
5.5 Superposition Property 107
5.6 Scaling of End-to-End Delay Bound 109
5.6.1 Delay Bound From the Concatenation Property 110
5.6.2 Delay Bound from Node-by-Node Analysis 111
5.7 Calculus on Traffic and Service Envelope Processes 112
5.8 Summary and Bibliographic Comments 114
Problems 116

6 Independent Case Analysis 119
6.1 Introduction 119
6.2 Analysis Based on Stochastic Strict Server 121
6.2.1 Backlog and Delay Bounds 122
6.2.2 Output Characterization 124
6.2.3 Concatenation Property 128
6.2.4 Leftover Service Characterization 130
6.2.5 Superposition Property 133
6.2.6 Scaling of End-to-End Delay Bound 134
6.3 Calculus with Moment Generating Functions 136
6.3.1 Moment Generating Function Basics 136
6.3.2 Basic Properties and Performance Bounds 138
6.4 Summary and Bibliographic Comments 141
Problems 142

7 Analysis on Scheduling Disciplines 145
7.1 Introduction to gSBB 145
7.2 Properties of gSBB 148
7.3 Analysis on Different Scheduling Disciplines 150
7.3.1 General Results 151
7.3.2 First In First Out (FIFO) 152
7.3.3 Strict Priority (SP) 153
7.3.4 Generalized Processor Sharing (GPS) 154
7.4 Application to Measurement-Based Admission Control 155
7.4.1 Introduction to MBAC 156
7.4.2 Introduction to Flow-Aware Networking 157
7.4.3 Dynamic Priority Scheduling-Based MBAC 158
7.4.4 Analysis 160
7.5 Summary and Bibliographic Comments 163
Problems 164

8 Traffic Conformance Study 165
8.1 Network Model 166
8.1.1 Conformance Deterioration and Stochastic Burstiness Increase 167
8.1.2 Property of Token Bucket Shaper 168
8.2 Conformance Study of Per-Flow Scheduling Network 169
8.2.1 Single-Node Case 169
8.2.2 Multi-node Case 170
8.3 Conformance Study of Aggregate Scheduling Network 171
8.3.1 Per-Flow in Single-Node Case 171
8.3.2 Per-Flow in Multi-node Case 176
8.3.3 Per-Aggregate Case 178
8.4 Simulation Results 179
8.4.1 Per-Flow Scheduling Network in Single-Node Case 179
8.4.2 Aggregate Scheduling Network in Single-Node Case 182
8.4.3 Aggregate Scheduling Network in Multi-node Case 183
8.5 Summary and Bibliographic Comments 185
Problems 187

9 LRD Isolation in Generalized Processor Sharing 189
9.1 Introduction 190
9.1.1 GPS Fundamentals 190
9.1.2 LRD Traffic Characterization 191
9.1.3 LRD Isolation of Flows 192
9.2 Analysis of LRD Traffic 193
9.2.1 Single Arrival Process 193
9.2.2 Aggregate Process 196
9.3 Sample Path Behavior of LRD Traffic in a GPS System 199
9.3.1 GPS Decomposition 199
9.3.2 A General Bound 200
9.3.3 An Alternate Bound 201
9.4 Technique to Check and Ensure LRD Isolation 202
9.4.1 Limitations of Existing Methods 203
9.4.2 Necessary and Sufficient Condition 205
9.4.3 Weight Adjustment and Assignment 206
9.5 Sample Path Behavior of LRD Traffic in a PGPS System 208
9.6 Summary and Bibliographic Comments 211
Problems 212

A Open Research Challenges 213
A.1 Applicability Condition Study of Stochastic Network Calculus 214
A.2 Advanced Properties 215
A.3 Network Information Theory for the Internet 217

References 219

Index 227

List of Figures

1.1 Analysis of a simple network 4
1.2 Maximum horizontal and vertical distances between two functions 10

2.1 Token bucket regulator 20
2.2 Dual-token bucket constrained traffic 22
2.3 Comparison of node-by-node analysis and edge-to-edge analysis 33
2.4 Two servers in tandem 38

3.1 Conversions between stochastic arrival curve models 62
3.2 Relationships between stochastic arrival curve models 62

4.1 Conversions between stochastic service curves 80
4.2 Relationships between stochastic service curves 80

5.1 Concatenation property of stochastic service curve 97
5.2 Stochastic servers in tandem 110

6.1 Comparison of Lemmas 6.1 and 1.5 120

7.1 Comparison of bounding functions for DBB, SBB, and gSBB . . 147
7.2 MBAC structure 157

8.1 Network model 166
8.2 Aggregate scheduling in multi-node case 177
8.3 Network topology used in simulation 180
8.4 Queue length tail distribution after crossing a single node in a per-flow scheduling network 180
8.5 Non-conformance probability after crossing a single node in a per-flow scheduling network 181
8.6 Queue length tail distribution after crossing a single node in an aggregate scheduling network 182

8.7 Non-conformance probability after crossing a single node in an aggregate scheduling network 183
8.8 Queue length tail distribution after crossing multi-nodes in an aggregate scheduling network 184
8.9 Non-conformance probability after crossing multi-nodes in an aggregate scheduling network 184

9.1 Decomposing a GPS system into N fictitious SSQs 200
9.2 PGPS server 209

List of Tables

2.1 Latency and rate terms of some schedulers 25

5.1 Properties provided by a combination of traffic model and server model 114

9.1 Preference for Lemma 9.7 or Lemma 9.9 in different scenarios 198