QoS & Traffic Management in IP & ATM Networks

- Today's First Comprehensive Guide to QoS and Traffic Management
- Shows How to Integrate Voice and Data on an IP or ATM Network
- Offers Real-Life Examples of How to Relieve and Avoid Network Traffic Jams
- Written by a Pioneer and Leader in ATM Traffic Management

DAVID McDYSAN

QoS & Traffic Management in IP & ATM Networks

David McDysan

McGraw-Hill
New York San Francisco Washington, D.C.
Auckland Bogotá Caracas Lisbon London Madrid
Mexico City Milan Montreal New Delhi San Juan
Singapore Sydney Tokyo Toronto

McGraw-Hill

A Division of The McGraw-Hill Companies

1 2 3 4 5 6 7 8 9 0 AGM/AGM 9 0 4 3 2 1 0 9

ISBN 0-07-134959-6

The sponsoring editor for this book was Steven Elliot, the editing supervisor was Penny Linskey, and the production supervisor was Claire Stanley. It was set Vendome ICG and Eras by the author using Microsoft Word, Powerpoint, Excel, and Lotus Freelance.

Printed and bound by Quebecor/Martinsburg.

To my wife, Debbie and my parents,
Lowell and Martha

PREFACE

Quality of Service (QoS) and traffic management is becoming an increasingly important subject in the broadband Internet-driven information age. The pioneering work in the development of QoS-aware Asynchronous Transfer Mode (ATM) switches and protocols is the foundation upon which similar capabilities are being developed for an Internet Protocol (IP) based set of applications. Furthermore, the Internet is exploring new approaches to QoS and traffic management to meet the challenges of unprecedented growth and global scale.

Unfortunately, many of the classic works and theoretical results of traffic engineering, queuing theory and other important disciplines remain inaccessible to the web surfer, corporate information worker, network designer, network operator, and communications manager. This book aims to fill the void between high-level, simplistic treatments of the subject, graduate-level texts, and the specialist technical literature. Accordingly, a central theme of the book is old-fashioned common sense.

Traditionally, network managers must enlist a small community of expert traffic engineers and network designers to solve these problems. Although complex problems often require complex solutions, these experts could often do better in explaining their work to others. Toward this end, this book uses real-life analogies to illustrate important concepts in a sincere effort to appeal to a larger number of readers than traditional traffic engineering and queuing theory texts do. The objective of this book is to empower a larger population of network designers, network managers, and end users with an intuitive sense of QoS and traffic management to complement some basic analytical tools.

We are all familiar with the issues involved in QoS and traffic management from our everyday experiences with travel on the highways, trains, and airlines. Different levels of quality are an important way to address shortage of a particular resource or prioritize service, since some customers are willing to pay more in order to receive better service. For example, first class passengers receive preferential treatment and guaranteed reservations, albeit at a higher price. Thus, quality of service has strong policy and economic associations.

Traffic management deals with traffic levels for which a certain QoS applies. One way of avoiding congestion, and the resulting impact on quality, is to apply admission control. For example, there are only a certain number of seats on an airplane or train, or only a maximum number of vehicles can use a highway over a specific time interval. Thus, admitting only those passengers with tickets is an important means of providing the level of quality via a simple traffic control. In the vehicular analogy, signals on the on-ramps to freeways may modulate admitted traffic to optimize traffic flow. As we shall

see, direct analogies occur in IP and ATM networks to QoS and traffic management.

This book uses mathematics to quantify QoS parameters and traffic levels. Consistent with the objective of reaching the largest possible audience, we formulate the solution for many problems in spreadsheet format using the Microsoft Excel program. The text identifies trademarked names, like Excel, using initial caps. I will put selected spreadsheets on my home page as pointed to in the author pages on the McGraw-Hill web site, www.mcgraw-hill.com/computing/authors.

Who Should Read This Book

The primary objective of this book is to introduce the theory and practical use of traffic engineering to a broad audience. Beginning users can utilize it as an introduction to traffic engineering and a guide to explore specific subject areas in greater depth. Intermediate users can employ it as an authoritative reference. It also targets the expert who needs to communicate traffic engineering concepts to executives, customers, or colleagues. The book should serve as a reference guide that will assist the engineer or manager in understanding the basic concepts as well as applying the theory to real networking problems. The book targets the hot topics of integrated voice, video and data networking, QoS, and traffic management for IP and ATM networking technologies. Each chapter provides a bibliography for the reader who wishes to further explore certain topics.

How This Book Is Organized

The outline of the book progresses in a logical sequence of parts, stepping the reader through the basic concepts of networking, QOS, traffic, probability theory, queuing theory, congestion control, routing, network design, decision making, and simulation. Each part begins with an introductory chapter that draws analogies with real-life that illustrate important points. The introductory level chapters target a non-technical professional. A beginner could read the first chapter of each part and end up with knowledge of traffic engineering sufficient to understand the standards published on the subject or converse with an expert. These chapters contain only the simplest equations, if they contain any at all. Each successive chapter within

each part is more advanced and uses mathematics appropriate to the task.

The second, intermediate level chapter presents formulas in a cookbook style that is easily put to use by the practicing engineer or manager using commonly available computer tools like spreadsheets and macro languages. A technical background is desirable, but not necessary for the intermediate chapters. The book does provide references to other texts and articles so that a reader can obtain the necessary background.

The third and most advanced chapter of each part targets the advanced undergraduate student or the introductory graduate student in the disciplines of mathematics, communications engineering, or computer science. The reader of the advanced chapter in each part should have some background in the particular topic. It provides selected proofs and references for advanced study. A professional could consult these references and extend their capabilities through independent study. The references cite a number of graduate level texts already published in this area.

Each part and chapter begins with a quotation to break up an otherwise dry topic, as well as stimulate the reader to look at traffic engineering problems from a broader perspective, hence honing their intuition. The book is organized into seven Parts, each containing three Chapters. The remainder of this section summarizes the contents of each Part and its constituent Chapters.

PART 1 — Traffic Management in Communication Networks

Part 1 provides some basic background information and introduces the subjects of QoS and traffic management in IP and ATM networks. It describes the basic paradigms and operation of the IP and ATM protocols, emphasizing the traffic management aspects. A reader already familiar with these subjects could skip this part after reviewing the definitions of capacity and traffic parameters in Chapter 3.

1. *Introduction and Overview.* Chapter 1 begins with a brief history of circuit and packet switching as it relates to IP and ATM networking. It includes a high-level introduction and roadmap to the detailed concepts of OoS and traffic management covered in the remainder of the book. It also contains information on how to get the IP and ATM standards referenced in the book.
2. *Review of Circuit- and Packet-Switched Networks*. This chapter provides more in depth background information on circuit switch-

ing, connectionless packet switching used via IP, and connection-oriented packet switching used via ATM and MultiProtocol Label Switching (MPLS).

3. *Capacity, Throughput, and Service.* Understanding terminology using precise definitions is critical to the study of traffic engineering. Therefore, this chapter defines transmission bandwidth and its relationship to channel capacity. Additionally, it defines the concepts of a bottleneck, throughput, and efficiency. We also define source traffic characteristics like the peak and average rate in general terms. This chapter also precisely defines the means used in IP and ATM networks for measuring and monitoring traffic parameters.

PART 2 — Quality of Service (QoS) and Traffic Control

Part 2 focuses on the subject of QoS and introduces the techniques that IP and ATM networks employ to deliver the required quality for a specific level of traffic. It takes the generic notion of quality and quantifies it in terms of specific measures. Taken together, the first two parts define the basic concepts of traffic and quality applied to specific IP and ATM protocols. This provides the framework used in later parts of the book for computing the QoS delivered via specific systems and network designs in response to a well-defined level of traffic.

4. *Perception is Reality.* Since human perception is a principal determinant for the required level of quality, this chapter begins by reviewing how our senses react to voice and video communication signals. The coverage includes a description of the terminology employed in IP and ATM networks to request a specific level of QoS.
5. *Quality of Service (QoS) Defined.* This chapter precisely defines QoS in terms of parameters like loss, delay, availability, and variation in delay specified in IP and ATM standards. It also includes a description of how an end user can request service in terms of a traffic level and an associated QoS.
6. *Delivering QoS via Traffic Control.* IP routers and ATM switches utilize a range of implementations to deliver differentiated QoS and traffic control. This chapter summarizes these techniques, which include admission control, policing, and shaping. The subject of queuing and scheduling is briefly introduced, but covered in depth in Chapter 10.

PART 3 — The Traffic Phenomenon

Since most offered traffic has at least some random component, Part 3 provides background on the mathematics from probability theory used to compute the quality delivered by a particular design. A reader looking to get an overview could read only Chapter 7. Readers seeking a more in depth understanding of probability theory or statistics should review Chapters 8 and 9.

7. *Randomness in Our Everyday Lives.* Chapter 7 introduces the reader to the concept of randomness through some simple puzzles. It then introduces the concept of stochastic processes commonly used to model traffic phenomena.
8. *Random Traffic Models.* This chapter provides an overview of important results from probability theory used in the analysis of traffic systems. The text summarizes important results and distributions. It identifies the spreadsheet formulas and functions available to compute numerical results. It concludes with an introduction to stochastic processes.
9. *Advanced Traffic Models.* Traditionally, traffic analysis used the relatively simple Markov model summarized in this chapter. This chapter also defines more recently developed methods called self-similar processes that better model the observed characteristics for some traffic types encountered in IP and ATM networks.

PART 4 — Queuing Principles

This part builds upon the foundation of probability theory and random processes established in Part 3 to provide a basic tool kit for analyzing the performance of IP and ATM networks. Called queuing theory, these results describe the statistics of important QoS measures like delay, loss, and availability.

10. *Queuing — A Fancy Name for Waiting in Line.* This chapter defines the basics of all traffic handling systems: arrivals, waiting room, and service discipline. It also describes the role of prioritized and thresholded queues serviced by a scheduling policy employed by routers and switches to deliver differentiated QoS.
11. *Basic Queuing Theory Applied.* A significant body of knowledge exists regarding the performance of queuing systems driven by Markovian traffic. This chapter summarizes the important results as applied to the performance of connection-oriented networks,

buffer design in routers and switches, as well as voice and data integration.

12. *Intermediate Queuing Theory Applied.* The mathematics of queuing system analysis becomes quite complex for non-Markovian traffic models. Chapter 12 surveys this landscape and provides some useful approximations for estimating system performance.

PART 5 — Congestion Detection and Control

Now, after understanding the statistically predictable effect of random traffic on performance, this part addresses the unpredictable phenomenon of congestion. The discussion draws analogies with everyday life and frames the problem structure in terms of the two basic forms of response to congestion: avoidance and reaction. The chapters in the part use this basic structure to analyze the performance of open and closed loop congestion control schemes.

13. *Traffic Jam Up Ahead!* The congestion phenomena we all encounter in our everyday lives are similar in many ways to that experienced in networks. We introduce these concepts, and define the tradeoff between throughput and delay involved in the design of networks. This chapter also describes the effect of congestion on retransmission protocols used in data communication networks.
14. *Open-Loop Congestion Control.* The technique here is to avoid congestion via careful planning and forethought. This includes selective marking of traffic according to conformance with the traffic parameters associated with the flow or connection. The analysis includes performance of Weighted Fair Queuing (WFQ) on a nodal basis and the effective throughput achieved via retransmission protocols with a fixed window size.
15. *Closed-Loop Congestion Control.* When planning isn't possible, or unexpected traffic levels occur, a reactive approach is all that remains. Here, the discipline of control theory tells us much about the basic properties of closed loop congestion control schemes. The chapter applies these techniques to rate- and window-based flow control systems, including a detailed model of TCP/IP performance.

PART 6 — Routing and Network Design

In real networks, traffic and congestion control occur in a complex pattern of interconnected devices serving traffic flows from many

users. This part takes the results of the preceding chapters and applies them in a network context. Since extending the result of a single node to an interrelated set of nodes is a complicated problem, we analyze some simple network topologies to expose some general aspects of network design.

16. *Routing Background and Concepts.* This chapter introduces the subject via analogies to transportation systems and some simple network examples. It then describes the link-state routing algorithm and the concept of constrained routing. It concludes with a discussion on the tradeoffs involved in routing algorithm complexity and the efficient utilization of transmission links and switching resources.
17. *Routing Algorithms.* The mathematical description of routing draws upon the disciplines of graph theory, computer science, and operations research. This chapter introduces the terminology from these areas and applies them to generic network design problems.
18. *Performance of Network Routing Designs.* Analyzing the performance of a network of devices is a challenging problem. This chapter summarizes the networks of queues model for computing end-to-end performance for Markovian traffic. It also describes some analytical routing models for symmetric mesh and tree networks.

PART 7 — Putting It All Together

This final part introduces some additional techniques utilized in network design. It also highlights some important practical considerations involved in the design, deployment, and operation of an IP or ATM network. Finally, it concludes with some thoughts and possible future directions of QoS and traffic management in IP and ATM networks.

19. *Traffic Engineering Applied.* Chapter 19 describes some additional networking techniques involved in the design of large networks. It then provides an overview of the network planning and design process, highlighting the role of network design tools.
20. *Designing Real World Networks.* This chapter includes some additional material on network design. This includes the performance of hierarchical network design and the effect of overflow traffic. It also summarizes dynamic routing, and least cost design.
21. *Where to Go From Here.* Here, we introduce the use of event-based simulation tools to model complicated network configurations. This chapter also provides some practical guidelines for formulat-

ing traffic models. It concludes with a brief discussion on the future of QoS and traffic management.

Acknowledgements

A book like this requires a great deal of help and support. First, I would like to acknowledge Steve Elliot of McGraw-Hill who refined the proposal for this book and supported the effort. Secondly, I would like to thank my wife, Debbie, for giving me the time to write this book along with her continual support and encouragement. Finally, I would like to acknowledge the reviewers who carefully read the manuscript, corrected errors, suggested clarifications, and provided additional references. These reviewers were: Professor Thomas Chen of Southern Methodist University, Fatih Alagoz of George Washington University, Roland Smith of Nortel Networks, Byoung-Joon Lee of Cisco Systems, Furrukh Fahim of Fahim Associates, Dr. Cheng Chen of NEC, and Dr. Bharathi Devi, James Liou, and Syeda Sanjana of MCI WorldCom.

This book does not reflect any policy or position of MCI WorldCom. The ideas and concepts expressed herein are those of the author or the cited references.

CONTENTS