

## Preface

This book is written for computer system performance analysts. Its goal is to teach them to apply queueing network models in their work, as tools to assist in answering the questions of cost and performance that arise throughout the life of a computer system.

Our approach to the subject arises from our collective experience in contributing to the theory of queueing network modelling, in embodying this theory in performance analysis tools, in applying these tools in the field, and in teaching computer system analysis using queueing network models in academic and industrial settings. Some important beliefs underlying our approach are:

- Although queueing network models are not a panacea, they are the appropriate tool in a wide variety of computer system design and analysis applications.
- The single most important attribute of a computer system analyst is a thorough understanding of computer systems. We assume this of our readers.
- On the one hand, mathematical sophistication is not required to analyze computer systems intelligently and successfully using queueing network models. This is the case because the algorithms for evaluating queueing network models are well developed.
- On the other hand, the purchase of a queueing network modelling software package does not assure success in computer system analysis. This is the case because defining and parameterizing a queueing network model of a particular computer system is a blend of art and science, requiring training and experience.

Queueing network modelling is a methodology for the analysis of computer systems. A methodology is a way of thinking, not a substitute for thinking.

We have divided the book into six parts. In Part I we provide four types of background material: a general discussion of queueing network modelling, an overview of the way in which a modelling study is conducted, an introduction to the interesting performance quantities in computer systems and to certain relationships that must hold among them, and a discussion of the inputs and outputs of queueing network models.

In Part II we present the techniques that are used to evaluate queueing network models — to obtain outputs such as utilizations, residence times, queue lengths, and throughputs from inputs such as workload intensities and service demands.

In Part III we explore the need for detailed models of specific subsystems, and the construction of such models for memory, disk I/O, and processor subsystems.

In Part IV we study the parameterization of queueing network models of existing systems, evolving systems, and proposed systems.

In Part V we survey some non-traditional applications, such as the analysis of computer communication networks and database concurrency control mechanisms. We also examine the structure and use of queueing network modelling software packages.

In Part VI, the appendices, we provide a case study in obtaining queueing network parameter values from system measurement data, and programs implementing the queueing network evaluation techniques described in Part II.

Case studies appear throughout the book. They are included to illustrate various aspects of computer system analysis using queueing network models. They should *not* be misconstrued as making general statements about the relative performance of various systems; the results have significance only for the specific configurations and workloads under consideration.

We have summarized a number of important modelling techniques in the form of “Algorithms”. Our intention is to provide enough information that the reader can understand fully the essential aspects of each technique. We omit details of significance to the implementation of a technique when we feel that these details might obscure the more fundamental concepts.

It is our experience that practicing computer system analysts are relatively skilled in techniques such as workload characterization, system measurement, interpretation of performance data, and system tuning, and are at least acquainted with basic statistical methods and with simulation. Each of these subjects is well represented in the existing literature, and is given short shrift in the present book. Much interesting and important research work concerning queueing network modelling also is given short shrift; we discuss the one approach to each problem that we feel is best suited for application. For readers who desire to pursue a topic in greater detail than we have provided, each chapter concludes with a brief discussion of the relevant literature.

We owe a significant debt to Jeffrey P. Buzen and Peter J. Denning, who have been instrumental in the development of a pragmatic

philosophy of computer system analysis using queueing network models. Their influence is evident especially in our use of the *operational* framework for queueing network modelling, which conveys much greater intuition than the more traditional *stochastic* framework.

Jeffrey A. Brumfield, Jeffrey P. Buzen, Domenico Ferrari, Lenny Freilich, and Roger D. Stoesz have assisted us by reviewing our manuscript, as have several anonymous reviewers. Our work in computer system analysis using queueing network models has been supported in part by the National Science Foundation and by the Natural Sciences and Engineering Research Council of Canada. We thank our colleagues at the University of Washington and at the University of Toronto for their encouragement, and our families and friends for their forbearance.

Edward D. Lazowska, John Zahorjan,  
G. Scott Graham, and Kenneth C. Sevcik

Seattle and Toronto