Preface

Queueing is an aspect of modern life that we encounter at every step in our daily activities. Whether it happens at the checkout counter in the supermarket or in accessing the Internet, the basic phenomenon of queueing arises whenever a shared facility needs to be accessed for service by a large number of jobs or customers. The study of queueing is important as it provides both a theoritical background to the kind of service that we may expect from such a facility and the way in which the facility itself may be designed to provide some specified grade of service to its customers.

Our study of queueing was basically motivated by its use in the study of communication systems and computer networks. The various computers, routers and switches in such a network may be modelled as individual queues. The whole system may itself be modelled as a queueing network providing the required service to the messages, packets or cells that need to be carried. Application of queueing theory provides the theoritical framework for the design and study of such networks. The purpose of this book is to support a course on queueing systems at the senior undergraduate or graduate levels. Such a course would then provide the theoritical background on which a subsequent course on the performance modelling and analysis of computer networks may be based. This is indeed the strategy adopted for teaching computer networks and their performance modelling and analysis in the Department of Electrical Engineering at I.I.T., Kanpur. The material of this book was originally provided as lecture notes to the students specialising in the area of telecommunications and networking. These students were required to go through a sequence of two courses, the first one on queueing and the second on computer networks. The first course on queueing and the associated lecture notes were expected to provide the theoritical background for the second course and its related topics.

The phenomenon of queueing also arises in operations research and industrial engineering as the facilities studied in these areas may also be modelled as either individual queues or queueing networks. This text will also be useful for senior students in this area. This book assumes that the student is familiar with the basics of probability theory and its applications. It also assumes that the students know and can apply results from the theory of transforms, especially Laplace Transforms and Z-Transforms. It is normally expected that students in Electrical Engineering or Computer Science would be familiar with this required theoritical background before they reach their senior undergraduate or graduate levels. Students in other areas may require additional mathematical training to acquire this knowledge. An appropriate mathematical course covering these topics may be a prerequisite for such students who want to use this book for the study of queueing systems.

This book is basically concerned with the analysis of queueing systems. Some example scenarios are usually given for the systems being studied even though studies of potential queueing applications are not the primary objectives of this text. The analytical models have been carefully developed and presented. However, given the expected mathmatical background of the students using this text, obvious steps in the analysis have usually been omitted and left as an exercise for the readers. If this book is used for students studying queueing at a level lower than what was suggested earlier, the analytical steps might be omitted to focus more on the results which are finally obtained. For such students, it may be useful to just cover the basic queueing models of Chapter 2 and the basics of queueing networks as given in Chapter 5. For students or practicing engineers interested only in applying exact or approximate queueing network models for solving their application problems, the algorithms provided in Chapters 5 and 6 are recommended.

Chapter 1 provides an introduction to queues and queueing systems. To illustrate the simplicity of basic queueing analysis, this chapter also presents the analysis of a simple single server queue. This has been done with the aid of some simplifying assumptions and is intended to show that a basic queueing analysis may indeed be very easily done.

Chapter 2 presents the equilibrium solutions of basic queues that may be analysed using birth-death models. These queues are mostly ones with Poisson arrival processes and exponential service time distributions. We also discusses Little's result, which is used extensively in this book and elsewhere for the study of queues and queueing networks. Queues with bulk arrivals and using the method of stages to approximately solve queues with non-exponential service time distributions have also been considered in this chapter.

Chapters 3 and 4 deal with the single server queue with Poisson arrivals and general service time distributions. While Chapter 3 focusses on the basic queue of this type and presents an extensive analysis of its important performance parameters, Chapter 4 discusses some important variations of this queue. These variations include queues with vacations of different types and various priority queueing models. It also presents the analysis of some basic discrete-time queues. (All the other queueing models considered in this book are continuous-time in nature.)

Chapters 5 and 6 of the book deal with queueing networks. Chapter 5 considers simple open and closed network models that have a product-form solution and can be analysed exactly, provided suitable simplifying assumptions are satisfied. Jackson's Theorem is introduced here for a variety of queueing networks. We also present the convolution algorithm and the mean value algorithm for the exact solution of simple closed queueing networks. Norton's theorem for reducing closed queueing networks has also been discussed in this chapter.

Chapter 6 focusses on various algorithms that have been proposed as approximate methods of analysis for more complex queueing networks. A variety of such algorithms have been presented for both open and closed networks. Fork/Join nodes and their analysis in open and closed queueing network have also been considered. For open networks, we present the GI/G/m approximation of QNA that gives exceptionally good results in most situations. Networks where some or all the queues are of finite capacity have also been considered. The different blocking models that may be used have been presented and discussed in detail in this chapter. Even though these networks with blocking cannot be solved in an exact fashion (except with very restrictive assumptions) the Maximum Entropy Method does provide a reasonably good approximate solution. This has been presented for both open and closed networks with repetitive service blocking of different types. We have also presented approximate algorithms to handle open and closed networks with transfer blocking and open networks with rejection blocking.

Exact analysis of queueing systems is often difficult. Even with simplifying assumptions and approximations, it may not always be possible to obtain an analytical solution. In such cases, one has to take recourse to simulations for the study of these systems. Chapter 7 discusses the simulation approach and provides some of the basic knowledge necessary to meaningfully study queueing systems using simulations.

This book may be used as a textbook for a course on basic queueing theory by limiting oneself to the study of single queues of various types using Chapters 2, 3 and 4. The material on queueing networks in Chapters 5 and 6 may then be omitted from such a course. Chapter 6 gives summary descriptions of various approximation techniques that may be used to solve complex queueing networks. These would also be useful for practicing engineers who need to solve more complex queueing based network models. We have implemented the algorithms described in Chapters 5 and 6 in a queueing analysis package *QNAT*. A beta version of this software is being publicly distributed and may be downloaded from the Internet. The download details for this package may be obtained from the author. It should be noted that *QNAT* allows both analysis and simulation of all the queueing network models described in Chapters 5 and 6. For simulations, *QNAT* uses the discrete event simulation approach described in Chapter 7.

Sanjay Kumar Bose