Preface

In the early 1970s something of a revolution was taking place in the world of signal processing. This was mainly due to the introduction of digital signal processing devices which were rapidly replacing the old analogue ones. In parallel with this came the introduction of new, discrete-time techniques which could be used for the analysis and design of the new digital systems. The use of the z-transform and the fast Fourier transform (FFT) algorithms and techniques based on these became widespread.

A similar revolution was also taking place in the early 1970s in the computer and telecommunications industries, with the introduction of microcomputers, and computer networks such as the Aloha system and the ARPA network. However, this is where the parallel with signal processing ends. Despite the digital nature of these computer systems and networks, the techniques used for their analysis and design were firmly embedded in queuing theory, a topic that was mainly exclusive to the continuous-time domain. This is a situation that has persisted almost to the present day, and it is only now, in the 1990s, that the concept of modelling computer or telecommunication systems using discrete-time queuing theory is beginning to generate considerable interest. It is in anticipation of this interest escalating that this book has been written.

The reluctance to use discrete-time concepts is mainly due to the generally accepted view that discrete-time queueing systems can be more complex to analyse than equivalent continuous-time ones. In a continuous-time model, only a single state change can occur at any given time instant. In a discrete-time model, because of the finite size of a time-unit, multiple state changes can occur from one time-unit to the next. This complicates the resulting analysis of the model.

However, many digital systems such as communication or computer networks operate on the basis of time-slotting. Such systems inherently lead themselves to representation with discrete-time models by specifying a one-to-one correspondence between a time-unit in the model and a time-slot (or some convenient multiple thereof) in the physical system. In this way very accurate models, that faithfully reproduce the stochastic behaviour of a communication or computer network, can be constructed. One of the objectives of the present book is to present a unified approach to developing accurate models of communication or computer networks using discrete-time queueing theory.

Although the models so constructed cannot always be solved exactly, another objective of the book is to develop good approximation techniques to obtain solutions.

To motivate the above two objectives, I quote the words of Paul Schweitzer when taking part in a panel discussion on 'The State of the Art and Future Directions in Computer Modelling,' the text of which can be found in [IAZE 84]:

> 'In my view we have reached the end of the road for exact models, and future efforts should be devoted to developing better classes of approximation models ... it is better to have an approximate treatment of an accurate model than an accurate treatment of an inaccurate model'.

This book attempts to marry these two suggestions, and provide a unified method for developing accurate, discrete-time models of communication and computer networks, along with good approximation techniques for solving them.

The book has seven chapters and, following the Introduction (Chapter 1), can be divided into three parts which might be classed as: Basics (Chapters 2 and 3), Discrete-time queueing systems (Chapters 4 and 5), and Applications (Chapters 6 and 7).

Chapter 1 outlines the basic aims and philosophy of the book, and introduces discrete-time queues at the simplest possible level.

Chapters 2 and 3 cover probability theory and discrete-time Markov chains. It is expected that most readers already have a knowledge of these topics. The topics are, however, included for completeness, to establish a notation, and provide a reference source for subsequent chapters. The chapter on probability theory (Chapter 2) is fairly standard, with the possible exception of the material on the conditional binomial distribution, which later turns up in one of the models. Surprisingly, this material does not appear in the popular texts, such as those by Feller [FELL 71] or Papoulis [PAPO 84]. Chapter 3 on Markov chains is another standard treatment. There is however a strong emphasis here on how Markov chains relate to performance models for communication and computer networks. There is also an emphasis on the concept of time reversal.

Chapters 4 and 5 deal with discrete-time queues and discrete-time queueing networks, respectively. Although many of the results presented in Chapter 4 can be found elsewhere, these are spread over a number of diverse research papers that use widely different notations. I have derived the results from basics, mainly using nothing more complicated than discrete-time Markov chains, in order to give a consistent presentation throughout. Of particular importance here are the results on S-queues, which are based on a paper by Jean Walrand [WALR 83A].

Chapter 5 extends the results on discrete-time queues to discrete-time queueing networks. Again there is a strong influence from Walrand here. The classic product form distributions are derived for both open and closed networks of S-queues. A new development here is the discrete-time queueing network model for multiple-access protocols. In general, these models have state dependent routing probabilities and non-linear traffic equations, and approximation techniques must be used to solve them. An approximation technique known as equilibrium point analysis (EPA) is introduced. This method was motivated by the work of Shuji Tasaka [TASA 86]. The emphasis in this text however is on the interpretation of EPA in the context of discrete-time queueing networks. Important extensions of the method are derived that can be applied to networks with different customer classes.

Chapters 6 and 7 are concerned with applications to satellite networks and local networks, respectively. Most of the models here have been carefully chosen to illustrate a specific modelling concept or technique; for example, how to handle such things as buffered users, statistically different users, finite channel delays, timing deadlines (as in real-time), and unsolvable traffic equations are all covered. A number of the models presented here have not previously been published in the open literature.

The precedence relations among the chapters are strictly sequential, although, as previously noted, Chapters 2 and 3 can mostly be skipped by those familiar with the topics. Some exercises are included at the end of Chapters 3 through 7. These provide plausible extensions of the results contained in the text, although those exercises marked with an asterisk are considered to be fairly major undertakings.

The text should prove useful to both practitioners and researchers concerned with communication and computer network performance evaluation. It is hoped that it will also prove invaluable to 'get off the ground' prospective researchers who wish to investigate the relatively untapped area of discrete-time queueing systems and their application to the performance modelling of communication and computer networks.

Acknowledgement

I would like to thank colleagues and students in the Department of Electronic and Electrical Engineering at Loughborough University of Technology for many helpful suggestions, and also acknowledge the use of the facilities at the University. Special thanks are due to my friend and student Kamajith (Rohan) Rodrigo who expertly typed the manuscript, produced most of the diagrams, and corrected many of my errors along the way. Thanks are also due to Professors Jean Walrand and Isi Mitrani, both of whom helped to clarify points on which I was unsure. Finally, I thank my wife Christine, without whose support and understanding this book could not have been written.

Michael E. Woodward
Loughborough