# Modeling and Analysis of Computer Communications Networks

**Jeremiah F. Hayes**
*Concordia University*
*Montreal, Quebec, Canada*

PLENUM PRESS • NEW YORK AND LONDON

# Preface

In large measure the traditional concern of communications engineers has been the conveyance of voice signals. The most prominent example is the telephone network, in which the techniques used for transmission multiplexing and switching have been designed for voice signals. However, one of the many effects of computers has been the growing volume of the sort of traffic that flows in networks composed of user terminals, processors, and peripherals. The characteristics of this data traffic and the associated performance requirements are quite different from those of voice traffic. These differences, coupled with burgeoning digital technology, have engendered a whole new set of approaches to multiplexing and switching this traffic. The new techniques are the province of what has been loosely called computer communications networks.

The subject of this book is the mathematical modeling and analysis of computer communications networks, that is to say, the multiplexing and switching techniques that have been developed for data traffic. The basis for many of the models that we shall consider is queueing theory, although a number of other disciplines are drawn on as well. The level at which this material is covered is that of a first-year graduate course. It is assumed that at the outset the student has had a good undergraduate course in probability and random processes of the sort that are more and more common among electrical engineering and computer science departments. (For the purpose of review, but not first introduction, the required background material is given in a pair of appendices.) The material in the text is developed from this starting point. The objective is to develop in the student the ability to model and analyze computer communication networks. We also seek to impart a critical appreciation of the literature in the field.

In a book at this level, it is inevitable that the choice of the particular material to be included in the text is heavily influenced by the author's own research and professional experience. However, as the book evolved, the work of others assumed a more and more prominent role. In fact, one of the most rewarding aspects of writing the book was learning and appreciating the fine work that others have done, and getting a sense of how the field has progressed. We hope that we have conveyed some of this to the reader. Our appreciation of a good deal of the material outside our immediate interests was gained to a significant degree through a number of first-rate survey papers in the field. We shall gratefully acknowledge these papers at the appropriate points in the text.

The level of the course for which this book is intended places certain limitations on the material that could be covered. In several instances important work could not be included simply because students could not be expected to have the appropriate mathematical training. In other cases the analysis was simply too involved. In spite of these difficulties, we derive great satisfaction from the fact that we were able to include almost all of the work which we consider to be seminal in the field. We tried very hard to present this work in a form that would be palatable to the serious though inexperienced student. A number of exercises illustrating the material are included.

Since the focus is on modeling and analysis, somewhat less attention is given to the details of implementation and operation. However, we attempted to include enough of this material so as to place the mathematical models in the proper context. In the course of the discussion we direct the reader to a number of authoritative sources on implementation and operation.

There is more material in the text than can be covered in an ordinary graduate course. In order to assist the instructor in selecting material, we point out several natural groupings. Chapters 1, 2, and 3 serve as an introduction to the rest of the text. After completing Chapter 3 one could go on to Chapter 4 or to Chapter 10. The main results in Chapters 4, 5, and 6 are obtained through the imbedded Markov technique, thereby providing a unifying theme for the three chapters. The results in Chapters 6, 7, 8, and 9 are pertinent to the modeling of a particular form of computer communication network—the local area network. In fact, all of the well-known accessing techniques are considered. In order to cover Chapters 6 through 9, some material on the imbedded Markov chain is also necessary.

Chapters 10, 11, and 12 also provide a convenient grouping. Chapters 10 and 11 rely heavily on Jackson network theory. The linkage between Chapters 11 and 12 is that they treat the so-called "higher-level" protocols, in particular flow control and routing. Although the final chapter on network layout stands somewhat apart from the others, there are points of contact between network layout and routing. This later is the subject of Chapter 12.

In concluding these opening remarks we would like to express our deep gratitude to a number of people for their contributions to this work.

Jerry Foschini and John Silvester read an earlier version of the text and made a number of valuable suggestions for its improvement. Their diligent effort and encouragement were significant contributions. Thelma Hyland typed the manuscript. She was the soul of patience through the numerous revisions. A picture is worth a thousand words, and a thousand thanks are due to Miao Duen-Zhi for her assistance with the figures. The book evolved from notes of a course taught at McGill University over the period 1979–1983. A number of students suffered through the earlier (and later) versions of the notes. For their forbearance we express our deep gratitude. As mentioned earlier, the material in the book reflects our research. Our interest in computer communications began at Bell Labs, and was stimulated by Bob Lucky and Jack Salz. We would also like to acknowledge our first collaborator in the field, Dave Sherman. Finally, the author would like to pay tribute to his wife, Florence, and to his children, Mary, Ann, Jemmy, and Martin. They endured his absence and preoccupation during the long writing process.

JEREMIAH F. HAYES
*Concordia University*
*Montreal, Quebec*

# Contents

# 1

# Basic Orientation

## 1.1. Modeling and Analysis of Computer Communications Networks

Computer communications has recently emerged from its origins—the fields of computers and communications—as a separate, distinct area of study. We hope a complete picture of the field will become apparent in the course of the book; however, a useful short-term definition is that computer communications is the study of networks that are designed for the transmission of digital data. Such networks allow users to communicate with computers and computers to communicate with one another. The advent of distributed processing has been the great impetus to the development of computer communications. As we shall see in the course of our discussion, the properties of data signals are significantly different for voice signals. Because of this difference techniques employed in designing and operating computer communication networks are fundamentally different from those that have been used in the traditional voice network. However, these techniques are so well suited to digital technology that there is, at this writing, a full-scale effort to integrate voice traffic into the same networks.

The term "computer communications" encompasses a great range of activities and concepts. The immediate impression is that of daunting complexity as one is faced with the task of modeling and analysis. Nevertheless workers in the field have developed mathematical models which have provided a great deal of insight i[illegible]esign and performance of systems. The fundamental aim of this book is to present these models and to develop the basic tools for their analysis. To a large extent the models are based on queueing theory. Since it plays such a large role in computer communications, much of this book is concerned with queueing theory, beginning with its basic principles. Other disciplines have also had varying degrees of application, in particular, optimization theory, information theory, and statistical communication theory.

The basic objective of the mathematical modeling of computer communications system is the prediction of performance. In the generic queueing theory model, customers randomly arrive at a facility with service requirements that may be random in duration. The theory attempts to find probabilistic descriptions of such quantities as the sizes of waiting lines, the delay experienced by an arrival, and the availability of a service facility.

Queueing theory began with the work of the danish mathematician A. K. Erlang (1–3) on telephone switching systems. Although queueing finds wide use as one of the basic components of operations research, telecommunications remains its most successful application. In the voice telephone network, demands for service take the form of subscribers initiating a call. Erlang found that given a sufficiently large population, the random rate of such calls can be described by a Poisson process.† In this application the server is the telephone line. The analog of the service time of a customer is the duration of the call, which was found to have an exponential distribution. An important result obtained by Erlang, which is still used in engineering the telephone network, is the probability of a busy signal as a function of the traffic level and the number of lines that have been provided.

Although queueing theory was developed for voice traffic, it is applicable to computer communications. The generation of data messages at a computer terminal corresponds to the initiation of voice calls to the network. The time required to transmit the message over a communications facility corresponds to the required time of service. We shall deal exclusively with digital data; accordingly the transmission time is a function of the number of bits in the message and the data rate of the line. We shall consider digital transmission exclusively.

## 1.2. Computer Communications Networks

Broadly speaking, the function of a communications network is the transmission of information through shared transmission facilities. As we shall see in the course of the book, a large array of techniques are used to effect this sharing. In this introductory section we shall sketch them.

For purposes of explanation we may classify computer communication networks in two basic categories: local area networks and wide area networks. Local area networks have a limited geographical extent, typically within a kilometer. In a widely encountered example of a local area network the users are sitting at terminals accessing a main frame computer. A number of users share a common transmission line. Because of the localized nature

†Terms such as "Poisson process" and "exponential distribution" will be defined in Chapter 3

of the network, the network can have relatively simple topologies such as the ring, the bus, or the star (see Figure 2.20a, 2.20b, and 2.20c, respectively). A number of accessing techniques can be employed in conjunction with these topologies. For example, with the ring and the bus, token passing grants access to one user at a time. Users having access simply transmit data messages. The protocol governing the technique provides means for passing exclusive access from one user to another. In a contrasting technique used with the bus topology, users contend for the line by transmitting messages in a random fashion. The controlling protocol provides a mechanism for resolving conflict. In the modeling of these protocols we shall be concerned with quantities such as the time required to gain access and the overhead required by the protocol. The overhead has a large effect on throughput, which is the volume of traffic that can go through the network.

For networks covering a wider area the mesh configuration shown in Figure 2.19 is typical. User data are encapsulated in entities which we shall call packets. Since packets contain source and destination addresses, transmission lines can be shared in a flexible manner. Packets may travel from source to destination through a number of intermediate nodes. At each of these nodes the packet is processed. The addressing indicates the rout to be followed. Further, at the intermediate nodes the packet may be checked for errors which were incurred in transmission.

In wide area networks we are interested in performace at a number of levels. We shall consider the means of sharing the links between network nodes among a number of users. The total time required by a packet to travel through the network is of considerable interest. Because of the dispersed nature of the control of store-and-forward networks, congested spots can appear within the network. This phenomenon necessitates protocols for the control of flow into and within the network. Hand-in-hand with flow control protocols are routing protocols, which choose the sequence of nodes a packet must follow. Again the focus of the modeling is on throughput and delay.

## 1.3. Summary of Text

In this section we shall summarize the text. Before doing so, a general comment is in order. As mentioned above, the primary subject of the text is mathematical modeling and analysis of computer communications networks. Accordingly, we shall not be concerned with details of implementation. We shall present only the detail required to place a problem in context and to carry out the modeling and analysis.

Let us now proceed to consider the remaining chapters of the book in sequence. In order to smooth the flow in this first exposure, several terms are defined only implicitly or not at all. For definitions the reader is referred to the section of the book under discussion.

*Chapter 2: Protocols and Facilities*

Chapter 2 is mainly descriptive in tone. The idea is to establish the framework under which the modeling and analysis is carried out. For purposes of explanation we consider the basic components of a computer communication system to be what we shall call facilities and protocols, corresponding roughly to hardware and software in computer systems. The facilities are the communications systems. The flow of information on the facility is controlled by protocols. Under facilities we shall consider the telephone network, satellite systems, local area networks, mobile radio, and CATV. Now, given the means of communication, protocols shape the flow of information. For example, protocols translate electrical impulses into binary digits. In making this dichotomy there is something of a pedagogical artifice since the operation of the communication facilities involves control. However, in our discussion we shall focus on protocols only as they relate to data transmission.

We shall set out discussion of protocols in the framework of a system of protocols called open systems interconnection (OSI). This system is composed of seven entities called layers or levels. Each level contains a closely related set of functions. Careful attention is paid to the interface between levels, where information is exchanged between the sets of functions that make up the different layers. As part of the discussion of the protocols, we consider both baseband and passband data transmission techniques. In general terms we also consider packetizing, flow control, and routing. These subjects are treated from the aspect of an overview since they will be dealt with in detail in the remainder of the text. Protocols that are used in local area networks are treated in the same way since they are given a great deal of attention in the rest of the book.

*Chapter 3: Pure Birth and Birth–Death Processes*

Chapter 3 begins the work of establishing the mathematical models for the networks. The properties of the Poisson arrival process are derived from first principles. This is generalized to the pure birth process and then to the birth–death process. This process leads directly to queueing theoretic models. We develop the principles of the $M/M/S$ queue with infinite and finite storage capacity. By Little's formula the average delay can be related to the average number of messages in the system. The departure process from the $M/M/S$ queue is shown to be Poisson by means of Burke's theorem, a result which has considerable implications when we consider networks of queues in Chapters 10 and 11. In the final section of the chapter nonexponentially distributed service times are dealt with by means of Erlang's method of stages.

*Chapter 4: Imbedded Markov Chains: The M/G/1 Queue*

Messages with a general length distribution are considered under the assumption of Poisson arrival to a single server. The analysis is carried out by means of the imbedded Markov chain—a far simpler technique than the method of stages considered in the previous chapter. The results of the analysis are in the form of transforms of probability distributions of queue sizes and message delay. From these transforms averages and higher-order moments can be readily obtained. Also considered are methods for finding probability distributions of queue size and delay.

*Chapter 5: Time-Division Multiplexing*

The imbedded Markov chain approach is applied to the study of synchronous and asynchronous time-division multiplexed (TDM) systems. Bulk departure in TDM systems is analyzed by application of Rouche's theorem. As in Chapter 4, the results of analysis are queue size and message delay. Automatic repeat-request protocols are considered in the same general framework of the imbedded Markov chain.

In the Chapters 6, 7, 8, and 9, accessing techniques for local area networks are analyzed. All of these techniques can be viewed as ways of sharing a server, the transmission line among a number of message sources. Since the locations of these sources are different from one another, part of the capacity of the channel must be used to coordinate the flow of traffic. This overhead plays an important role in the analytical results on accessing techniques. In fact, the objective of the design of access protocols is the balancing of overhead and performance.

*Chapter 6: Priority Queues and Ring Systems*

As we have indicated, a prominent class of local area network has the ring topology. Certain of the accessing techniques used on ring systems can be viewed as giving one class of messages priority over another. The imbedded Markov chain technique is used to analyze systems with two or more priority classes. Again performance is measured in terms of queue size and message delay.

*Chapter 7: Polling and Token Passing*

An important accessing technique is polling, which can be viewed as a server visiting queues cyclically, giving equal service to each. The polling technique is appropriate to systems with a tree topology. The same basic model applies as well to an accessing technique called token passing. This technique is used on both ring and bus systems. The salient result of performance studies is the strong sensitivity to overhead. In this case overhead is the time required to pass channel access from one message source to another either by polling or by passing a token.

*Chapter 8: Random Access Systems*

The results of the previous chapter show the deleterious effect of overhead on performance, particularly at light loading. Random access techniques seek to minimize overhead by minimizing coordination between sources. As shown by the analysis presented in Chapter 8, random access systems are characterized by performance that is good at light loading, but which deteriorates rapidly with increasing load. The results of analysis also show that these systems also exhibit unstable behavior as the load increases. A comparison is made between two contending techniques for accessing in local area networks: token passing and carrier sense multiple access.

*Chapter 9: Probing and Tree Search Techniques*

In response to the effect of overhead in polling systems and instability in random access systems, a class of adaptive techniques, based on tree search, have been devised. The basic idea of the techniques is to minimize the overhead at light loading and, through adaptivity, to allow graceful degradation as the loading increases. Analysis of the tree search technique based on a Markov chain model of the system shows that the desired results have been obtained. The throughput is increased and the system is stable.

In Chapters 10, 11, and 12, the general subject under study is networks of queues. The application is the mesh network encountered in wide area networks. The particular aspect of these networks that is considered is the performance of higher-level protocols, routing, and flow control.

*Chapter 10: Networks of Queues*

Under certain assumptions networks whose nodes contain queues can be analyzed in a manner similar to birth–death processes. These networks of queues are the so-called Jackson networks. The importance of these networks lies in the fact that the steady-state probability distribution for the network assumes a particularly simple form of the product of the marginal probability distributions. This product form permits the computation of quantities of interest. In order to apply Jackson network theory to computer communications networks certain assumptions are necessary, one of which is Kleinrock's often quoted (and much abused) independence assumption. The main results of analysis is the calculation of message delay. This result is applied to the allocation of transmission capacity within the network so as to minimize cost while maintaining a certain standard of performance in terms of message delay.

*Chapter 11: Congestion and Flow Control*

Because of the random nature of traffic flow, it is possible for points of congestion to build up in the interior of a network. In this chapter techniques for controlling the flow into networks are modeled and analyzed. Flow control is studied by means of extended classes of Jackson networks with the objective of exploring both subjects. Both link and network levels are treated by means of closed networks of queues. Of particular interest are computational methods for closed networks of queues. Finally, Jackson theory is applied to the study of networks in which blocking due to limited storage capacity is possible.

*Chapter 12: Routing–Flow Allocation*

In general a message may take more than one path through a network in traveling from source to destination. In large measure the algorithms used to determine the route of a message are based on shortest-path routing. Classical algorithms of this sort which are traffic independent are given. This is followed by the consideration of techniques for minimizing the average delay of a message, taking into account queueing as well as other factors. These algorithms may be centralized or distributed. In a final section flow control and routing are considered in the same general context.

*Chapter 13: Network Layout and Reliability*

The final chapter of the text is something of a departure from the preceding from the points of view of both the subject matter and the mathematical foundations. The first subject taken up is minimum spanning trees. We begin with an algorithm which does not take capacity contraints into account. This is followed by the branch and bound technique, which finds minimum spanning trees under capacity constraints. While this technique is optimum, it requires too much computation for reasonable size networks. The Esau–Williams algorithm is suboptimum with the same objective.

The layout of mesh networks is the next topic of discussion. The complexity of the problem is such that only suboptimum techniques are feasible. One such technique which is based on certain graph theoretic ideas is considered.

In the final section of the text the reliability of networks is treated. Bounds for connectability are found under the assumption that network links fail independently of one another. A brief survey of other approaches to reliability is made as well.

## References

1. E. Brockmeyer, H. L. Halstrom, and A. Jensen, "The Life and Works of A. K. Erland" *Transactions of the Danish Academy of Technical Science ATS*, No. 2 (1948).
2. A. K. Erland, "The Theory of Probabilities and Telephone Conversations," *Nyt Tidsskrift Matematik B*, **20**, 33–39 (1909).
3. A. K. Erland, "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges," *Electroteknikeren*, **13**, 5–13 (1917) [in English: *PO Electrical Engineering Journal*, **10**, 189–197 (1917–1918).