

remedies for queueing delays. Problem sets are included in all chapters, which can help readers master queueing techniques.

SUPPLEMENTAL CASE STUDIES

In addition to the studies provided in this book, the following cases are excellent for in class discussion.

- Browne, J. 1984. *Management and Analysis of Service Operations*. New York: North-Holland. (contains several case studies drawn from the Port Authority of New York; can be used in Chapters 7, 11)
- Edie, L. C. 1954. "Traffic Delays at Toll Booths," *Journal of the Operations Research Society*, 2, 107–138. (Chapters 3, 7)
- Green, L. and P. Kolesar. 1989. "Testing the Validity of a Queueing Model of Police Patrol," *Management Science*, 35, 127–148. (Chapter 3)
- Horonjeff, R. 1969. "Analyses of Passenger and Baggage Flows in Airport Terminal Buildings," *Journal of Aircraft*, 6, 446–451. (Chapter 6)
- Ignall, E. J., P. Kolesar, A. J. Swersey, W. E. Walker, E. H. Blum, and G. Carter. 1975. "Improving the Deployment of New York City Fire Companies," *Interfaces*, 5:2, 48–61. (Chapter 11)
- Kolesar, P. 1984. "Stalking the Endangered CAT: A Queueing Analysis of Congestion at Automatic Teller Machines," *Interfaces*, 14:6, 16–26.
- Kolesar, P. J., K. L. Rider, T. B. Crabill, and W. E. Walker. 1975. "A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars," *Operations Research*, 23, 1045–1062. (Chapter 7)
- Linder, R. W. 1969. "The Development of Manpower and Facilities Planning Methods for Airline Telephone Reservations Offices," *Operational Research Quarterly*, 20, 3–21. (Chapter 7)
- Oliver, R. M. and A. H. Samuel. 1967. "Reducing Letter Delays in Post Offices," *Operations Research*, 10, 839–892. (Chapter 10)
- Vickrey, W. S. 1955. "Revising New York's Subway Fare Structure," *Journal of the Operations Research Society of America*, 2, 38–68. (Chapter 8)
- Welch, N. and J. Gussow. 1986. "Expansion of Canadian National Railway's Line Capacity," *Interfaces*, 16:1, 51–64. (Chapter 10)

Think back over your week and consider how much time you spent waiting in line—at the supermarket, in traffic, at the post office. Every moment you spent waiting for some type of service you were part of a queue. But queues are not always this obvious and they do not have to involve people. A suit waiting to be dry-cleaned is part of a queue; a memo waiting to be typed is part of a queue; and a lawsuit waiting to be heard in court is part of a queue. A queue is a group of people, tasks, or objects waiting to be served. Waiting is the essence of queueing.

Though we are most aware of our own waiting time, queueing is foremost a problem for industry and government. The success of any organization depends on maximizing the utilization of its resources. Every minute that an employee spends waiting for another department and every minute that a job spends waiting to be processed is money wasted. The success of any organization also depends on attracting and keeping customers. Every minute that a customer spends waiting to be served translates into lost business and lost revenue.

Continued growth in the service industries in the United States has heightened the need to manage and control queueing effectively. Though queueing is also an important concern to manufacturers, it is felt especially strongly in the service sector because of the heavy reliance on customer interaction. In their article "Will Services Follow Manufacturing into Decline?" James Quinn and Christopher Gagnon highlight the need to keep service industries competitive:

Although they probably know better, many executives still think of the service sector in terms of people making hamburgers or shining shoes. These images belie the complexity, power, technical sophistication, and economic value of activities that now account for more than 68% of the nation's GNP and 71% of its employment. Worse, they help perpetuate a set of myths about service industries that lead managers and policymakers to ignore their full potential. Worse still, such inattention and complacency threaten to undermine the competitive ability of these industries at a time when their importance to the national economy has never been greater.¹

Though not always recognized as such, queueing problems in the service sector constantly appear in the headlines of major newspapers. Note these examples:

Airline Delays

Operating the Eastern shuttle on time is one of the toughest jobs in the industry. The New York-Washington corridor traverses some of the most congested airspace in the world. Shuttle pilots must dodge a multitude of other planes, especially near New York's three major airports, La Guardia, Kennedy and Newark. . . . Congestion in the sky can quickly cause a frustrating backup of 20 planes waiting to take off at either La Guardia or at National airport in Washington.²

Medical Transplants

"Modern Technology has provided us with a double-edged sword," said Dr. Gary Friedlander, chief of orthopedics at Yale University. "On the one hand, organ transplants are miracles that will save many lives. On the other hand, we are faced with thousands of people whose lives potentially could be saved with a transplant but who will die because they are not available." The hard truth is there are nowhere near enough donor organs and probably never will be.³

Court Congestion

New York City's criminal-justice system is in a state of crisis, just barely able to cope with a growing flood of new drug cases generated by the Police Department's drive against crack. . . . So desperate is the situation . . . that without a major new infusion of money, manpower, new courtrooms and more jail space, the system will be swamped by this summer. . . . As a result, judges and district attorneys say, more and more accused drug traffickers and users will simply be returned to the streets, their cases dismissed or the charges greatly reduced.⁴

Fortunately, queues are not problems without solutions. Queueing problems go unsolved only because the organizations causing the problems do not directly pay the price. The server does not bear the cost and aggravation of waiting. And though long waits likely translate into lost business, organizations do not always see the cause and effect. Lost business can be blamed on any number of things—competition, weather, or

other "acts of God"—which have nothing to do with queueing. Casting blame elsewhere perpetuates the problem.

In *Quality Is Free*, Philip Crosby challenges the notion that product quality depends on a trade-off with cost. He writes: "Quality is free. . . . What costs money are the unquality things—all the actions that involve not doing jobs right the first time."⁵ Just as no one benefits from poor quality, no one directly benefits from queueing. The time customers spend waiting does not benefit the organization providing the service, and it surely does not benefit the customer. Time spent queueing is time wasted forever.

The point to recognize is that queueing is an important problem that affects us all both directly and indirectly. But just as importantly queueing is a solvable problem. The solution may not be simple, but with some understanding and an ability to put itself in the customer's shoes, an organization can substantially reduce waiting times at little or no expense.

1.1 HISTORY OF QUEUEING

Since the day that humans gathered into societies, there have been queues. History books are replete with occurrences of citizens waiting for the dole, courtiers waiting for an audience with the king, or the unemployed waiting for a bowl of soup. Though queueing is by no means new, the study of queues is modern, dating only to the beginning of the twentieth century and the work of A. K. Erlang, a man who spent years investigating a then cutting-edge technology—telephones (see Brockmeyer et al. 1948).

Theoretical analysis of queueing systems grew considerably with the advent of operations research in the late 1940s and early 1950s. The first textbook on the subject, *Queues, Inventories, and Maintenance*, was written in 1958 by Morse. Saaty wrote his famous *Elements of Queueing Theory with Applications* in 1961, and Kleinrock completed his *Queueing Systems* in 1976. Today, more than 40 books have been written on the subject.

Over the last 30 years the mathematics of queueing systems has advanced tremendously. Journals such as *Operations Research* and *Naval Research Logistics Quarterly* frequently contain contributions on the subject, and the total number of papers exceeds 1000. Less has been written on the scientific aspects of queueing systems—that is, on how real queues actually behave. Most of the work in this area is found in applications-oriented journals, such as *Interfaces*, *Transportation Science*, and *International Journal of Production Research*.

1.2 ELEMENTS OF QUEUEING SYSTEMS

It may seem that a queue of lawsuits waiting for trial has little in common with a queue of shoppers waiting at the checkout counter. In fact, both possess the same basic elements. They are both processes by which "customers" wait for service. This concept may seem

straightforward, but the terms *customer*, *server*, and *queue* do require elaboration if they are to be applied to disparate systems.

The customer is the person or thing that waits for service. In the case of a queue of people waiting to buy stamps, the “customer” is a customer in the ordinary sense of the word. But identifying the customer is not always trivial. At a supermarket checkout counter, the customer may be either the shopper in line or the items being purchased, depending on one’s point of view. Both are waiting for service. *A customer does not have to be a person.*

The server is the person or thing providing the service. Like the customer, identifying the server is sometimes obvious and sometimes not. For the stamp queue, the server might simply be the postal clerk. But if one takes a broader perspective, the server might encompass the clerk, a cash register, and other machines that serve the customer in concert. All are necessary to perform the service and all are part of the server.

The queue is the group of customers waiting to be served. The queue does not have to be an orderly line. It does not even have to be visible. The queue is simply the group of customers that have requested, but have not yet received, service. The request for service can take many different forms: the physical arrival of a person at the back of a line, a phone call, a submission of a piece of work, or the like.

Thus, a queueing system has just three basic elements: customer, server, and queue. Despite the apparent simplicity, identifying the basic elements is not always easy and is almost always open to debate. There are even times when the roles of customer and server are interchanged (an idle server might be viewed as a “customer” waiting for a true customer to serve).

1.3 QUEUEING CHARACTERISTICS

The best way to understand how a queue operates is to examine the characteristics of the basic queueing elements. Here are some of the things to consider.

1.3.1 Customer

The *arrival process* depicts the timing of customer arrivals at the queue. Do customers arrive independently of each other, or do they arrive in groups? Do customers arrive at a fairly constant rate, or is there some pattern to their arrivals (for example, a “rush hour”)? Is the arrival process predictable or random? Do different types of customers arrive at different times of the day?

In addition to understanding customer arrivals, one should also understand customer reneging, balking, and jockeying. **Reneging** is the act of leaving a queue before being served; **balking** is the act of not joining a queue upon arrival. Conceptually, the two concepts are the same. The only difference is the timing of when the customer leaves

Sec. 1.4 Examples of Queueing Systems

(either immediately in the case of balking or later in the case of reneging). For this reason, both actions will be called *reneging* in this book. **Jockeying** is the act of switching from one queue to another. Reneging, balking, and jockeying are three of the most difficult aspects of a queueing system to measure because the customer may never be recorded by the system. How long must a queue be before customers renege? Are some types of customers more likely to renege than others? Will a customer that reneges come back or never return? If the customer returns, when will that be?

1.3.2 Server

The *service process* represents the time taken to serve customers, commonly referred to as the *service time*. Is the service time constant, or does it vary from customer to customer? Are customers served in bulk, as at a traffic signal? Does the service time depend on the type of customer, and can it be predicted in advance? Does the service time depend on the server or the time of day? Is there any way to shorten the service time by improving server efficiency?

A second important aspect of the server is its *configuration*. How many servers are there? Which servers work in parallel, performing identical tasks, and which perform in series, performing different tasks? What are the rules governing how customers move from server to server?

1.3.3 Queue

The *queue discipline* specifies the order in which customers in the queue are served. Are customers served on a first-come, first-served (FCFS) basis, or perhaps on a last-come, first-served (LCFS) basis? Do different customers receive different priority, and if so, what is the system for assigning priority? Is service for a customer ever interrupted to serve another with higher priority?

A second important characteristic of the queue is its general *organization*. Is there a single queue that feeds all servers, separate queues at each server, or some variation of the two? Is there a limit on the total number of customers in the queue? What is the system for keeping track of customers and ensuring orderliness?

These are some of the key questions to ask in assessing a queueing system. Many of these aspects can be changed in order to improve the performance of the system; others are beyond control. Regardless, understanding these aspects is the first step toward improving a queueing system’s performance.

1.4 EXAMPLES OF QUEUEING SYSTEMS

To give you a better feel for the basic queue elements, this section describes three familiar queueing systems: bank, supermarket, and motor vehicles department. These systems represent three common configurations: parallel servers fed by a single queue; parallel servers, each fed by its own queue; and a combination of parallel and serial servers.

1.4.1 Bank

The primary functions of commercial banks are borrowing and lending money. Borrowers are served by loan officers, and lenders (that is, depositors) are served by account officers and cashiers. Depending on the transaction, a bank client might visit any of several places inside the bank (Fig. 1.1).

Let us concentrate on the depositor who wishes to make a withdrawal. The depositor has two choices for service: an ordinary (human) teller or an automated (computer) teller. For the first case, the client first walks to a desk and fills out a withdrawal form, then joins a first-come, first-served queue to wait for the first available teller. In the second case, the client does not fill out a withdrawal form, but instead joins a single queue waiting for the automated teller.

The two types of tellers are represented by the queue diagram in Fig. 1.2. The upside-down triangle is the symbol for a queue, and the circle is the symbol for a server. The arrows represent the movement of customers from queue to server. The ordinary tellers work in parallel and are fed by a single queue. The automated teller works alone and is also fed by a single queue.

What Is the Customer? Either the depositor or the transaction can be viewed as the customer. If a depositor wishes to perform several transactions, each transaction might be considered a separate customer.

What Is the Server? The server is either an automated teller or the human teller (perhaps in concert with a computer console and other machines, some of which may be shared with other tellers). Less directly, a centralized computer also serves the customer.

The configuration, queue discipline, and queue organization have been outlined. Consider on your own the other key characteristics of the system: What is the arrival process—Are there times of the day or week when customers arrive at an especially fast rate? What is the reneging process—Do customers renege from the ordinary teller queue to join the automated queue? What is the service process—Do some transactions take longer than others, and why? In light of these characteristics, consider why the system is designed the way it is.

1.4.2 Supermarket

Supermarkets sell a wide variety of products ranging from produce to packaged food to stationery. A supermarket operates on a self-service basis. Shoppers take a cart as they enter the store and walk up and down the aisles, retrieving goods along the way. After selecting their goods, shoppers join the queue for one of the checkout counters and wait for a checker to compute the bill, accept payment, and bag the groceries.

Figure 1.3 shows the checkout configuration for a market with eight checkout counters. Each counter has a cash register, a scale (for weighing produce), and a laser scanner (for reading bar codes from packaged goods). Two of these counters can be used for quick-check (12 items or less). Thus, customers making small purchases have some

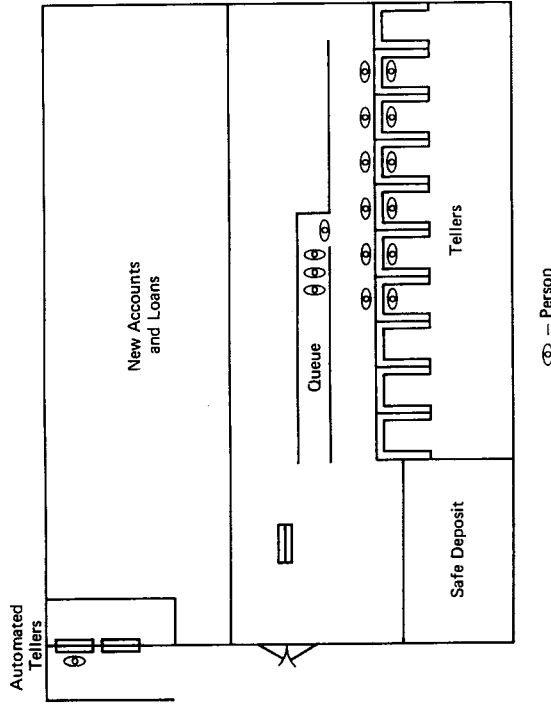


Figure 1.1 Bank layout showing parallel servers fed by a single queue.

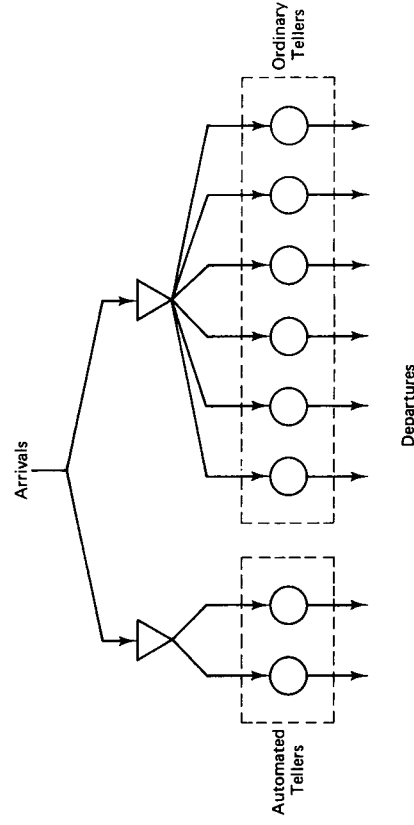


Figure 1.2 Queue diagram of bank.

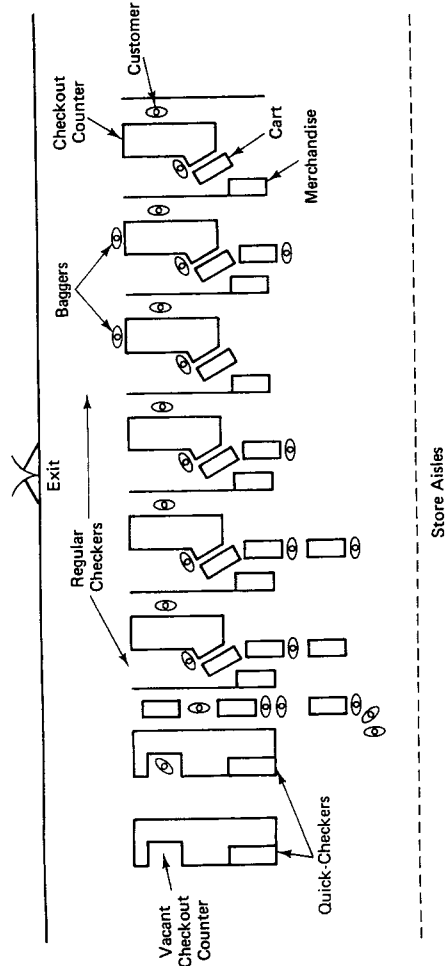


Figure 1.3 Grocery store layout showing parallel servers fed by separate queues.

priority over other customers. The number of counters in operation varies with demand. Each counter can be served by one (checker only) or two (checker and bagger) people, also depending on demand. These features are illustrated by the queue diagram in Fig. 1.4, which shows a system with parallel servers, each fed by its own queue.

What Is the Customer? The customer is either the shopper or the set of items purchased (along with the payment transaction). In the latter case, the queue diagram might be more detailed and account for queues occurring at the various pieces of equipment (for example, scale versus laser scanner).

What Is the Server? The server is the checker along with the cash register, scanner, and scale. When used, the bagger is also part of the server.

Again, the configuration, queue organization, and queue discipline have been outlined. Consider on your own: What is the arrival process—Are there certain times of the day, week, or month when the arrival rate is particularly large? What is the reneging process—Will a customer renege after walking in the door, after selecting the products, or some time between? What is the service process—Does service time vary among customers, can it be predicted in advance, and what can the customer do to influence it? And, with multiple queues, what is the jockeying process—At what points are customers likely to jockey over to a shorter line? How do grocery carts affect jockeying? In light of these characteristics, consider why queues are organized differently at supermarkets than at banks, and why some supermarkets are organized differently than others.

1.4.3 Motor Vehicles Department

The California Motor Vehicles Department issues driver's licenses, identification cards, and vehicle registrations. To be issued a license, first-time California drivers must pass a road test as well as a written test. A license is renewed by passing the written test alone.

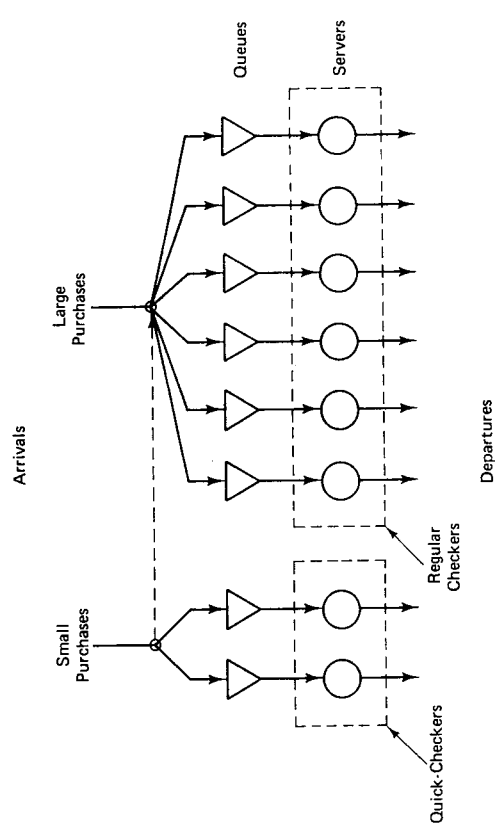


Figure 1.4 Queue diagram of grocery store.

Vehicle registrations are normally handled through the mail, though late registrations must be completed at the motor vehicles office.

A driver wishing to renew a license first fills out a form, then joins an FCFS queue to wait for the first available clerk to type the information. The clerk provides the driver with a copy of the written test and directs him or her to a desk to complete the test. Once finished, the driver joins a second FCFS queue to wait for the first available grader to score the test and check the driver's eyesight. If the driver passes the test, he or she then joins a third queue to have his or her picture taken. The grader also takes the photograph, but a single camera is shared among all graders. The actual license arrives two to three weeks later in the mail. If a driver is new to California, then he or she must also complete a road test subsequent to passing the written test. Figure 1.5 shows the layout for a motor vehicles office. Figure 1.6 is the flow diagram and shows that the system has a combination of serial and parallel servers.

What Is the Customer? The driver is the customer, and all customers are nearly identical. However, for the last queue, the grader and the driver wait for the camera together. That is, both are, in a sense, the customer.

What Is the Server? Three servers perform three distinct tasks: clerk/typewriter types the form, grader scores test and administers eye exam, and camera takes picture. Though the grader is needed to take the picture, it is the camera that provides the service.

Again, the server configuration, queue configuration, and queue discipline have already been outlined. Consider on your own: What is the arrival process—Does the department of motor vehicles have any information that can help predict, or control, arrivals? What is the reneging process—If a person reneges, will he or she come back

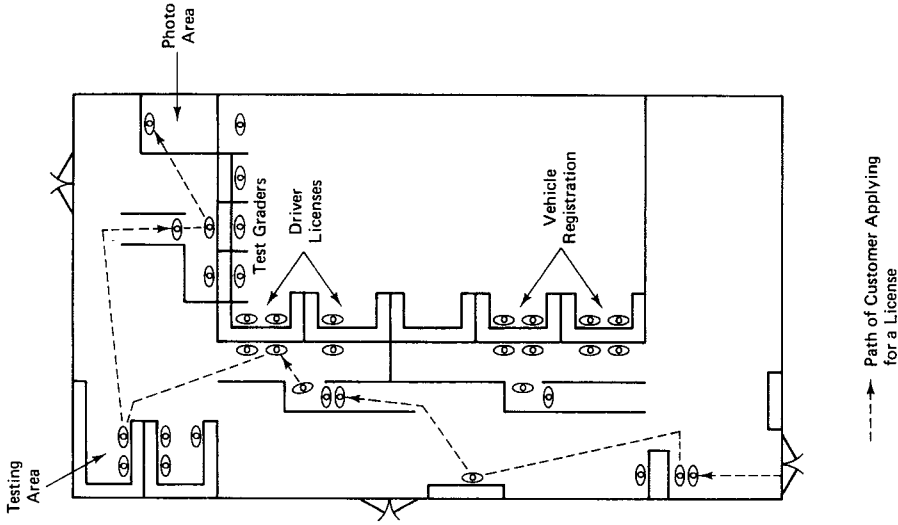


Figure 1.5 Layout showing serial servers in motor vehicles department.

later? What is the service process—Does service time vary from customer to customer? In light of these characteristics, consider why servers are assigned distinct tasks rather than identical tasks.

1.4.4 Other Queues

The bank, supermarket, and motor vehicles department examples were chosen for their familiarity, not because they are in any way unique. Other queues are not so familiar. For example, income tax returns pass through several queues as they are processed (cashing

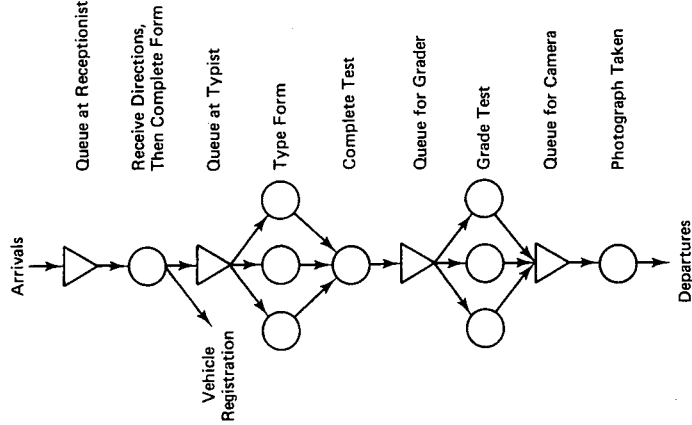


Figure 1.6 Queue diagram of motor vehicles department.

check, recording information, verifying information, and so on). Phone calls are transformed into data packets that queue at switching centers. Silicon wafers form queues at fabrication machines inside manufacturing plants. Other queues may be familiar but are more difficult to classify. For example, congestion forms on freeways when the lanes do not have sufficient capacity to accommodate all of the vehicles. The server is not a machine or a person but a restriction that impedes the flow of vehicles, the service time being defined by the minimum spacing between vehicles.

1.5 ATTITUDES TOWARD QUEUEING

An East German visits a friend in Moscow. A little girl answers the door. The East German asks her:

- Where is your father?
- He is not home.
- When will he be at home?
- At eight hours, forty minutes and twenty-three seconds.

- Where is he?
- He is going around the world thirty-three and a half times.
- What about your mother? Is she at home?
- No, she isn't.
- When is she expected?
- I don't have any idea.
- How come you know the hour, minute, and second when your father will return, but you have no idea when your mother will return? Where is she?
- She is at the market in the line for meat.⁶

One does not have to live in Eastern Europe to appreciate the humor in this story. Dislike (some would even say hatred) of queues is probably universal. Though there is much that different cultures hold in common, behavior in queues is not always the same. Adherence to the principle of first-come, first-served is one example. In the words of the sociologist Edward Hall, "To Americans, to be first is to be more deserving. If an American has been sitting at a table in a restaurant for some time and a latecomer is served before he is, his blood pressure will rise noticeably. Yet in most places outside of Europe [FCFS] ordering in situations of this type is unknown. Instead the laws of selection apply; that is, service is dependent upon a person's rank" (1959, p. 157).

Leon Mann's study of queues to buy tickets for Australian football matches is indicative. The great popularity of the game dictated that fans had to wait overnight, and even up to six days, to be assured a seat. Mann describes the fans' efforts to maintain order in the absence of strict control: "People in the middle of the queue worked together to erect barricades from material left in the park" as constraints against would-be infiltrators (1969, p. 347). He also notes that "at times of maximum danger, and in the hour before the ticket windows opened, there was a visible bunching together, or shrinkage, in the physical length of the queue, literally a closing of the ranks" to prevent infiltration. In addition to "vociferous catcalls and jeering," Mann observes that "the most extreme constraint was physical force. During the early hours of August 15, five men were taken to hospitals after four separate brawls broke out in ticket lines."

Though first-come, first-served is the norm for queues of people waiting for service, such is not the case for all queues. In business, it is considered proper to give priority to the employee with the highest rank (for example, the department manager gets her memos typed first). Other times, priority is given to the work with the most urgent deadline. Favors can also influence priorities (such as tipping the maître d'hôtel to obtain preferential seating at a restaurant). And whereas people tend to queue in orderly lines for tellers or ticketing (Fig. 1.7), they do not in all cases, such as waiting to enter an elevator or cross a street (Fig. 1.8). In extreme situations, lack of control can lead to panic, as in stampedes to enter stadiums for rock-and-roll concerts and soccer matches.

Maintaining a pleasant waiting environment can also have a tremendous impact on our attitudes. Systems for maintaining order, such as roped barriers and signs, could have eliminated the anxiety witnessed in the Australian football queues. They can also prevent crowding, as when customers bunch together to prevent infiltration of the line. Adequate



Figure 1.7 Customers waiting in orderly queue to purchase tickets.
Source: Courtesy of Project for Public Spaces, Inc., New York.

lighting, ventilation, and sound control, as well as smoking prohibitions, are also important. Better yet is a system in which a patron can drop off a piece of work (the work being the customer) and retrieve it later, without ever having to be present in the queue.

The attitude of the server toward the customer is also significant. Servers should be aware that the act of waiting automatically puts the customer in a subordinate position. Thus, lack of regard for customer needs not only delays the customer but angers him or her. Monopolies, both private and public, are notorious for an indifferent attitude.



Figure 1.8 Queue at an intersection is not FCFS.
Source: Courtesy of Project for Public Spaces, Inc., New York.

1.6 SYSTEMS APPROACH

The ideas presented in this book fit within a larger scheme known as the systems approach (see Churchman 1968; de Neufville and Stafford 1971). The end objective of the approach is to decide the best way to operate a system—in the case of this book, the system being a queueing system. This objective is achieved with these four steps:

1. Formulation
2. Modeling
3. Evaluation
4. Decision

1.6.1 Formulation

The purpose of the formulation step is to assess the characteristics of the situation at hand and then isolate and define a “problem.” One might think of this as putting a box around a part of the system. Anything inside the box is a matter of interest; anything outside is taken as given. In the three queueing examples (bank, supermarket, and motor vehicles department), the system was defined by isolating the activities that occur at a given location and ignoring things that happen elsewhere. Problem definition is an essential first step toward problem solution.

The mathematician G. Polya (1945) emphasizes answering these questions in defining a problem:

- What is the unknown?
- What are the data?
- What is the condition?

The unknown are those things that can be controlled. In the parlance of operations research, the unknown are the decision variables. The number of servers, the server configuration, and the queue organization are three examples. The data are the quantitative characteristics of the situation—perhaps a history of customer arrivals and service times. These data are the facts that can be used to arrive at the best possible decision. The condition is a word description of the situation, such as the type of service requested, restrictions on server configurations, and so on.

1.6.2 Modeling

The purpose of modeling is to develop a representation of the system that provides a better understanding of how the system operates. A model might be physical (for example, a scale model of a bank), conceptual (a word description of how a bank operates), or mathematical (an equation showing average waiting time as a function of customer arrival rate). Of these three types, mathematical models are most applicable to queues, and this is

where the book concentrates. However, physical and conceptual models will also be mentioned.

Creating good models is by no means a simple task. It requires a mixture of scientific talent and a good dose of creativity. Fundamentally, the two main considerations are that the model should be realistic and the model should be meaningful. The model should also be suited to the problem at hand and possess just the right amount of detail. If the model is not realistic, it cannot be trusted, and if the model is not meaningful, it cannot be understood. Unfortunately, these two qualifications usually conflict. One can go to great effort to create a very realistic model, only to find that it provides no insights into how the system operates. The important message is that a balance should be achieved.

Creating good models is not a matter of creating the “right” models, for there is rarely one right answer. We have already seen that the customer can be defined in any of several ways. How the customer is represented is but the first point in the modeling process, and the first point where the modeler’s judgment must be exercised. The same type of judgment must be exercised throughout the process.

1.6.3. Evaluation

After the model is created, the next step is to use it to evaluate the system. The evaluation step can be divided into two parts: generation of alternatives and evaluation of alternatives.

The purpose of the first step is to find alternative ways to run the system. The model might provide insights into alternative configurations—more servers, serial servers, or the like. Optimization techniques might also be applied to a simplified model to identify how many servers to operate or server schedules. Or the alternative might come from one’s own imagination. The best ideas often result from one’s own observations of the system.

After the alternatives are identified, they can be analyzed more rigorously with a detailed model. This usually involves predicting how well the alternatives will work according to quantitative *measures of performance (MOP)*. What is the cost of the system? What is the average waiting time? How many customers will renege? These are some of the possible MOPs. A measure of performance should, first of all, be measurable—something that can actually be recorded by observing the queue. And the measure of performance should provide an important indicator of how well the alternative meets major system objectives, such as maximizing return on investment.

1.6.4 Decision

The decision phase is the least structured of the steps in systems analysis. The main objective here is to consider the information obtained from the evaluation phase and select the best alternative. Because objectives may be conflicting, and some alternatives may score well according to some measures of performance and not according to others, considerable judgment must be exercised. The most important thing to recognize is that

the best alternative is not an automatic by-product of the evaluation process. The model does not pick the best alternative. Rather, the model is used to provide the information needed for a *person* to select the best alternative. There is no simple prescription.

1.6.5 Systems Analysis in Queueing

This book does not cover all the steps of the systems analysis process as it applies to queues. The emphasis here is on the observation and modeling aspects. Other pertinent subjects, such as evaluation techniques and decision making, are covered in books on operations research, decision analysis, and systems analysis (references are provided at the end of this chapter). Anyone who wishes to develop and analyze queueing models should also be exposed to these areas.

Despite the appeal of the systems approach, do not feel compelled to follow it whenever a problem is confronted. With a little understanding of how queues behave, the solution might become obvious. If you already know how to fix a problem, there is no point in studying it.

1.7 BOOK ORGANIZATION

This book is written with the intention of developing the understanding needed to diagnose and correct queueing problems. It will provide awareness of what solutions can be applied to a given situation and exposure to various modeling techniques (simulation, fluid approximations, stochastic models, and so on). The book is written in a cumulative fashion, each chapter dependent on the material in the preceding chapter.

The first half of the book, Chaps. 2 to 6, explores the behavior of simple queueing systems. Understanding how a queue operates must always begin with observation and measurement, the subject of Chap. 2. This is followed by a group of four chapters on modeling queueing systems. The first of these, Chap. 3, introduces the Poisson process as a model for customer arrivals. Chapter 4 explains the philosophy behind simulation and describes how to use simulation to model a random queueing process. In Chap. 5, the behavior of queueing systems that operate in steady state is described. And Chap. 6 examines queueing systems where the customer arrival rate is not constant but varies over time.

Chapters 7 through 9 concentrate on solving queueing problems, and are somewhat less mathematical than the previous four. Chapter 7 provides methods for reducing waiting time through changes in the service process. Chapter 8 also provides suggestions for reducing waiting time, but through changes in the arrival process. Chapter 9 examines the queue discipline and ways to reduce the cost of delay by adjusting the order in which customers are served.

Chapter 10 covers queueing systems with multiple servers, examining queueing behavior and solutions to queueing problems. Included in Chap. 10 is a description of some of the computer packages available for simulating the performance of queueing networks.

Finally, Chap. 11 looks at the design aspects of queueing systems. It includes such topics as keeping track of customers, queueing layout, queueing location, and queueing environment.

Chapters 2 to 5 include a series of exercises in which students can observe a queueing system, measure its performance, and create a queueing model. The aim here is to provide some experience in model development. Chapters 7 to 10 consider a series of case studies, describing the operation of a variety of real queueing systems. The studies serve to illustrate the concepts presented in each chapter. In particular, they can provide the focus for class discussion on ways to eliminate queueing problems. Chapter 11 includes a design exercise addressed at improving all aspects of queueing performance.

FURTHER READING

- BROCKMEYER, E., H. A. HALSTROM, AND A. JENSEN. 1948. *The Life and Works of A. K. Erlang*. Copenhagen: Danish Academy of Technical Science.
- CHURCHMAN, C. W. 1968. *The Systems Approach*. New York: Dell.
- _____, R. L. ACKOFF, AND E. L. ARNOFF. 1957. *Introduction to Operations Research*. New York: John Wiley.
- DE NEUFVILLE, R., AND J. H. STAFFORD. 1971. *Systems Analysis for Engineers and Managers*. New York: McGraw-Hill.
- EDIE, L. C. 1954. "Traffic Delays at Toll Booths," *Journal of the Operations Research Society of America*, 2, 107-138.
- HALL, E. T. 1959. *The Silent Language*. Garden City, N.Y.: Doubleday.
- HALL, R. W. 1985. "What's So Scientific about MS/OR?" *Interfaces*, 15, 40-45.
- KLEINROCK, L. 1975, 1976. *Queueing Systems*. New York: John Wiley.
- MANN, L. 1969. "Queue Culture: The Waiting Line as a Social System," *American Journal of Sociology*, 75, 340-354.
- MORSE, P. 1958. *Queues, Inventories and Maintenance*. New York: John Wiley.
- OLIVER, R. M., AND A. H. SAMUEL. 1967. "Reducing Letter Delays in Post Offices," *Operations Research*, 10, 839-892.
- POLYA, G. 1945. *How to Solve It*. Princeton, N.J.: Princeton University Press.
- RAIFFA, H. 1968. *Decision Analysis, Introductory Lectures on Choices under Uncertainty*. Menlo Park, Calif.: Addison-Wesley.
- SAATY, T. L. 1961. *Elements of Queueing Theory with Applications*. New York: McGraw-Hill.
- SCHWARTZ, B. 1975. *Queueing and Waiting, Studies in the Social Organization of Access and Delay*. Chicago: University of Chicago Press.

NOTES

1. Reprinted by permission of the *Harvard Business Review*. Excerpt from "Will Services Follow Manufacturing into Decline?" by James Brian Quinn and Christopher E. Gagnon (November/December 1986), p. 95. Copyright © 1986 by the President and Fellows of Harvard College. All rights reserved.

2. W. Carley, "A Ride on the Eastern Shuttle Is Rush Hour in Crowded Skies," *The Wall Street Journal*, September 24, 1985, p. 31. Reprinted by permission of *The Wall Street Journal*, © Dow Jones and Company, Inc., 1985. All rights reserved.
3. R. Kotulak, "Transplant Gift in Short Supply," *Chicago Tribune*, July 16, 1986, p. 1. © Copyrighted, Chicago Tribune Company. All rights reserved. Used with permission.
4. D. E. Pitt, "Drug Cases Clog New York Courts," *The New York Times*, April 4, 1989. © Copyright 1989 by The New York Times Company. Reprinted by permission.
5. P. B. Crosby, *Quality Is Free* (New York: McGraw-Hill, 1979), p. 1.
6. From *First Prize: Fifteen Years!* by C. Banc and A. Dundes. Fairleigh Dickinson University Press, Rutherford, N.J., 1986, p. 58. Reprinted by permission of Associated University Presses.

PROBLEMS

1. Based on your understanding of each of the following queueing systems, describe the server, customer, queue organization, queue discipline, and reneging and jockeying processes.
 - a. Toll booths on a highway
 - b. Traffic intersection controlled by a traffic light
 - c. Taxi service at a hotel
 - d. Elevators in a large building
 - e. Copying service at a copy shop
 - f. Mail sorting by the postal service
 - g. Medical service at a health center
 - h. Airplanes landing at an airport
 - i. Emergency telephone lines
 - j. A sit-down restaurant
 - k. Time-shared computer system
2. Describe a situation (or situations) in which waiting or delay is harmful to productivity. Discuss how the situation might be improved.

EXERCISE: QUEUEING TOUR

Select five queueing systems to visit and observe (your instructor may provide suggestions). Answer the following questions for each site. Be succinct.

1. Describe the server(s) and service process. How many servers are there?
2. Describe the customer(s).
3. Diagram the server configuration and the queue organization.
4. Record the service time for five successive customers. Describe the factors that influence the service time.
5. Count and record the number of customers that arrive over five successive 1-minute intervals. Describe the factors that influence the arrival process.
6. If you observe any reneging or jockeying, describe the process.
7. Describe the queue discipline.
8. Record the time of day and the date.

Observation and Measurement

A doctor examining a patient looks for symptoms of whether the patient is healthy or ill. Blood pressure, body temperature, and pulse are just a few of the things to check. An analyst examining a queueing system also looks for symptoms of health or illness. These symptoms are known as measures of performance (MOP).

Observing the queueing system is the key to the formulation stage of systems analysis. It directly answers the questions: What are the data? and What are the conditions? Understanding how the system currently operates should also be the basis for choosing a new form of operation. Observation should provide insights to help answer the question: What is the unknown? That is, it should help identify, define, and isolate the problem.

Chapter 1 describes how to assess the characteristics of a queueing system. This chapter describes how to measure the performance of the system. It details which queueing symptoms to examine and how they should be measured. It begins by listing some of the key MOPs that apply to almost any queueing system. Next, a major topic that will be used throughout the book is introduced: cumulative arrival and departure diagrams. These diagrams are used to display and analyze the most important MOPs. The diagrams are followed by a discussion of how to calculate averages and standard deviations of the performance measures and how to create an empirical probability distribution. Little's formula is introduced. The chapter concludes with a review of techniques for