CHAPTER 13

Waiting Lines and Queuing Theory Models

LEARNING OBJECTIVES

After completing this chapter, students will be able to:

1. Describe the trade-off curves for cost-of-waiting time and cost of service.
2. Understand the three parts of a queuing system: the calling population, the queue itself, and the service facility.
3. Describe the basic queuing system configurations.
4. Understand the assumptions of the common models dealt with in this chapter.
5. Analyze a variety of operating characteristics of waiting lines.

CHAPTER OUTLINE

13.1 Introduction

The study of *waiting lines*, called **queuing theory**, is one of the oldest and most widely used quantitative analysis techniques. Waiting lines are an everyday occurrence, affecting people shopping for groceries, buying gasoline, making a bank deposit, or waiting on the telephone for the first available airline reservationist to answer. Queues,* another term for waiting lines, may also take the form of machines waiting to be repaired, trucks in line to be unloaded, or airplanes lined up on a runway waiting for permission to take off. The three basic components of a queuing process are arrivals, service facilities, and the actual waiting line.

In this chapter we discuss how analytical models of waiting lines can help managers evaluate the cost and effectiveness of service systems. We begin with a look at waiting line costs and then describe the characteristics of waiting lines and the underlying mathematical assumptions used to develop queuing models. We also provide the equations needed to compute the operating characteristics of a service system and show examples of how they are used. Later in the chapter, you will see how to save computational time by applying queuing tables and by running waiting line computer programs.

13.2 Waiting Line Costs

One of the goals of queuing analysis is finding the best level of service for an organization.

Most **waiting line** problems are centered on the question of finding the ideal level of services that a firm should provide. Supermarkets must decide how many cash register checkout positions should be opened. Gasoline stations must decide how many pumps should be opened and how many attendants should be on duty. Manufacturing plants must determine the optimal number of mechanics to have on duty each shift to repair machines that break down. Banks must decide how many teller windows to keep open to serve customers during various hours of the day. In most cases, this level of service is an option over which management has control. An extra teller, for example, can be borrowed from another chore or can be hired and trained quickly if demand warrants it. This may not always be the case, though. A plant may not be able to locate or hire skilled mechanics to repair sophisticated electronic machinery.

When an organization *does* have control, its objective is usually to find a happy medium between two extremes. On the one hand, a firm can retain a large staff and provide *many* service facilities. This may result in excellent customer service, with seldom more than one or two customers in a queue. Customers are kept happy with the quick response and appreciate the convenience. This, however, can become expensive.

The other extreme is to have the *minimum* possible number of checkout lines, gas pumps, or teller windows open. This keeps the **service cost** down but may result in customer dissatisfaction. How many times would you return to a large discount department store that had only one cash register open during the day you shop? As the average length of the queue increases and poor service results, customers and goodwill may be lost.

Managers must deal with the trade-off between the cost of providing good service and the cost of customer waiting time. The latter may be hard to quantify.

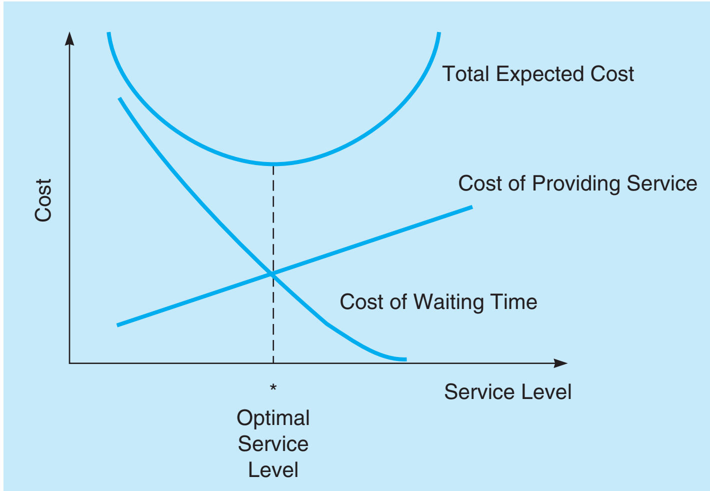
Most managers recognize the trade-off that must take place between the cost of providing good service and the cost of customer waiting time. They want queues that are short enough so that customers don't become unhappy and either storm out without buying or buy but never return. But they are willing to allow some waiting in line if it is balanced by a significant savings in service costs.

HISTORY How Queuing Models Began

Queuing theory had its beginning in the research work of a Danish engineer named A. K. Erlang. In 1909, Erlang experimented with fluctuating demand in telephone traffic. Eight years later, he published a report addressing the delays in automatic dialing equipment. At the end of World War II, Erlang's early work was extended to more general problems and to business applications of waiting lines.

*The word *queue* is pronounced like the letter *Q*, that is, "kew."

FIGURE 13.1
Queuing Costs and Service Levels



Total expected cost is the sum of service plus waiting costs.

One means of evaluating a service facility is thus to look at a *total expected cost*, a concept illustrated in Figure 13.1. Total expected cost is the sum of expected *service costs* plus expected **waiting costs**.

Service costs are seen to increase as a firm attempts to raise its level of service. For example, if three teams of stevedores, instead of two, are employed to unload a cargo ship, service costs are increased by the additional price of wages. As service improves in speed, however, the cost of time spent waiting in lines decreases. This waiting cost may reflect lost productivity of workers while their tools or machines are awaiting repairs or may simply be an estimate of the costs of customers lost because of poor service and long queues.

Three Rivers Shipping Company Example

As an illustration, let's look at the case of the Three Rivers Shipping Company. Three Rivers runs a huge docking facility located on the Ohio River near Pittsburgh. Approximately five ships arrive to unload their cargoes of steel and ore during every 12-hour work shift. Each hour that a ship sits idle in line waiting to be unloaded costs the firm a great deal of money, about $1,000 per hour. From experience, management estimates that if one team of stevedores is on duty to handle the unloading work, each ship will wait an average of 7 hours to be unloaded. If two teams are working, the average waiting time drops to 4 hours; for three teams, it's 3 hours; and for four teams of stevedores, only 2 hours. But each additional team of stevedores is also an expensive proposition, due to union contracts.

The goal is to find the service level that minimizes total expected cost.

Three Rivers's superintendent would like to determine the optimal number of teams of stevedores to have on duty each shift. The objective is to minimize total expected costs. This analysis is summarized in Table 13.1. To minimize the sum of service costs and waiting costs, the firm makes the decision to employ two teams of stevedores each shift.

13.3 Characteristics of a Queuing System

In this section we take a look at the three parts of a queuing system: (1) the arrivals or inputs to the system (sometimes referred to as the **calling population**), (2) the queue or the waiting line itself, and (3) the service facility. These three components have certain characteristics that must be examined before mathematical queuing models can be developed.

Arrival Characteristics

The input source that generates arrivals or customers for the service system has three major characteristics. It is important to consider the *size* of the calling population, the *pattern* of arrivals at the queuing system, and the *behavior* of the arrivals.

TABLE 13.1 Three Rivers Shipping Company Waiting Line Cost Analysis

	NUMBER OF TEAMS OF STEVEDORES WORKING			
	1	2	3	4
(a) Average number of ships arriving per shift	5	5	5	5
(b) Average time each ship waits to be unloaded (hours)	7	4	3	2
(c) Total ship hours lost per shift ($a \times b$)	35	20	15	10
(d) Estimated cost per hour of idle ship time	$1,000	$1,000	$1,000	$1,000
(e) Value of ship's lost time or waiting cost ($c \times d$)	$35,000	$20,000	$15,000	$10,000
(f) Stevedore team salary,* or service cost	$6,000	$12,000	$18,000	$24,000
(g) Total expected cost ($e + f$)	$41,000	$32,000 Optimal cost	$33,000	$34,000

*Stevedore team salaries are computed as the number of people in a typical team (assumed to be 50), times the number of hours each person works per day (12 hours), times an hourly salary of $10 per hour. If two teams are employed, the rate is just doubled.

Unlimited (or infinite) calling populations are assumed for most queuing models.

SIZE OF THE CALLING POPULATION Population sizes are considered to be either **unlimited** (essentially *infinite*) or **limited** (*finite*). When the number of customers or arrivals on hand at any given moment is just a small portion of potential arrivals, the calling population is considered unlimited. For practical purposes, examples of unlimited populations include cars arriving at a highway tollbooth, shoppers arriving at a supermarket, or students arriving to register for classes at a large university. Most queuing models assume such an infinite calling population. When this is not the case, modeling becomes much more complex. An example of a finite population is a shop with only eight machines that might break down and require service.

Arrivals are random when they are independent of one another and cannot be predicted exactly.

PATTERN OF ARRIVALS AT THE SYSTEM Customers either arrive at a service facility according to some known schedule (for example, one patient every 15 minutes or one student for advising every half hour) or else they arrive *randomly*. Arrivals are considered random when they are independent of one another and their occurrence cannot be predicted exactly. Frequently in queuing problems, the number of arrivals per unit of time can be estimated by a probability distribution known as the **Poisson distribution**. See Section 2.14 for details about this distribution.

The concepts of balking and reneging.

BEHAVIOR OF THE ARRIVALS Most queuing models assume that an arriving customer is a patient customer. Patient customers are people or machines that wait in the queue until they are served and do not switch between lines. Unfortunately, life and quantitative analysis are complicated by the fact that people have been known to balk or renege. **Balking** refers to customers who refuse to join the waiting line because it is too long to suit their needs or interests. **Reneging** customers are those who enter the queue but then become impatient and leave without completing their transaction. Actually, both of these situations just serve to accentuate the need for queuing theory and waiting line analysis. How many times have you seen a shopper with a basket full of groceries, including perishables such as milk, frozen food, or meats, simply abandon the shopping cart before checking out because the line was too long? This expensive occurrence for the store makes managers acutely aware of the importance of service-level decisions.

Waiting Line Characteristics

The models in this chapter assume unlimited queue length.

The waiting line itself is the second component of a queuing system. The length of a line can be either *limited* or *unlimited*. A queue is limited when it cannot, by law of physical restrictions, increase to an infinite length. This may be the case in a small restaurant that has only 10 tables and can serve no more than 50 diners an evening. Analytic queuing models are treated in this chapter under an assumption of **unlimited queue length**. A queue is unlimited when its size is unrestricted, as in the case of the tollbooth serving arriving automobiles.

A second waiting line characteristic deals with **queue discipline**. This refers to the rule by which customers in the line are to receive service. Most systems use a queue discipline known as

Most queuing models use the FIFO rule. This is obviously not appropriate in all *service systems, especially those dealing with emergencies.*

the **first-in, first-out** (FIFO) rule. In a hospital emergency room or an express checkout line at a supermarket, however, various assigned priorities may preempt FIFO. Patients who are critically injured will move ahead in treatment priority over patients with broken fingers or noses. Shoppers with fewer than 10 items may be allowed to enter the express checkout queue but are *then* treated as first come, first served. Computer programming runs are another example of queuing systems that operate under priority scheduling. In most large companies, when computer-produced paychecks are due out on a specific date, the payroll program has highest priority over other runs.[*]

Service Facility Characteristics

The third part of any queuing system is the service facility. It is important to examine two basic properties: (1) the configuration of the service system and (2) the pattern of service times.

Number of service channels in a queuing system is the number of servers.

BASIC QUEUING SYSTEM CONFIGURATIONS Service systems are usually classified in terms of their number of channels, or number of servers, and number of phases, or number of service stops, that must be made. A **single-channel system**, with one server, is typified by the drive-in bank that has only one open teller, or by the type of drive-through fast-food restaurant that has become so popular in the United States. If, on the other hand, the bank had several tellers on duty and each customer waited in one common line for the first available teller, we would have a **multichannel system** at work. Many banks today are multichannel service systems, as are most large barber shops and many airline ticket counters.

Single-phase *means the customer receives service at only one station before leaving the system.* **Multiphase** *implies two or more stops before leaving the system.*

A **single-phase system** is one in which the customer receives service from only one station and then exits the system. A fast-food restaurant in which the person who takes your order also brings you the food and takes your money is a single-phase system. So is a driver's license agency in which the person taking your application also grades your test and collects the license fee. But if the restaurant requires you to place your order at one station, pay at a second, and pick up the food at a third service stop, it becomes a **multiphase system**. Similarly, if the driver's license agency is large or busy, you will probably have to wait in a line to complete the application (the first service stop), then queue again to have the test graded (the second service stop), and finally go to a third service counter to pay the fee. To help you relate the concepts of channels and phases, Figure 13.2 presents four possible configurations.

SERVICE TIME DISTRIBUTION Service patterns are like arrival patterns in that they can be either constant or random. If service time is constant, it takes the same amount of time to take care of each customer. This is the case in a machine-performed service operation such as an automatic car wash. More often, service times are randomly distributed. In many cases it can be assumed that random service times are described by the **negative exponential probability distribution**. See Section 2.13 for details about this distribution.

Service times often follow the negative exponential distribution.

It is important to confirm that the queuing assumptions of Poisson arrivals and exponential services are valid before applying the model.

The exponential distribution is important to the process of building mathematical queuing models because many of the models' theoretical underpinnings are based on the assumption of Poisson arrivals and exponential services. Before they are applied, however, the quantitative analyst can and should observe, collect, and plot service time data to determine if they fit the exponential distribution.

Identifying Models Using Kendall Notation

D. G. Kendall developed a notation that has been widely accepted for specifying the pattern of arrivals, the service time distribution, and the number of channels in a queuing model. This notation is often seen in software for queuing models. The basic three-symbol **Kendall notation** is in the form

Arrival distribution/Service time distribution/Number of service channels open

where specific letters are used to represent probability distributions. The following letters are commonly used in Kendall notation:

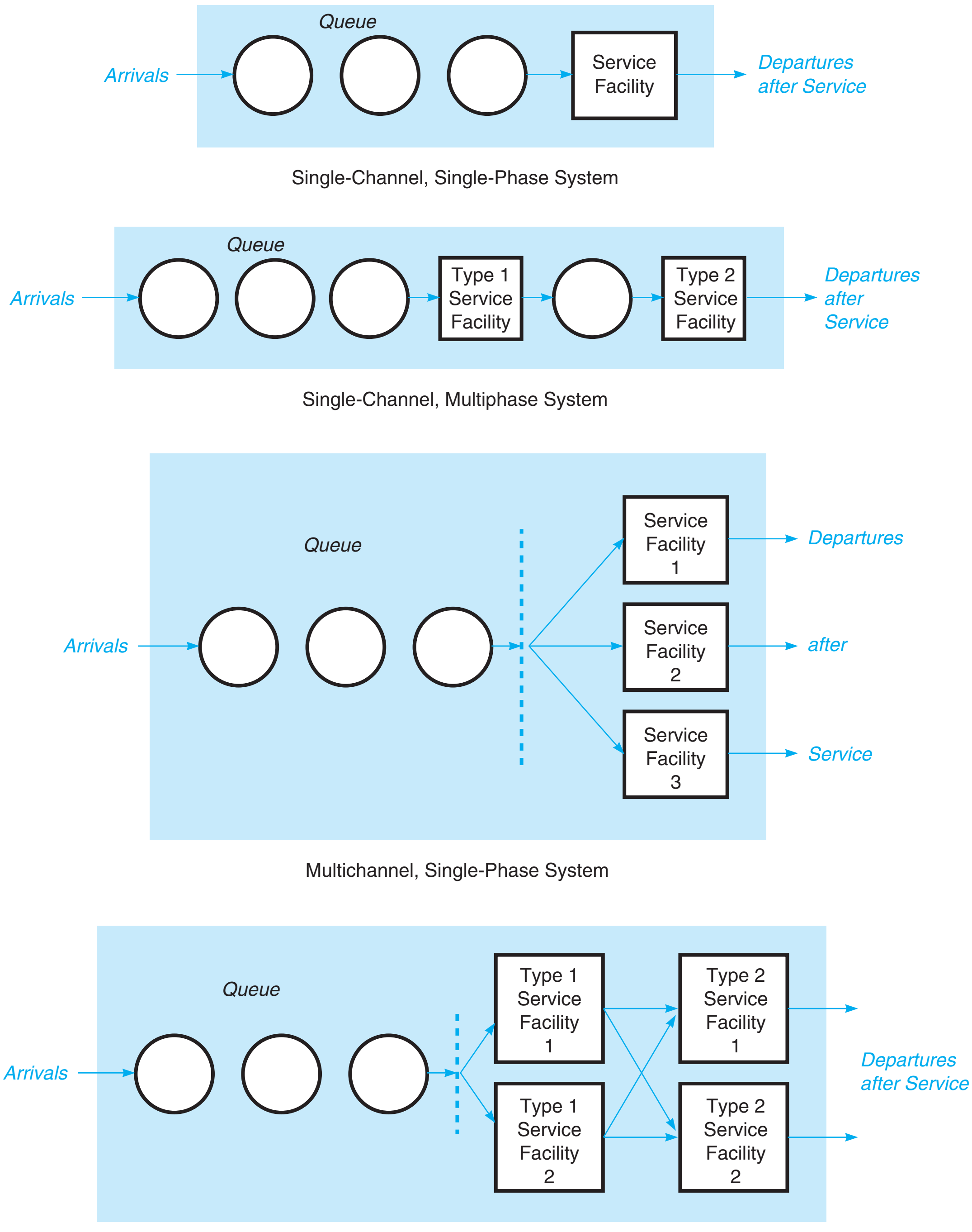
M = Poisson distribution for number of occurrences (or exponential times)

D = constant (deterministic) rate

G = general distribution with mean and variance known

[*]The term *FIFS* (*f*irst *i*n, *f*irst *s*erved) is often used in place of FIFO. Another discipline, LIFS (*l*ast *i*n, *f*irst *s*erved), is common when material is stacked or piled and the items on top are used first.

FIGURE 13.2 Four Basic Queuing System Configurations



Thus, a single channel model with Poisson arrivals and exponential service times would be represented by

$$M/M/1$$

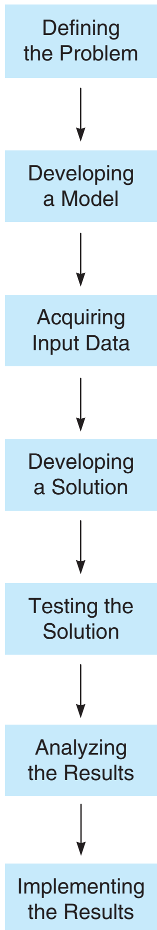
An M/M/2 model has Poisson arrivals, exponential service times, and two channels.

When a second channel is added, we would have

$$M/M/2$$

If there are m distinct service channels in the queuing system with Poisson arrivals and exponential service times, the Kendall notation would be $M/M/m$. A three-channel system with Poisson arrivals and constant service time would be identified as $M/D/3$. A four-channel system with Poisson arrivals and service times that are normally distributed would be identified as $M/G/4$.

MODELING IN THE REAL WORLD — Montgomery County's Public Health Services



Defining the Problem

In 2002, the Centers for Disease Control and Prevention began to require that public health departments create plans for smallpox vaccinations. A county must be prepared to vaccinate every person in an infected area in a few days. This was of particular concern after the terrorist attack of September 11, 2001.

Developing a Model

Queuing models for capacity planning and discrete event simulation models were developed through a joint effort of the Montgomery County (Maryland) Public Health Service and the University of Maryland, College Park.

Acquiring Input Data

Data were collected on the time required for the vaccinations to occur or for medicine to be dispensed. Clinical exercises were used for this purpose.

Developing a Solution

The models indicate the number of staff members needed to achieve specific capacities to avoid congestion in the clinics.

Testing the Solution

The smallpox vaccination plan was tested using a simulation of a mock vaccination clinic in a full-scale exercise involving residents going through the clinic. This highlighted the need for modifications in a few areas. A computer simulation model was then developed for additional testing.

Analyzing the Results

The results of the capacity planning and queuing model provide very good estimates of the true performance of the system. Clinic planners and managers can quickly estimate capacity and congestion as the situation develops.

Implementing the Results

The results of this study are available on a web site for public health professionals to download and use. The guidelines include suggestions for workstation operations. Improvements to the process are continuing.

Source: Based on Kay Aaby, et al. "Montgomery County's Public Health Service Operations Research to Plan Emergency Mass Dispensing and Vaccination Clinics," *Interfaces* 36, 6 (November–December 2006): 569–579.

IN ACTION — Slow Down the Ski Lift to Get Shorter Lines

At a small, five-lift ski resort, management was worried that lift lines were getting too long. While this is sometimes a nice problem to have because it means business is booming, it is a problem that can backfire. If word spreads that a ski resort's lift lines are too long, customers may choose to ski elsewhere, where lines are shorter.

Because building new ski lifts requires significant financial investment, management decided to hire an external consultant with queuing system experience to study the problem. After several weeks of investigating the problem, collecting data, and measuring the length of the lift lines at various times, the consultant presented recommendations to the ski resort's management.

Surprisingly, instead of building new ski lifts, the consultant proposed that management slow down its five ski lifts at the resort to half their current speed and to double the number of chairs on each of the lifts. This meant, for example, that instead of 40 feet between lift chairs, there would be only 20 feet between lift chairs, but because they were moving more slowly, there was still the same amount of time for customers to board the lift. So if a particular lift previously took 4 minutes to get to the top, it would now take 8 minutes. It was reasoned that skiers wouldn't notice the difference in time because they were on the lifts and enjoying the view on the way to the top; this proved to be a valid assumption. Moreover, at any given time, twice as many people could actually be on the ski lifts, and fewer people were in the lift lines. The problem was solved!

Source: Anonymous ski resort and golf course consulting company, private communication, 2009.

There is a more detailed notation with additional terms that indicate the maximum number in the system and the population size. When these are omitted, it is assumed there is no limit to the queue length or the population size. Most of the models we study here will have those properties.

13.4 Single-Channel Queuing Model with Poisson Arrivals and Exponential Service Times ($M/M/1$)

In this section we present an analytical approach to determine important measures of performance in a typical service system. After these numeric measures have been computed, it will be possible to add in cost data and begin to make decisions that balance desirable service levels with waiting line service costs.

Assumptions of the Model

These seven assumptions must be met if the single-channel, single-phase model is to be applied.

The single-channel, single-phase model considered here is one of the most widely used and simplest queuing models. It involves assuming that seven conditions exist:

1. Arrivals are served on a FIFO basis.
2. Every arrival waits to be served regardless of the length of the line; that is, there is no balking or reneging.
3. Arrivals are independent of preceding arrivals, but the average number of arrivals (the arrival rate) does not change over time.
4. Arrivals are described by a Poisson probability distribution and come from an infinite or very large population.
5. Service times also vary from one customer to the next and are independent of one another, but their average rate is known.
6. Service times occur according to the negative exponential probability distribution.
7. The average service rate is greater than the average arrival rate.

When these seven conditions are met, we can develop a series of equations that define the queue's **operating characteristics**. The mathematics used to derive each equation is rather complex and outside the scope of this book, so we will just present the resulting formulas here.

Queuing Equations

We let

$$\lambda = \text{mean number of arrivals per time period (for example, per hour)}$$
$$\mu = \text{mean number of people or items served per time period}$$

When determining the arrival rate (λ) and the service rate (μ), the same time period must be used. For example, if λ is the average number of arrivals per hour, then μ must indicate the average number that could be served per hour.

The queuing equations follow.

These seven queuing equations for the single-channel, single-phase model describe the important operating characteristics of the service system.

1. The average number of customers or units in the system, L, that is, the number in line plus the number being served:

$$L = \frac{\lambda}{\mu - \lambda} \qquad \textbf{(13-1)}$$

2. The average time a customer spends in the system, W, that is, the time spent in line plus the time spent being served:

$$W = \frac{1}{\mu - \lambda} \qquad \textbf{(13-2)}$$

3. The average number of customers in the queue, L_q:

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{13-3}$$

4. The average time a customer spends waiting in the queue, W_q:

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} \tag{13-4}$$

5. The **utilization factor** for the system, ρ (the Greek lowercase letter rho), that is, the probability that the service facility is being used:

$$\rho = \frac{\lambda}{\mu} \tag{13-5}$$

6. The percent idle time, P_0, that is, the probability that no one is in the system:

$$P_0 = 1 - \frac{\lambda}{\mu} \tag{13-6}$$

7. The probability that the number of customers in the system is greater than k, $P_{n>k}$:

$$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1} \tag{13-7}$$

Arnold's Muffler Shop Case

We now apply these formulas to the case of Arnold's Muffler Shop in New Orleans. Arnold's mechanic, Reid Blank, is able to install new mufflers at an average rate of 3 per hour, or about 1 every 20 minutes. Customers needing this service arrive at the shop on the average of 2 per hour. Larry Arnold, the shop owner, studied queuing models in an MBA program and feels that all seven of the conditions for a single-channel model are met. He proceeds to calculate the numerical values of the preceding operating characteristics:

$$\lambda = 2 \text{ cars arriving per hour}$$

$$\mu = 3 \text{ cars serviced per hour}$$

$$L = \frac{\lambda}{\mu - \lambda} = \frac{2}{3 - 2} = \frac{2}{1} = 2 \text{ cars in the system on the average}$$

$$W = \frac{1}{\mu - \lambda} = \frac{1}{3 - 2} = 1 \text{ hour that an average car spends in the system}$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{2^2}{3(3 - 2)} = \frac{4}{3(1)} = \frac{4}{3} = 1.33 \text{ cars waiting in line on the average}$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{2}{3(3 - 2)} = \frac{2}{3} \text{ hour} = 40 \text{ minutes} = \text{average waiting time per car}$$

Note that W and W_q are in *hours*, since λ was defined as the number of arrivals per *hour*.

$$\rho = \frac{\lambda}{\mu} = \frac{2}{3} = 0.67 = \text{percentage of time mechanic is busy, or the probability that the server is busy}$$

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{2}{3} = 0.33 = \text{probability that there are 0 cars in the system}$$

Probability of More than k Cars in the System

k	$P_{n>k} = \left(\frac{2}{3}\right)^{k+1}$	
0	0.667	*Note that this is equal to $1 - P_0 = 1 - 0.33 = 0.667$.*
1	0.444	
2	0.296	
3	0.198	*Implies that there is a 19.8% chance that more than 3 cars are in the system.*
4	0.132	
5	0.088	
6	0.058	
7	0.039	

USING EXCEL QM ON THE ARNOLD'S MUFFLER SHOP QUEUE To use Excel QM for this problem, from the Excel QM menu, select *Waiting Lines - Single Channel (M/M/1)*. When the spreadsheet appears, enter the arrival rate (2) and service rate (3). All the operating characteristic will automatically be computed, as demonstrated in Program 13.1.

INTRODUCING COSTS INTO THE MODEL Now that the characteristics of the queuing system have been computed, Arnold decides to do an economic analysis of their impact. The waiting line

PROGRAM 13.1 Excel QM Solution to Arnold's Muffler Example

	A	B	C	D	E
1	Arnold's Muffler Shop				
2					
3	Waiting Lines		M/M/1 (Single Server Model)		
4	The arrival RATE and service RATE both must be rates and use the same time unit. Given a time such as 10 minutes, convert it to a rate such as 6 per hour.				
5					
6	Data			Results	
7	Arrival rate (λ)	2		Average server utilization(ρ)	0.6666667
8	Service rate (μ)	3		Average number of customers in the queue(L_q)	1.3333333
9				Average number of customers in the system(L_s)	2
10				Average waiting time in the queue(W_q)	0.6666667
11				Average time in the system(W_s)	1
12				Probability (% of time) system is empty (P_0)	0.3333333
13					
14					
15	Probabilities				
16	Number	Probabilit y	Cumulative Probability		
17	0	0.333333	0.333333		
18	1	0.222222	0.555556		
19	2	0.148148	0.703704		
20	3	0.098765	0.802469		
21	4	0.065844	0.868313		
22	5	0.043896	0.912209		

Conducting an economic analysis is the next step. It permits cost factors to be included.

model was valuable in predicting potential waiting times, queue lengths, idle times, and so on. But it did not identify optimal decisions or consider cost factors. As stated earlier, the solution to a queuing problem may require management to make a trade-off between the increased cost of providing better service and the decreased waiting costs derived from providing that service. These two costs are called the waiting cost and the service cost.

The total service cost is

$$\text{Total service cost} = (\text{Number of channels})(\text{Cost per channel})$$
$$\text{Total service cost} = mC_s \tag{13-8}$$

where

m = number of channels

C_s = service cost (labor cost) of each channel

The waiting cost when the waiting time cost is based on time in the system is

$$\begin{aligned}\text{Total waiting cost} &= (\text{Total time spent waiting by all arrivals})(\text{Cost of waiting})\\ &= (\text{Number of arrivals})(\text{Average wait per arrival})C_w\end{aligned}$$

so,

$$\text{Total waiting cost} = (\lambda W)C_w \tag{13-9}$$

If the waiting time cost is based on time in the queue, this becomes

$$\text{Total waiting cost} = (\lambda W_q)C_w \tag{13-10}$$

These costs are based on whatever time units (often hours) are used in determining λ. Adding the total service cost to the total waiting cost, we have the total cost of the queuing system. When the waiting cost is based on the time in the system, this is

$$\text{Total cost} = \text{Total service cost} + \text{Total waiting cost}$$
$$\text{Total cost} = mC_s + \lambda W C_w \tag{13-11}$$

When the waiting cost is based on time in the queue, the total cost is

$$\text{Total cost} = mC_s + \lambda W_q C_w \tag{13-12}$$

At times we may wish to determine the daily cost, and then we simply find the total number of arrivals per day. Let us consider the situation for Arnold's Muffler Shop.

Customer waiting time is often considered the most important factor.

Arnold estimates that the cost of customer waiting time, in terms of customer dissatisfaction and lost goodwill, is \$50 per hour of time spent *waiting* in line. (After customers' cars are actually being serviced on the rack, customers don't seem to mind waiting.) Because on the average a car has a $^2\!/_3$ hour wait and there are approximately 16 cars serviced per day (2 per hour times 8 working hours per day), the total number of hours that customers spend waiting for mufflers to be installed each day is $^2\!/_3 \times 16 = {}^{32}\!/_3$, or $10\,^2\!/_3$ hours. Hence, in this case,

$$\text{Total daily waiting cost} = (8 \text{ hours per day})\lambda W_q C_w = (8)(2)\left(^2\!/_3\right)(\$50) = \$533.33$$

The only other cost that Larry Arnold can identify in this queuing situation is the pay rate of Reid Blank, the mechanic. Blank is paid \$15 per hour:

$$\text{Total daily service cost} = (8 \text{ hours per day})mC_s = 8(1)(\$15) = \$120$$

Waiting cost plus service cost equal total cost.

The total daily cost of the system as it is currently configured is the total of the waiting cost and the service cost, which gives us

$$\text{Total daily cost of the queuing system} = \$533.33 + \$120 = \$653.33$$

Now comes a decision. Arnold finds out through the muffler business grapevine that the Rusty Muffler, a cross-town competitor, employs a mechanic named Jimmy Smith who can efficiently install new mufflers at the rate of 4 per hour. Larry Arnold contacts Smith and inquires as to his interest in switching employers. Smith says that he would consider leaving the Rusty Muffler but only if he were paid a \$20 per hour salary. Arnold, being a crafty businessman, decides that it may be worthwhile to fire Blank and replace him with the speedier but more expensive Smith.

He first recomputes all the operating characteristics using a new service rate of 4 mufflers per hour:

$$\lambda = 2 \text{ cars arriving per hour}$$

$$\mu = 4 \text{ cars serviced per hour}$$

$$L = \frac{\lambda}{\mu - \lambda} = \frac{2}{4 - 2} = 1 \text{ car in the system on the average}$$

$$W = \frac{1}{\mu - \lambda} = \frac{1}{4 - 2} = \frac{1}{2} \text{ hour in the system on the average}$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{2^2}{4(4 - 2)} = \frac{4}{8} = \frac{1}{2} \text{ cars waiting in line on the average}$$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{2}{4(4 - 2)} = \frac{2}{8} = \frac{1}{4} \text{ hour} = 15 \text{ minutes average waiting time per car in the queue}$$

$$\rho = \frac{\lambda}{\mu} = \frac{2}{4} = 0.5 = \text{percentage of time mechanic is busy}$$

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - 0.5 = 0.5 = \text{probability that there are 0 cars in the system}$$

Probability of More Than k Cars in the System

k	$P_{n>k} = \left(2/4\right)^{k+1}$
0	0.500
1	0.250
2	0.125
3	0.062
4	0.031
5	0.016
6	0.008
7	0.004

It is quite evident that Smith's speed will result in considerably shorter queues and waiting times. For example, a customer would now spend an average of $1/2$ hour in the system and $1/4$ hour waiting in the queue, as opposed to 1 hour in the system and $2/3$ hour in the queue with Blank as mechanic. The total daily waiting time cost with Smith as the mechanic will be

$$\text{Total daily waiting cost} = (8 \text{ hours per day})\lambda W_q C_w = (8)(2)\left(\frac{1}{4}\right)(\$50) = \$200 \text{ per day}$$

Notice that the total time spent waiting for the 16 customers per day is now

$$(16 \text{ cars per day}) \times \left(1/4 \text{ hour per car}\right) = 4 \text{ hours}$$

instead of 10.67 hours with Blank. Thus, the waiting is much less than half of what it was, even though the service rate only changed from 3 per hour to 4 per hour.

Here is a comparison for total costs using the two different mechanics.

The service cost will go up due to the higher salary, but the overall cost will decrease, as we see here:

$$\text{Service cost of Smith} = 8 \text{ hours/day} \times \$20/\text{hour} = \$160 \text{ per day}$$

$$\begin{aligned}\text{Total expected cost} &= \text{Waiting cost} + \text{Service cost} = \$200 + \$160 \\ &= \$360 \text{ per day}\end{aligned}$$

Because the total daily expected cost with Blank as mechanic was \$653.33, Arnold may very well decide to hire Smith and reduce costs by \$653.33 − \$360 = \$293.33 per day.

IN ACTION Ambulances in Chile Evaluate and Improve Performance Metrics Through Queuing Theory

Researchers in Chile turned to queuing theory to evaluate and improve ambulance services. They investigated the key performance indicators that are of value to an ambulance company manager (e.g., number of ambulances deployed, ambulance base locations, operating costs) and the key performance indicators that are of value to a customer (e.g., waiting time until service, queue prioritization system).

The somewhat surprising result that the researchers brought to light was that seemingly simple metrics such as waiting time in queue, W_q, were of the most value for ambulance operations. Reducing wait times saves the ambulance company money in gasoline and transportation costs, and, more importantly, it can save lives for its customers. In other words, sometimes simpler is better.

Source: Based on Marcos Singer and Patricio Donoso. "Assessing an Ambulance Service with Queuing Theory," *Computers & Operations Research* 35, 8 (2008): 2549–2560.

Enhancing the Queuing Environment

Although reducing the waiting time is an important factor in reducing the waiting time cost of a queuing system, a manager might find other ways to reduce this cost. The total waiting time cost is based on the total amount of time spent waiting (based on W or W_q) and the cost of waiting (C_w). Reducing either of these will reduce the overall cost of waiting. Enhancing the queuing environment by making the wait less unpleasant may reduce C_w as customers will not be as upset by having to wait. There are magazines in the waiting room of doctors' offices for patients to read while waiting. There are tabloids on display by the checkout lines in grocery stores, and customers read the headlines to pass time while waiting. Music is often played while telephone callers are placed on hold. At major amusement parks there are video screens and televisions in some of the queue lines to make the wait more interesting. For some of these, the waiting line is so entertaining that it is almost an attraction itself.

All of these things are designed to keep the customer busy and to enhance the conditions surrounding the waiting so that it appears that time is passing more quickly than it actually is. Consequently, the cost of waiting (C_w) becomes lower and the total cost of the queuing system is reduced. Sometimes, reducing the total cost in this way is easier than reducing the total cost by lowering W or W_q. In the case of Arnold's Muffler Shop, Arnold might consider putting a television in the waiting room and remodeling this room so customers feel more comfortable while waiting for their cars to be serviced.

13.5 Multichannel Queuing Model with Poisson Arrivals and Exponential Service Times (*M*/*M*/m)

The next logical step is to look at a *multichannel queuing system*, in which two or more servers or channels are available to handle arriving customers. Let us still assume that customers awaiting service form one single line and then proceed to the first available server. An example of such a multichannel, single-phase waiting line is found in many banks today. A common line is formed and the customer at the head of the line proceeds to the first free teller (Refer to Figure 13.2 for a typical multichannel configuration.)

The multichannel model also assumes Poisson arrivals and exponential service times.

The multiple-channel system presented here again assumes that arrivals follow a Poisson probability distribution and that service times are distributed exponentially. Service is first come, first served, and all servers are assumed to perform at the same rate. Other assumptions listed earlier for the single-channel model apply as well.

Equations for the Multichannel Queuing Model

If we let

m = number of channels open,

λ = average arrival rate, and

μ = average service rate at each channel

the following formulas may be used in the waiting line analysis:

1. The probability that there are zero customers or units in the system:

$$P_0 = \frac{1}{\left[\displaystyle\sum_{n=0}^{n=m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{m\mu}{m\mu - \lambda}} \text{ for } m\mu > \lambda \qquad \textbf{(13-13)}$$

2. The average number of customers or units in the system:

$$L = \frac{\lambda\mu(\lambda/\mu)^m}{(m-1)!(m\mu - \lambda)^2}P_0 + \frac{\lambda}{\mu} \qquad \textbf{(13-14)}$$

3. The average time a unit spends in the waiting line or being serviced (namely, in the system):

$$W = \frac{\mu(\lambda/\mu)^m}{(m-1)!(m\mu - \lambda)^2}P_0 + \frac{1}{\mu} = \frac{L}{\lambda} \qquad \textbf{(13-15)}$$

4. The average number of customers or units in line waiting for service:

$$L_q = L - \frac{\lambda}{\mu} \qquad \textbf{(13-16)}$$

5. The average time a customer or unit spends in the queue waiting for service:

$$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda} \qquad \textbf{(13-17)}$$

6. Utilization rate:

$$\rho = \frac{\lambda}{m\mu} \qquad \textbf{(13-18)}$$

These equations are obviously more complex than the ones used in the single-channel model, yet they are used in exactly the same fashion and provide the same type of information as did the simpler model.

Arnold's Muffler Shop Revisited

For an application of the multichannel queuing model, let's return to the case of Arnold's Muffler Shop. Earlier, Larry Arnold examined two options. He could retain his current mechanic, Reid Blank, at a total expected cost of \$653 per day; or he could fire Blank and hire a slightly more expensive but faster worker named Jimmy Smith. With Smith on board, service system costs could be reduced to \$360 per day.

The muffler shop considers opening a second muffler service channel that operates at the same speed as the first one.

A third option is now explored. Arnold finds that at minimal after-tax cost he can open a *second* garage bay in which mufflers can be installed. Instead of firing his first mechanic, Blank, he would hire a second worker. The new mechanic would be expected to install mufflers at the same rate as Blank—about $\mu = 3$ per hour. Customers, who would still arrive at the rate of $\lambda = 2$ per hour, would wait in a single line until one of the two mechanics is free. To find out how this option compares with the old single-channel waiting line system, Arnold computes several operating characteristics for the $m = 2$ channel system:

$$P_0 = \frac{1}{\left[\sum_{n=0}^{1} \frac{1}{n!}\left(\frac{2}{3}\right)^n\right] + \frac{1}{2!}\left(\frac{2}{3}\right)^2\left(\frac{2(3)}{2(3) - 2}\right)}$$

$$= \frac{1}{1 + \frac{2}{3} + \frac{1}{2}\left(\frac{4}{9}\right)\left(\frac{6}{6 - 2}\right)} = \frac{1}{1 + \frac{2}{3} + \frac{1}{3}} = \frac{1}{2} = 0.5$$

$= \text{probability of 0 cars in the system}$

$$L = \left(\frac{(2)(3)(2/3)^2}{1![2(3) - 2]^2}\right)\left(\frac{1}{2}\right) + \frac{2}{3} = \frac{8/3}{16}\left(\frac{1}{2}\right) + \frac{2}{3} = \frac{3}{4} = 0.75$$

$= \text{average number of cars in the system}$

$$W = \frac{L}{\lambda} = \frac{3/4}{2} = \frac{3}{8}\text{ hours} = 22\tfrac{1}{2}\text{ minutes}$$

$= \text{average time a car spends in the system}$

$$L_q = L - \frac{\lambda}{\mu} = \frac{3}{4} - \frac{2}{3} = \frac{1}{12} = 0.083$$

$= \text{average number of cars in the queue}$

$$W_q = \frac{L_q}{\lambda} = \frac{0.083}{2} = 0.0415\text{ hour} = 2\tfrac{1}{2}\text{ minutes}$$

$= \text{average time a car spends in the queue}$

Dramatically lower waiting time results from opening the second service bay.

These data are compared with earlier operating characteristics in Table 13.2. The increased service from opening a second channel has a dramatic effect on almost all characteristics. In particular, time spent waiting in line drops from 40 minutes with one mechanic (Blank) or 15 minutes with Smith down to only $2\tfrac{1}{2}$ minutes! Similarly, the average number of cars in the queue falls to 0.083 (about $1/12$ of a car).* But does this mean that a second bay should be opened?

To complete his economic analysis, Arnold assumes that the second mechanic would be paid the same as the current one, Blank, namely, \$15 per hour. The daily waiting cost now will be

$$\text{Total daily waiting cost} = (8\text{ hours per day})\lambda W_q C_w = (8)(2)(0.0415)(\$50) = \$33.20$$

TABLE 13.2 Effect of Service Level on Arnold's Operating Characteristics

	LEVEL OF SERVICE		
OPERATING CHARACTERISTIC	ONE MECHANIC (REID BLANK) $\mu = 3$	TWO MECHANICS $\mu = 3$ FOR EACH	ONE FAST MECHANIC (JIMMY SMITH) $\mu = 4$
Probability that the system is empty (P_0)	0.33	0.50	0.50
Average number of cars in the system (L)	2 cars	0.75 car	1 car
Average time spent in the system (W)	60 minutes	22.5 minutes	30 minutes
Average number of cars in the queue (L_q)	1.33 cars	0.083 car	0.50 car
Average time spent in the queue (W_q)	40 minutes	2.5 minutes	15 minutes

*You might note that adding a second mechanic does not cut queue waiting time and length just in half, but makes it even smaller. This is because of the *random* arrival and service processes. When there is only one mechanic and two customers arrive within a minute of each other, the second will have a long wait. The fact that the mechanic may have been idle for 30 to 40 minutes before they both arrive does not change this average waiting time. Thus, single-channel models often have high wait times relative to multichannel models.

PROGRAM 13.2
Excel QM Solution to Arnold's Muffler Multichannel Example

	A	B	C	D	E
1	**Arnold's Muffler Shop Multichannel**				
2					
3	**Waiting Lines**		**M/M/s**		
4	The arrival RATE and service RATE both must be rates and use the same time unit. Given a				
5					
6	**Data**			**Results**	
7	Arrival rate (λ)	2		Average server utilization(ρ)	0.33333
8	Service rate (μ)	3		Average number of customers in the queue	0.08333
9	Number of servers(s)	2		Average number of customers in the system	0.75
10				Average waiting time in the queue(W_q)	0.04167
11				Average time in the system(W_s)	0.375
12	Enter the arrival rate, service rate, and number of servers (channels).			Probability (% of time) system is empty (P_0)	0.5
13					
14	**Probabilities**				
15	Number	Probability	Cumulative Probability		
16	0	0.500000	0.500000		
17	1	0.333333	0.833333		
18	2	0.111111	0.944444		
19	3	0.037037	0.981481		

Notice that the total waiting time for the 16 cars per day is (16 cars/day) × (0.0415 hour/car) = 0.664 hours per day instead of the 10.67 hours with only one mechanic.

The service cost is doubled, as there are two mechanics, so this is

$$\text{Total daily service cost} = (8 \text{ hours per day})mC_s = (8)2(\$15) = \$240$$

The total daily cost of the system as it is currently configured is the total of the waiting cost and the service cost, which is

$$\text{Total daily cost of the queuing system} = \$33.20 + \$240 = \$273.20$$

As you recall, total cost with just Blank as mechanic was found to be $653 per day. Cost with just Smith was just $360. Opening a second service bay will save about $380 per day compared to the current system, and it will save about $87 per day compared to the system with the faster mechanic. Thus, because the after-tax cost of a second bay is very low, Arnold's decision is to open a second service bay and hire a second worker who is paid the same as Blank. This may have additional benefits because word may spread about the very short waits at Arnold's Muffler Shop, and this may increase the number of customers who choose to use Arnold's.

USING EXCEL QM FOR ANALYSIS OF ARNOLD'S MULTICHANNEL QUEUING MODEL To use Excel QM for this problem, from the Excel QM menu, select *Waiting Lines - Multiple Channel Model (M/M/s)*. When the spreadsheet appears, enter the arrival rate (2), service rate (3), and number of servers (2). Once these are entered, the solution shown in Program 13.2 will be displayed.

13.6 Constant Service Time Model ($M/D/1$)

Constant service rates speed the process compared to exponentially distributed service times with the same value of μ.

Some service systems have constant service times instead of exponentially distributed times. When customers or equipment are processed according to a fixed cycle, as in the case of an automatic car wash or an amusement park ride, constant service rates are appropriate. Because constant rates are certain, the values for L_q, W_q, L, and W are always less than they would be in the models we have just discussed, which have variable service times. As a matter of fact, both the average queue length and the average waiting time in the queue are *halved* with the constant service rate model.

IN ACTION Queuing at the Polls

The long lines at the polls in recent presidential elections have caused some concern. In 2000, some voters in Florida waited in line over 2 hours to cast their ballots. The final tally favored the winner by 527 votes of the almost 6 million votes cast in that state. Some voters growing tired of waiting and simply leaving the line without voting may have affected the outcome. A change of 0.01% could have caused different results. In 2004, there were reports of voters in Ohio waiting in line 10 hours to vote. If as few as 19 potential voters at each precinct in the 12 largest counties in Ohio got tired of waiting and left without voting, the outcome of the election may have been different. There were obvious problems with the number of machines available in some of the precincts. One precinct needed 13 machines but had only 2. Inexplicably, there were 68 voting machines in warehouses that were not even used. Other states had some long lines as well.

The basic problem stems from not having enough voting machines in many of the precincts, although some precincts had sufficient machines and no problems. Why was there such a problem in some of the voting precincts? Part of the reason is related to poor forecasting of the voter turnout in various precincts. Whatever the cause, the misallocation of voting machines among the precincts seems to be a major cause of the long lines in the national elections. Queuing models can help provide a scientific way to analyze the needs and anticipate the voting lines based on the number of machines provided.

The basic voting precinct can be modeled as a multichannel queuing system. By evaluating a range of values for the forecast number of voters at the different times of the day, a determination can be made on how long the lines will be, based on the number of voting machines available. While there still could be some lines if the state does not have enough voting machines to meet the anticipated need, the proper distribution of these machines will help to keep the waiting times reasonable.

Source: Based on Alexander S. Belenky and Richard C. Larson. "To Queue or Not to Queue," *OR/MS Today* 33, 3 (June 2006): 30–34.

Equations for the Constant Service Time Model

Constant service model formulas follow:

1. Average length of the queue:

$$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)} \tag{13-19}$$

2. Average waiting time in the queue:

$$W_q = \frac{\lambda}{2\mu(\mu - \lambda)} \tag{13-20}$$

3. Average number of customers in the system:

$$L = L_q + \frac{\lambda}{\mu} \tag{13-21}$$

4. Average time in the system:

$$W = W_q + \frac{1}{\mu} \tag{13-22}$$

Garcia-Golding Recycling, Inc.

Garcia-Golding Recycling, Inc., collects and compacts aluminum cans and glass bottles in New York City. Its truck drivers, who arrive to unload these materials for recycling, currently wait an average of 15 minutes before emptying their loads. The cost of the driver and truck time wasted while in queue is valued at $60 per hour. A new automated compactor can be purchased that will process truck loads at a constant rate of 12 trucks per hour (i.e., 5 minutes per truck). Trucks arrive according to a Poisson distribution at an average rate of 8 per hour. If the new compactor is

PROGRAM 13.3 Excel QM Solution for Constant Service Time Model with Garcia-Golding Recycling Example

	A	B	C	D	E
1	**Garcia-Golding Recycling**				
2					
3	**Waiting Lines**		**M/D/1 (Constant Service Times)**		
4	The arrival RATE and service RATE both must be rates and use the same time unit. Given a time				
5					
6	**Data**			**Results**	
7	Arrival rate (λ)	8		**Average server utilization(ρ)**	**0.66667**
8	Service rate (μ)	12		**Average number of customers in the queue(L_q)**	**0.66667**
9				**Average number of customers in the system(L_s)**	**1.33333**
10				**Average waiting time in the queue(W_q)**	**0.08333**
11				**Average time in the system(W_s)**	**0.16667**
12				**Probability (% of time) system is empty (P_0)**	**0.33333**

put in use, its cost will be amortized at a rate of \$3 per truck unloaded. A summer intern from a local college did the following analysis to evaluate the costs versus benefits of the purchase:

Cost analysis for the recycling example.

$$\textit{Current}\text{ waiting cost/trip} = \left(\tfrac{1}{4}\text{ hour waiting now}\right)(\$60/\text{hour cost})$$
$$= \$15/\text{trip}$$

$$\textit{New}\text{ system: } \lambda = 8 \text{ trucks/hour arriving,}$$
$$\mu = 12 \text{ trucks/hour served}$$

$$\text{Average waiting time in queue} = W_q = \frac{\lambda}{2\mu(\mu - \lambda)} = \frac{8}{2(12)(12 - 8)}$$
$$= \frac{1}{12}\text{ hour}$$

$$\text{Waiting cost/trip with new compactor} = \left(\tfrac{1}{12}\text{ hour wait}\right)(\$60/\text{hour cost}) = \$5/\text{trip}$$
$$\text{Savings with new equipment} = \$15\text{ (current system)} - \$5\text{ (new system)}$$
$$= \$10/\text{trip}$$
$$\text{Cost of new equipment amortized} = \underline{\$3/\text{trip}}$$
$$\text{Net savings} = \$7/\text{trip}$$

USING EXCEL QM FOR GARCIA-GOLDING'S CONSTANT SERVICE TIME MODEL To use Excel QM for this problem, from the Excel QM menu, select *Waiting Lines - Constant Service Time Model (M/D/1)*. When the spreadsheet appears, enter the arrival rate (8) and the service rate (12). Once these are entered, the solution shown in Program 13.3 will be displayed.

13.7 Finite Population Model (*M*/*M*/1 with Finite Source)

When there is a limited population of potential customers for a service facility, we need to consider a different queuing model. This model would be used, for example, if you were considering equipment repairs in a factory that has five machines, if you were in charge of maintenance for a fleet of 10 commuter airplanes, or if you ran a hospital ward that has 20 beds. The limited population model permits any number of repair people (servers) to be considered.

The reason this model differs from the three earlier queuing models is that there is now a *dependent* relationship between the length of the queue and the arrival rate. To illustrate the extreme situation, if your factory had five machines and all were broken and awaiting repair, the arrival rate would drop to zero. In general, as the waiting line becomes longer in the limited population model, the arrival rate of customers or machines drops lower.

In this section, we describe a finite calling population model that has the following assumptions:

1. There is only one server.
2. The population of units seeking service is finite.*
3. Arrivals follow a Poisson distribution, and service times are exponentially distributed.
4. Customers are served on a first-come, first-served basis.

Equations for the Finite Population Model

Using

$$\lambda = \text{mean arrival rate}, \mu = \text{mean service rate}, N = \text{size of the population}$$

the operating characteristics for the finite population model with a single channel or server on duty are as follows:

1. Probability that the system is empty:

$$P_0 = \frac{1}{\sum_{n=0}^{N} \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n} \qquad \textbf{(13-23)}$$

2. Average length of the queue:

$$L_q = N - \left(\frac{\lambda + \mu}{\lambda}\right)(1 - P_0) \qquad \textbf{(13-24)}$$

3. Average number of customers (units) in the system:

$$L = L_q + (1 - P_0) \qquad \textbf{(13-25)}$$

4. Average waiting time in the queue:

$$W_q = \frac{L_q}{(N - L)\lambda} \qquad \textbf{(13-26)}$$

5. Average time in the system:

$$W = W_q + \frac{1}{\mu} \qquad \textbf{(13-27)}$$

6. Probability of n units in the system:

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_0 \quad \text{for } n = 0, 1, \ldots, N \qquad \textbf{(13-28)}$$

Department of Commerce Example

Past records indicate that each of the five high-speed "page" printers at the U.S. Department of Commerce, in Washington, D.C., needs repair after about 20 hours of use. Breakdowns have been determined to be Poisson distributed. The one technician on duty can service a printer in an average of 2 hours, following an exponential distribution.

*Although there is no definite number that we can use to divide finite from infinite populations, the general rule of thumb is this: If the number in the queue is a significant proportion of the calling population, use a finite queuing model. *Finite Queuing Tables*, by L. G. Peck and R. N. Hazelwood (New York: John Wiley & Sons, Inc., 1958), eliminates much of the mathematics involved in computing the operating characteristics for such a model.

To compute the system's operation characteristics we first note that the mean arrival rate is $\lambda = {}^{1}\!/_{20} = 0.05$ printer/hour. The mean service rate is $\mu = {}^{1}\!/_{2} = 0.50$ printer/hour. Then

1. $P_0 = \dfrac{1}{\displaystyle\sum_{n=0}^{5} \frac{5!}{(5-n)!}\left(\frac{0.05}{0.5}\right)^n} = 0.564$ (we leave these calculations for you to confirm)

2. $L_q = 5 - \left(\dfrac{0.05 + 0.5}{0.05}\right)(1 - P_0) = 5 - (11)(1 - 0.564) = 5 - 4.8$
   $= 0.2$ printer

3. $L = 0.2 + (1 - 0.564) = 0.64$ printer

4. $W_q = \dfrac{0.2}{(5 - 0.64)(0.05)} = \dfrac{0.2}{0.22} = 0.91$ hour

5. $W = 0.91 + \dfrac{1}{0.50} = 2.91$ hours

If printer downtime costs \$120 per hour and the technician is paid \$25 per hour, we can also compute the total cost per hour:

$$\begin{aligned}\text{Total hourly cost} &= (\text{Average number of printers down})(\text{Cost per downtime hour}) \\ &\quad + \text{Cost per technician hour} \\ &= (0.64)(\$120) + \$25 = \$76.80 + \$25.00 = \$101.80\end{aligned}$$

SOLVING THE DEPARTMENT OF COMMERCE FINITE POPULATION MODEL WITH EXCEL QM To use Excel QM for this problem, from the Excel QM menu, select *Waiting Lines - Limited Population Model (M/M/s)*. When the spreadsheet appears, enter the arrival rate (8), service rate (12), number of servers, and population size. Once these are entered, the solution shown in Program 13.4 will be displayed. Additional output is also available.

PROGRAM 13.4 **Excel QM Solution for Finite Population Model with Department of Commerce Example**

	A	B	C	D	E
1	Department of Commerce				
2					
3	Waiting Lines		M/M/s with a finite population		
4	The arrival rate is for each member of the population. If they go for service every 20				
5					
6	Data			Results	
7	Arrival rate (λ) per customer	0.05		Average server utilization(ρ)	0.43605
8	Service rate (μ)	0.5		Average number of customers in the queue(L_q)	0.20347
9	Number of servers	1		Average number of customers in the system(L_s)	0.63952
10	Population size (N)	5		Average waiting time in the queue(W_q)	0.93326
11				Average time in the system(W_s)	2.93326
12				Probability (% of time) system is empty (P_0)	0.56395
13				Effective arrival rate	0.21802

13.8 Some General Operating Characteristic Relationships

A **steady state** *is the normal operating condition of the queuing system.*

A queuing system is in a **transient state** *before the steady state is reached.*

Certain relationships exist among specific operating characteristics for any queuing system in a **steady state**. A steady state condition exists when a queuing system is in its normal stabilized operating condition, usually after an initial or **transient state** that may occur (e.g., having customers waiting at the door when a business opens in the morning). Both the arrival rate and the service rate should be stable in this state. John D. C. Little is credited with the first two of these relationships, and hence they are called **Little's Flow Equations**:

$$L = \lambda W \text{ (or } W = L/\lambda) \tag{13-29}$$

$$L_q = \lambda W_q \text{ (or } W_q = L_q/\lambda) \tag{13-30}$$

A third condition that must always be met is

Average time in system = average time in queue + average time receiving service

$$W = W_q + 1/\mu \tag{13-31}$$

The advantage of these formulas is that once one of these four characteristics is known, the other characteristics can easily be found. This is important because for certain queuing models, one of these may be much easier to determine than the others. These are applicable to all of the queuing systems discussed in this chapter except the finite population model.

13.9 More Complex Queuing Models and the Use of Simulation

Many practical waiting line problems that occur in production and operations service systems have characteristics like those of Arnold's Muffler Shop, Garcia-Golding Recycling Inc., or the Department of Commerce. This is true when the situation calls for single- or multichannel waiting lines, with Poisson arrivals and exponential or constant service times, an infinite calling population, and FIFO service.

More sophisticated models exist to handle variations of basic assumptions, but when even these do not apply we can turn to computer simulation, the topic of Chapter 14.

Often, however, *variations* of this specific case are present in an analysis. Service times in an automobile repair shop, for example, tend to follow the normal probability distribution instead of the exponential. A college registration system in which seniors have first choice of courses and hours over all other students is an example of a first-come, first-served model with a preemptive priority queue discipline. A physical examination for military recruits is an example of a multiphase system—one that differs from the single-phase models discussed in this chapter. A recruit first lines up to have blood drawn at one station, then waits to take an eye exam at the next station, talks to a psychiatrist at the third, and is examined by a doctor for medical problems at the fourth. At each phase, the recruit must enter another queue and wait his or her turn.

Models to handle these cases have been developed by operations researchers. The computations for the resulting mathematical formulations are somewhat more complex than the ones covered in this chapter,* and many real-world queuing applications are too complex to be modeled analytically at all. When this happens, quantitative analysts usually turn to *computer simulation*.

Simulation, the topic of Chapter 14, is a technique in which random numbers are used to draw inferences about probability distributions (such as arrivals and services). Using this approach, many hours, days, or months of data can be developed by a computer in a few seconds. This allows analysis of controllable factors, such as adding another service channel, without actually doing so physically. Basically, whenever a standard analytical queuing model provides only a poor approximation of the actual service system, it is wise to develop a simulation model instead.

*Often, the *qualitative* results of queuing models are as useful as the quantitative results. Results show that it is inherently more efficient to pool resources, use central dispatching, and provide single multiple-server systems rather than multiple single-server systems.

Summary

Waiting lines and service systems are important parts of the business world. In this chapter we describe several common queuing situations and present mathematical models for analyzing waiting lines following certain assumptions. Those assumptions are that (1) arrivals come from an infinite or very large population, (2) arrivals are Poisson distributed, (3) arrivals are treated on a FIFO basis and do not balk or renege, (4) service times follow the negative exponential distribution or are constant, and (5) the average service rate is faster than the average arrival rate.

The models illustrated in this chapter are for single-channel, single-phase and multichannel, single-phase problems. After a series of operating characteristics are computed, total expected costs are studied. As shown graphically in Figure 13.1, total cost is the sum of the cost of providing service plus the cost of waiting time.

Key operating characteristics for a system are shown to be (1) utilization rate, (2) percent idle time, (3) average time spent waiting in the system and in the queue, (4) average number of customers in the system and in the queue, and (5) probabilities of various numbers of customers in the system.

The chapter emphasizes that a variety of queuing models exist that do not meet all of the assumptions of the traditional models. In these cases we use more complex mathematical models or turn to a technique called computer simulation. The application of simulation to problems of queuing systems, inventory control, machine breakdown, and other quantitative analysis situations is the topic discussed in Chapter 14.

Glossary

Balking The case in which arriving customers refuse to join the waiting line.

Calling Population The population of items from which arrivals at the queuing system come.

FIFO A queue discipline (meaning first-in, first-out) in which the customers are served in the strict order of arrival.

Kendall Notation A method of classifying queuing systems based on the distribution of arrivals, the distribution of service times, and the number of service channels.

Limited, or Finite, Population A case in which the number of customers in the system is a significant proportion of the calling population.

Limited Queue Length A waiting line that cannot increase beyond a specific size.

Little's Flow Equations A set of relationships that exist for any queuing system in a steady state.

$M/D/1$ Kendall notation for the constant service time model.

$M/M/1$ Kendall notation for the single-channel model with Poisson arrivals and exponential service times.

$M/M/m$ Kendall notation for the multichannel queuing model (with m servers) and Poisson arrivals and exponential service times.

Multichannel Queuing System A system that has more than one service facility, all fed by the same single queue.

Multiphase System. A system in which service is received from more than one station, one after the other.

Negative Exponential Probability Distribution A probability distribution that is often used to describe random service times in a service system.

Operating Characteristics Descriptive characteristics of a queuing system, including the average number of customers in a line and in the system, the average waiting times in a line and in the system, and percent idle time.

Poisson Distribution A probability distribution that is often used to describe random arrivals in a queue.

Queue Discipline The rule by which customers in a line receive service.

Queuing Theory The mathematical study of waiting lines or queues.

Reneging The case in which customers enter a queue but then leave before being serviced.

Service Cost The cost of providing a particular level of service.

Single-Channel Queuing System A system with one service facility fed by one queue.

Single-Phase System A queuing system in which service is received at only one station.

Steady State The normal, stabilized operating condition of a queuing system.

Transient State The initial condition of a queuing system before a steady state is reached.

Unlimited, or Infinite, Population A calling population that is very large relative to the number of customers currently in the system.

Unlimited Queue Length A queue that can increase to an infinite size.

Utilization Factor (ρ) The proportion of the time that service facilities are in use.

Waiting Cost. The cost to the firm of having customers or objects waiting to be serviced.

Waiting Line (Queue) One or more customers or objects waiting to be served.

Key Equations

λ = mean number of arrivals per time period

μ = mean number of people or items served per time period

Equations 13-1 through 13-7 describe operating characteristics in the single-channel model that has Poisson arrival and exponential service rates.

(13-1) L = average number of units (customers) in the system

$$= \frac{\lambda}{\mu - \lambda}$$

(13-2) W = average time a unit spends in the system (Waiting time+Service time)

$$= \frac{1}{\mu - \lambda}$$

(13-3) L_q = average number of units in the queue

$$= \frac{\lambda^2}{\mu(\mu - \lambda)}$$

(13-4) W_q = average time a unit spends waiting in the queue

$$= \frac{\lambda}{\mu(\mu - \lambda)}$$

(13-5) ρ = utilization factor for the system $= \dfrac{\lambda}{\mu}$

(13-6) P_0 = probability of 0 units in the system (that is, the service unit is idle)

$$= 1 - \frac{\lambda}{\mu}$$

(13-7) $P_{n>k}$ = probability of more than k units in the system

$$= \left(\frac{\lambda}{\mu}\right)^{k+1}$$

Equations 13-8 through 13-12 are used for finding the costs of a queuing system.

(13-8) Total service cost $= mC_s$

where

m = number of channels

C_s = service cost (labor cost) of each channel

(13-9) Total waiting cost per time period $= (\lambda W)C_w$

C_w = cost of waiting

Waiting time cost based on time in the system.

(13-10) Total waiting cost per time period $= (\lambda W_q)C_w$

Waiting time cost based on time in the queue.

(13-11) Total cost $= mC_s + \lambda W C_w$

Waiting time cost based on time in the system.

(13-12) Total cost $= mC_s + \lambda W_q C_w$

Waiting time cost based on time in the queue.

Equations 13-13 through 13-18 describe operating characteristics in multichannel models that have Poisson arrival and exponential service rates, where m = the number of open channels.

(13-13) $$P_0 = \frac{1}{\left[\displaystyle\sum_{n=0}^{n=m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{m\mu}{m\mu - \lambda}}$$

for $m\mu > \lambda$

Probability that there are no people or units in the system.

(13-14) $$L = \frac{\lambda\mu(\lambda/\mu)^m}{(m-1)!(m\mu - \lambda)^2}P_0 + \frac{\lambda}{\mu}$$

Average number of people or units in the system.

(13-15) $$W = \frac{\mu(\lambda/\mu)^m}{(m-1)!(m\mu - \lambda)^2}P_0 + \frac{1}{\mu} = \frac{L}{\lambda}$$

Average time a unit spends in the waiting line or being serviced (namely, in the system).

(13-16) $$L_q = L - \frac{\lambda}{\mu}$$

Average number of people or units in line waiting for service.

(13-17) $$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda}$$

Average time a person or unit spends in the queue waiting for service.

(13-18) $$\rho = \frac{\lambda}{m\mu}$$

Utilization rate.

Equations 13-19 through 13-22 describe operating characteristics in single-channel models that have Poisson arrivals and constant service rates.

(13-19) $$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}$$

Average length of the queue.

(13-20) $$W_q = \frac{\lambda}{2\mu(\mu - \lambda)}$$

Average waiting time in the queue.

(13-21) $$L = L_q + \frac{\lambda}{\mu}$$

Average number of customers in the system.

(13-22) $$W = W_q + \frac{1}{\mu}$$

Average waiting time in the system.

Equations 13-23 through 13-28 describe operating characteristics in single-channel models that have Poisson arrivals and exponential service rates and a finite calling population.

(13-23) $P_0 = \dfrac{1}{\sum_{n=0}^{N} \dfrac{N!}{(N-n)!}\left(\dfrac{\lambda}{\mu}\right)^n}$

Probability that the system is empty.

(13-24) $L_q = N - \left(\dfrac{\lambda + \mu}{\lambda}\right)(1 - P_0)$

Average length of the queue.

(13-25) $L = L_q + (1 - P_0)$

Average number of units in the system.

(13-26) $W_q = \dfrac{L_q}{(N - L)\lambda}$

Average time in the queue.

(13-27) $W = W_q + \dfrac{1}{\mu}$

Average time in the system.

(13-28) $P_n = \dfrac{N!}{(N-n)!}\left(\dfrac{\lambda}{\mu}\right)^n P_0$ for $n = 0, 1, \ldots, N$

Probability of n units in the system.

Equations 13-29 to 13-31 are Little's Flow Equations, which can be used when a steady state condition exists.

(13-29) $L = \lambda W$

(13-30) $L_q = \lambda W_q$

(13-31) $W = W_q + 1/\mu$

Solved Problems

Solved Problem 13-1

The Maitland Furniture store gets an average of 50 customers per shift. The manager of Maitland wants to calculate whether she should hire 1, 2, 3, or 4 salespeople. She has determined that average waiting times will be 7 minutes with 1 salesperson, 4 minutes with 2 salespeople, 3 minutes with 3 salespeople, and 2 minutes with 4 salespeople. She has estimated the cost per minute that customers wait at $1. The cost per salesperson per shift (including benefits) is $70.

How many salespeople should be hired?

Solution

The manager's calculations are as follows:

	NUMBER OF SALESPEOPLE			
	1	2	3	4
(a) Average number of customers per shift	50	50	50	50
(b) Average waiting time per customer (minutes)	7	4	3	2
(c) Total waiting time per shift ($a \times b$) (minutes)	350	200	150	100
(d) Cost per minute of waiting time (estimated)	$1.00	$1.00	$1.00	$1.00
(e) Value of lost time ($c \times d$) per shift	$ 350	$ 200	$ 150	$ 100
(f) Salary cost per shift	$ 70	$ 140	$ 210	$ 280
(g) Total cost per shift	$ 420	$ 340	$ 360	$ 380

Because the minimum total cost per shift relates to two salespeople, the manager's optimum strategy is to hire 2 salespeople.

Solved Problem 13-2

Marty Schatz owns and manages a chili dog and soft drink store near the campus. Although Marty can service 30 customers per hour on the average (μ), he only gets 20 customers per hour (λ). Because Marty could wait on 50% more customers than actually visit his store, it doesn't make sense to him that he should have any waiting lines.

Marty hires you to examine the situation and to determine some characteristics of his queue. After looking into the problem, you find this to be an $M/M/1$ system. What are your findings?

Solution

$$L = \frac{\lambda}{\mu - \lambda} = \frac{20}{30 - 20} = 2 \text{ customers in the system on the average}$$

$$W = \frac{1}{\mu - \lambda} = \frac{1}{30 - 20} = 0.1 \text{ hour (6 minutes) that the average customer spends in the total system}$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{20^2}{30(30 - 20)} = 1.33 \text{ customers waiting for service in line on the average}$$

$$w_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{20}{30(30 - 20)} = {}^1\!/_{15} \text{ hour} = \text{(4 minutes)} = \text{average waiting time of a customer in the queue awaiting service}$$

$$\rho = \frac{\lambda}{\mu} = \frac{20}{30} = 0.67 = \text{percentage of the time that Marty is busy waiting on customers}$$

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \rho = 0.33 = \text{probability that there are no customers in the system (being waited on or waiting in the queue) at any given time}$$

Probability of *k* or More Customers Waiting in Line and/or Being Waited On

k	$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$
0	0.667
1	0.444
2	0.296
3	0.198

Solved Problem 13-3

Refer to Solved Problem 13-2. Marty agreed that these figures seemed to represent his approximate business situation. You are quite surprised at the length of the lines and elicit from him an estimated value of the customer's waiting time (in the queue, not being waited on) at 10 cents per minute. During the 12 hours that he is open he gets $(12 \times 20) = 240$ customers. The average customer is in a queue 4 minutes, so the total customer waiting time is $(240 \times 4 \text{ minutes}) = 960$ minutes. The value of 960 minutes is (\$0.10)(960 minutes) = \$96. You tell Marty that not only is 10 cents per minute quite conservative, but he could probably save most of that \$96 of customer ill will if he hired another salesclerk. After much haggling, Marty agrees to provide you with all the chili dogs you can eat during a week-long period in exchange for your analysis of the results of having two clerks wait on the customers.

Assuming that Marty hires one additional salesclerk whose service rate equals Marty's rate, complete the analysis.

Solution

With two cash registers open, the system becomes two channel, or $m = 2$. The computations yield

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=m-1} \frac{1}{n!}\left[\frac{20}{30}\right]^n\right] + \frac{1}{2!}\left[\frac{20}{30}\right]^2\left[\frac{2(30)}{2(30) - 20}\right]}$$

$$= \frac{1}{(1)(2/3)^0 + (1)(2/3)^1 + (1/2)(4/9)(6/4)} = 0.5$$

$$= \text{probability of no customers in the system}$$

$$L = \left[\frac{(20)(30)(20/30)^2}{(2-1)![(2)(30-20)]^2}\right]0.5 + \frac{20}{30} = 0.75 \text{ customer in the system on the average}$$

$$W = \frac{L}{\lambda} = \frac{3/4}{20} = \frac{3}{80}\text{ hour} = 2.25 \text{ minutes that the average customer spends in the total system}$$

$$L_q = L - \frac{\lambda}{\mu} = \frac{3}{4} - \frac{20}{30} = \frac{1}{12} = 0.083 \text{ customer waiting for service in line on the average}$$

$$W_q = \frac{L_q}{\lambda} = \frac{1/2}{20} = \frac{1}{240}\text{ hour} = \frac{1}{4}\text{ minute} = \text{average waiting time of a customer in the queue itself (not being serviced)}$$

$$\rho = \frac{\lambda}{m\mu} = \frac{20}{2(30)} = \frac{1}{3} = 0.33 = \text{utilization rate}$$

You now have (240 customers) $\times$ (1/240 hour) $=$ 1 hour total customer waiting time per day.

Total cost of 60 minutes of customer waiting time is (60 minutes)($0.10 per minute) $=$ $6.

Now you are ready to point out to Marty that the hiring of one additional clerk will save $96 – $6 = $90 of customer ill will per 12-hour shift. Marty responds that the hiring should also reduce the number of people who look at the line and leave as well as those who get tired of waiting in line and leave. You tell Marty that you are ready for two chili dogs, extra hot.

Solved Problem 13-4

Vacation Inns is a chain of hotels operating in the southwestern part of the United States. The company uses a toll-free telephone number to take reservations for any of its hotels. The average time to handle each call is 3 minutes, and an average of 12 calls are received per hour. The probability distribution that describes the arrivals is unknown. Over a period of time it is determined that the average caller spends 6 minutes either on hold or receiving service. Find the average time in the queue, the average time in the system, the average number in the queue, and the average number in the system.

Solution

The probability distributions are unknown, but we are given the average time in the system (6 minutes). Thus, we can use Little's Flow Equations:

$$W = 6\text{ minutes} = 6/60\text{ hour} = 0.1\text{ hour}$$
$$\lambda = 12\text{ per hour}$$
$$\mu = 60/3 = 20\text{ per hour}$$
$$\text{Average time in queue} = W_q = W - 1/\mu = 0.1 - 1/20 = 0.1 - 0.05 = 0.05\text{ hour}$$
$$\text{Average number in system} = L = \lambda W = 12(0.1) = 1.2\text{ callers}$$
$$\text{Average number in queue} = L_q = \lambda W_q = 12(0.05) = 0.6\text{ caller}$$

Self-Test

- Before taking the self-test, refer to the learning objectives at the beginning of the chapter, the notes in the margins, and the glossary at the end of the chapter.
- Use the key at the back of the book to correct your answers.
- Restudy pages that correspond to any questions that you answered incorrectly or material you feel uncertain about.

1. Most systems use the queue discipline known as the FIFO rule.
   a. True
   b. False
2. Before using exponential distributions to build queuing models, the quantitative analyst should determine if the service time data fit the distribution.
   a. True
   b. False

3. In a multichannel, single-phase queuing system, the arrival will pass through at least two different service facilities.
   a. True
   b. False
4. Which of the following is *not* an assumption in $M/M/1$ models?
   a. arrivals come from an infinite or very large population
   b. arrivals are Poisson distributed
   c. arrivals are treated on a FIFO basis and do not balk or renege
   d. service times follow the exponential distribution
   e. the average arrival rate is faster than the average service rate
5. A queuing system described as $M/D/2$ would have
   a. exponential service times.
   b. two queues.
   c. constant service times.
   d. constant arrival rates.
6. Cars enter the drive-through of a fast-food restaurant to place an order, and then they proceed to pay for the food and pick up the order. This is an example of
   a. a multichannel system.
   b. a multiphase system.
   c. a multiqueue system.
   d. none of the above.
7. The utilization factor for a system is defined as
   a. mean number of people served divided by the mean number of arrivals per time period.
   b. the average time a customer spends waiting in a queue.
   c. proportion of the time the service facilities are in use.
   d. the percentage of idle time.
   e. none of the above.
8. Which of the following would not have a FIFO queue discipline?
   a. fast-food restaurant
   b. post office
   c. checkout line at grocery store
   d. emergency room at a hospital
9. A company has one computer technician who is responsible for repairs on the company's 20 computers. As a computer breaks, the technician is called to make the repair. If the repairperson is busy, the machine must wait to be repaired. This is an example of
   a. a multichannel system.
   b. a finite population system.
   c. a constant service rate system.
   d. a multiphase system.
10. In performing a cost analysis of a queuing system, the waiting time cost (C_w) is sometimes based on the time in the queue and sometimes based on the time in the system. The waiting cost should be based on time in the system for which of the following situations?
   a. waiting in line to ride an amusement park ride
   b. waiting to discuss a medical problem with a doctor
   c. waiting for a picture and an autograph from a rock star
   d. waiting for a computer to be fixed so it can be placed back in service
11. Customers enter the waiting line at a cafeteria on a first-come, first-served basis. The arrival rate follows a Poisson distribution, and service times follow an exponential distribution. If the average number of arrivals is 6 per minute and the average service rate of a single server is 10 per minute, what is the average number of customers in the system?
   a. 0.6
   b. 0.9
   c. 1.5
   d. 0.25
   e. none of the above
12. In the standard queuing model, we assume that the queue discipline is ____________.
13. The service *time* in the $M/M/1$ queuing model is assumed to be ____________.
14. When managers find standard queuing formulas inadequate or the mathematics unsolvable, they often resort to ____________ to obtain their solutions.

Discussion Questions and Problems

Discussion Questions

13-1 What is the waiting line problem? What are the components in a waiting line system?

13-2 What are the assumptions underlying common queuing models?

13-3 Describe the important operating characteristics of a queuing system.

13-4 Why must the service rate be greater than the arrival rate in a single-channel queuing system?

13-5 Briefly describe three situations in which the FIFO discipline rule is not applicable in queuing analysis.

13-6 Provide examples of four situations in which there is a limited, or finite, population.

13-7 What are the components of the following systems? Draw and explain the configuration of each.
(a) barbershop
(b) car wash
(c) laundromat
(d) small grocery store

13-8 Give an example of a situation in which the waiting time cost would be based on waiting time in the queue. Give an example of a situation in which the waiting time cost would be based on waiting time in the system.

13-9 Do you think the Poisson distribution, which assumes independent arrivals, is a good estimation of arrival

rates in the following queuing systems? Defend your position in each case.
(a) cafeteria in your school
(b) barbershop
(c) hardware store
(d) dentist's office
(e) college class
(f) movie theater

Problems*

: 13-10 The Schmedley Discount Department Store has approximately 300 customers shopping in its store between 9 A.M. and 5 P.M. on Saturdays. In deciding how many cash registers to keep open each Saturday, Schmedley's manager considers two factors: customer waiting time (and the associated waiting cost) and the service costs of employing additional checkout clerks. Checkout clerks are paid an average of $8 per hour. When only one is on duty, the waiting time per customer is about 10 minutes (or $\frac{1}{6}$ hour); when two clerks are on duty, the average checkout time is 6 minutes per person; 4 minutes when three clerks are working; and 3 minutes when four clerks are on duty.

Schmedley's management has conducted customer satisfaction surveys and has been able to estimate that the store suffers approximately $10 in lost sales and goodwill for every *hour* of customer time spent waiting in checkout lines. Using the information provided, determine the optimal number of clerks to have on duty each Saturday to minimize the store's total expected cost.

• 13-11 The Rockwell Electronics Corporation retains a service crew to repair machine breakdowns that occur on an average of $\lambda = 3$ per day (approximately Poisson in nature).

The crew can service an average of $\mu = 8$ machines per day, with a repair time distribution that resembles the exponential distribution.
(a) What is the utilization rate of this service system?
(b) What is the average downtime for a machine that is broken?
(c) How many machines are waiting to be serviced at any given time?
(d) What is the probability that more than one machine is in the system? Probability that more than two are broken and waiting to be repaired or being serviced? More than three? More than four?

• 13-12 From historical data, Harry's Car Wash estimates that dirty cars arrive at the rate of 10 per hour all day Saturday. With a crew working the wash line, Harry figures that cars can be cleaned at the rate of one every 5 minutes. One car at a time is cleaned in this example of a single-channel waiting line.

Assuming Poisson arrivals and exponential service times, find the
(a) average number of cars in line.
(b) average time a car waits before it is washed.
(c) average time a car spends in the service system.
(d) utilization rate of the car wash.
(e) probability that no cars are in the system.

: 13-13 Mike Dreskin manages a large Los Angeles movie theater complex called Cinema I, II, III, and IV. Each of the four auditoriums plays a different film; the schedule is set so that starting times are staggered to avoid the large crowds that would occur if all four movies started at the same time. The theater has a single ticket booth and a cashier who can maintain an average service rate of 280 movie patrons per hour. Service times are assumed to follow an exponential distribution. Arrivals on a typically active day are Poisson distributed and average 210 per hour.

To determine the efficiency of the current ticket operation, Mike wishes to examine several queue operating characteristics.
(a) Find the average number of moviegoers waiting in line to purchase a ticket.
(b) What percentage of the time is the cashier busy?
(c) What is the average time that a customer spends in the system?
(d) What is the average time spent waiting in line to get to the ticket window?
(e) What is the probability that there are more than two people in the system? More than three people? More than four?

: 13-14 A university cafeteria line in the student center is a self-serve facility in which students select the food items they want and then form a single line to pay the cashier. Students arrive at a rate of about four per minute according to a Poisson distribution. The single cashier ringing up sales takes about 12 seconds per customer, following an exponential distribution.
(a) What is the probability that there are more than two students in the system? More than three students? More than four?
(b) What is the probability that the system is empty?
(c) How long will the average student have to wait before reaching the cashier?
(d) What is the expected number of students in the queue?
(e) What is the average number in the system?
(f) If a second cashier is added (who works at the same pace), how will the operating characteristics computed in parts (b), (c), (d), and (e) change? Assume that customers wait in a single line and go to the first available cashier.

*Note: Q means the problem may be solved with QM for Windows; X means the problem may be solved with Excel QM; and Q/X means the problem may be solved with QM for Windows and/or Excel QM.

13-15 The wheat harvesting season in the American Midwest is short, and most farmers deliver their truckloads of wheat to a giant central storage bin within a two-week span. Because of this, wheat-filled trucks waiting to unload and return to the fields have been known to back up for a block at the receiving bin. The central bin is owned cooperatively, and it is to every farmer's benefit to make the unloading/storage process as efficient as possible. The cost of grain deterioration caused by unloading delays, the cost of truck rental, and idle driver time are significant concerns to the cooperative members. Although farmers have difficulty quantifying crop damage, it is easy to assign a waiting and unloading cost for truck and driver of $18 per hour. The storage bin is open and operated 16 hours per day, 7 days per week, during the harvest season and is capable of unloading 35 trucks per hour according to an exponential distribution. Full trucks arrive all day long (during the hours the bin is open) at a rate of about 30 per hour, following a Poisson pattern.

To help the cooperative get a handle on the problem of lost time while trucks are waiting in line or unloading at the bin, find the

(a) average number of trucks in the unloading system.
(b) average time per truck in the system.
(c) utilization rate for the bin area.
(d) probability that there are more than three trucks in the system at any given time.
(e) total daily cost to the farmers of having their trucks tied up in the unloading process.

The cooperative, as mentioned, uses the storage bin only two weeks per year. Farmers estimate that enlarging the bin would cut unloading costs by 50% next year. It will cost $9,000 to do so during the off-season. Would it be worth the cooperative's while to enlarge the storage area?

13-16 Ashley's Department Store in Kansas City maintains a successful catalog sales department in which a clerk takes orders by telephone. If the clerk is occupied on one line, incoming phone calls to the catalog department are answered automatically by a recording machine and asked to wait. As soon as the clerk is free, the party that has waited the longest is transferred and answered first. Calls come in at a rate of about 12 per hour. The clerk is capable of taking an order in an average of 4 minutes. Calls tend to follow a Poisson distribution, and service times tend to be exponential. The clerk is paid $10 per hour, but because of lost goodwill and sales, Ashley's loses about $50 per hour of customer time spent waiting for the clerk to take an order.

(a) What is the average time that catalog customers must wait before their calls are transferred to the order clerk?
(b) What is the average number of callers waiting to place an order?
(c) Ashley's is considering adding a second clerk to take calls. The store would pay that person the same $10 per hour. Should it hire another clerk? Explain.

13-17 Automobiles arrive at the drive-through window at a post office at the rate of 4 every 10 minutes. The average service time is 2 minutes. The Poisson distribution is appropriate for the arrival rate and service times are exponentially distributed.

(a) What is the average time a car is in the system?
(b) What is the average number of cars in the system?
(c) What is the average time cars spend waiting to receive service?
(d) What is the average number of cars in line *behind* the customer receiving service?
(e) What is the probability that there are no cars at the window?
(f) What percentage of the time is the postal clerk busy?
(g) What is the probability that there are exactly two cars in the system?

13-18 For the post office in Problem 13-17, a second drive-through window is being considered. A single line would be formed and as a car reached the front of the line it would go to the next available clerk. The clerk at the new window works at the same rate as the current one.

(a) What is the average time a car is in the system?
(b) What is the average number of cars in the system?
(c) What is the average time cars spend waiting to receive service?
(d) What is the average number of cars in line *behind* the customer receiving service?
(e) What is the probability that there are no cars in the system?
(f) What percentage of the time are the clerks busy?
(g) What is the probability that there are exactly two cars in the system?

13-19 Juhn and Sons Wholesale Fruit Distributors employ one worker whose job is to load fruit on outgoing company trucks. Trucks arrive at the loading gate at an average of 24 per day, or 3 per hour, according to a Poisson distribution. The worker loads them at a rate of 4 per hour, following approximately the exponential distribution in service times.

Determine the operating characteristics of this loading gate problem. What is the probability that there will be more than three trucks either being loaded or waiting? Discuss the results of your queuing model computation.

13-20 Juhn believes that adding a second fruit loader will substantially improve the firm's efficiency. He estimates that a two-person crew, still acting like a single-server system, at the loading gate will double the loading rate from 4 trucks per hour to 8 trucks per hour. Analyze the effect on the queue of such a change and compare the results with those found in Problem 13-19.

13-21 Truck drivers working for Juhn and Sons (see Problems 13-19 and 13-20) are paid a salary of \$10 per hour on average. Fruit loaders receive about \$6 per hour. Truck drivers waiting in the queue or at the loading gate are drawing a salary but are productively idle and unable to generate revenue during that time. What would be the *hourly* cost savings to the firm associated with employing two loaders instead of one?

13-22 Juhn and Sons Wholesale Fruit Distributors (of Problem 13-19) are considering building a second platform or gate to speed the process of loading their fruit trucks. This, they think, will be even more efficient than simply hiring another loader to help out the first platform (as in Problem 13-20).

Assume that workers at each platform will be able to load 4 trucks per hour each and that trucks will continue to arrive at the rate of 3 per hour. Find the waiting line's new operating conditions. Is this new approach indeed speedier than the other two considered?

13-23 Bill First, general manager of Worthmore Department Store, has estimated that every hour of customer time spent waiting in line for the sales clerk to become available costs the store \$100 in lost sales and goodwill. Customers arrive at the checkout counter at the rate of 30 per hour, and the average service time is 3 minutes. The Poisson distribution describes the arrivals and the service times are exponentially distributed. The number of sales clerks can be 2, 3, or 4, with each one working at the same rate. Bill estimates the salary and benefits for each clerk to be \$10 per hour. The store is open 10 hours per day.

(a) Find the average time in the line if 2, 3, and 4 clerks are used.
(b) What is the total time spent waiting in line each day if 2, 3, and 4 clerks are used?
(c) Calculate the total of the daily waiting cost and the service cost if 2, 3, and 4 clerks are used. What is the minimum total daily cost?

13-24 Billy's Bank is the only bank in a small town in Arkansas. On a typical Friday, an average of 10 customers per hour arrive at the bank to transact business. There is one single teller at the bank, and the average time required to transact business is 4 minutes. It is assumed that service times can be described by the exponential distribution. Although this is the only bank in town, some people in the town have begun using the bank in a neighboring town about 20 miles away. A single line would be used, and the customer at the front of the line would go to the first available bank teller. If a single teller at Billy's is used, find

(a) the average time in the line.
(b) the average number in the line.
(c) the average time in the system.
(d) the average number in the system.
(e) the probability that the bank is empty.

13-25 Refer to the Billy's Bank situation in Problem 13-24. Billy is considering adding a second teller (who would work at the same rate as the first) to reduce the waiting time for customers, and he assumes that this will cut the waiting time in half. If a second teller is added, find

(a) the average time in the line.
(b) the average number in the line.
(c) the average time in the system.
(d) the average number in the system.
(e) the probability that the bank is empty.

13-26 For the Billy's Bank situation in Problems 13-24 and 13-25, the salary and benefits for a teller would be \$12 per hour. The bank is open 8 hours each day. It has been estimated that the waiting time cost per hour is \$25 per hour in the line.

(a) How many customers would enter the bank on a typical day?
(b) How much total time would the customers spend waiting in line during the entire day if one teller were used? What is the total daily waiting time cost?
(c) How much total time would the customers spend waiting in line during the entire day if two tellers were used? What is the total waiting time cost?
(d) If Billy wishes to minimize the total waiting time and personnel cost, how many tellers should be used?

13-27 Customers arrive at an automated coffee vending machine at a rate of 4 per minute, following a Poisson distribution. The coffee machine dispenses a cup of coffee in exactly 10 seconds.

(a) What is the average number of people waiting in line?
(b) What is the average number in the system?
(c) How long does the average person wait in line before receiving service?

13-28 The average number of customers in the system in the single-channel, single-phase model described in Section 13.4 is

$$L = \frac{\lambda}{\mu - \lambda}$$

Show that for $m = 1$ server, the multichannel queuing model in Section 13.5,

$$L = \frac{\lambda\mu\left(\frac{\lambda}{\mu}\right)^m}{(m-1)!(m\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$$

is identical to the single-channel system. Note that the formula for P_0 (Equation 13-13) must be utilized in this highly algebraic exercise.

13-29 One mechanic services 5 drilling machines for a steel plate manufacturer. Machines break down on an average of once every 6 working days, and breakdowns tend to follow a Poisson distribution. The mechanic can handle an average of one repair job per day. Repairs follow an exponential distribution.
(a) How many machines are waiting for service, on average?
(b) How many are in the system, on average?
(c) How many drills are in running order, on average?
(d) What is the average waiting time in the queue?
(e) What is the average wait in the system?

13-30 A technician monitors a group of five computers that run an automated manufacturing facility. It takes an average of 15 minutes (exponentially distributed) to adjust a computer that develops a problem. The computers run for an average of 85 minutes (Poisson distributed) without requiring adjustments. What is the
(a) average number of computers waiting for adjustment?
(b) average number of computers not in working order?
(c) probability the system is empty?
(d) average time in the queue?
(e) average time in the system?

13-31 The typical subway station in Washington, D.C., has 6 turnstiles, each of which can be controlled by the station manager to be used for either entrance or exit control—but never for both. The manager must decide at different times of the day just how many turnstiles to use for entering passengers and how many to be set up to allow exiting passengers.

At the Washington College Station, passengers enter the station at a rate of about 84 per minute between the hours of 7 and 9 A.M. Passengers exiting trains at the stop reach the exit turnstile area at a rate of about 48 per minute during the same morning rush hours. Each turnstile can allow an average of 30 passengers per minute to enter or exit. Arrival and service times have been thought to follow Poisson and exponential distributions, respectively. Assume riders form a common queue at both entry and exit turnstile areas and proceed to the first empty turnstile.

The Washington College Station manager does not want the average passenger at his station to have to wait in a turnstile line for more than 6 seconds, nor does he want more than 8 people in any queue at any average time.
(a) How many turnstiles should be opened in each direction every morning?
(b) Discuss the assumptions underlying the solution of this problem using queuing theory.

13-32 The Clear Brook High School band is holding a car wash as a fundraiser to buy new equipment. The average time to wash a car is 4 minutes, and the time is exponentially distributed. Cars arrive at a rate of one every 5 minutes (or 12 per hour), and the number of arrivals per time period is described by the Poisson distribution.
(a) What is the average time for cars waiting in the line?
(b) What is the average number of cars in the line?
(c) What is the average number of cars in the system?
(d) What is the average time in the system?
(e) What is the probability there are more than three cars in the system?

13-33 When additional band members arrived to help at the car wash (see Problem 13-32), it was decided that two cars should be washed at a time instead of just the one. Both work crews would work at the same rate.
(a) What is the average time for cars waiting in the line?
(b) What is the average number of cars in the line?
(c) What is the average number of cars in the system?
(d) What is the average time in the system?

Internet Homework Problems

See our Internet home page, at **www.pearsonhighered.com/render**, for additional homework problems, Problems 13-34 to 13-38.

Case Study

New England Foundry

For more than 75 years, New England Foundry, Inc., has manufactured wood stoves for home use. In recent years, with increasing energy prices, George Mathison, president of New England Foundry, has seen sales triple. This dramatic increase in sales has made it even more difficult for George to maintain quality in all the wood stoves and related products.

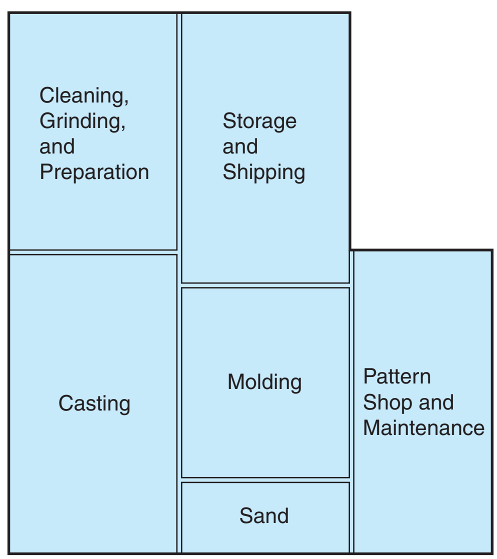
Unlike other companies manufacturing wood stoves, New England Foundry is *only* in the business of making stoves and stove-related products. Their major products are the Warmglo I, the Warmglo II, the Warmglo III, and the Warmglo IV. The Warmglo I is the smallest wood stove, with a heat output of 30,000 Btu, and the Warmglo IV is the largest, with a heat output of 60,000 Btu. In addition, New England Foundry, Inc., produces a large array of products that have been designed to be used with one of their four stoves, including warming shelves, surface thermometers, stovepipes, adaptors, stove gloves, trivets, mitten racks, andirons, chimneys, and heat shields. New England Foundry also publishes a newsletter and several paperback books on stove installation, stove operation, stove maintenance, and wood sources. It is George's belief that its wide assortment of products was a major contributor to the sales increases.

The Warmglo III outsells all the other stoves by a wide margin. The heat output and available accessories are ideal for the typical home. The Warmglo III also has a number of outstanding features that make it one of the most attractive and heat-efficient stoves on the market. Each Warmglo III has a thermostatically controlled primary air intake valve that allows the stove to adjust itself automatically to produce the correct heat output for varying weather conditions. A secondary air opening is used to increase the heat output in case of very cold weather. The internal stove parts produce a horizontal flame path for more efficient burning, and the output gases are forced to take an S-shaped path through the stove. The S-shaped path allows more complete combustion of the gases and better heat transfer from the fire and gases through the cast iron to the area to be heated. These features, along with the accessories, resulted in expanding sales and prompted George to build a new factory to manufacture Warmglo III stoves. An overview diagram of the factory is shown in Figure 13.3.

The new foundry uses the latest equipment, including a new Disamatic that helps in manufacturing stove parts. Regardless of new equipment or procedures, casting operations have remained basically unchanged for hundreds of years. To begin with, a wooden pattern is made for every cast-iron piece in the stove. The wooden pattern is an exact duplication of the cast-iron piece that is to be manufactured. New England Foundry has all of its patterns made by Precision Patterns, Inc., and these patterns are stored in the pattern shop and maintenance room. Then a specially formulated sand is molded around the wooden pattern. There can be two or more sand molds for each pattern. Mixing the sand and making the molds are done in the molding room. When the wooden pattern is removed, the resulting sand molds form a negative image of the desired casting. Next, the molds are transported to the casting room, where molten iron is poured into the molds and allowed to cool. When the iron has solidified, the molds are moved into the cleaning, grinding, and preparation room. The molds are dumped into large vibrators that shake most of the sand from the casting. The rough castings are then subjected to both sandblasting to remove the rest of the sand and grinding to finish some of the surfaces of the castings. The castings are then painted with a special heat-resistant paint, assembled into workable stoves, and inspected for manufacturing defects that may have gone undetected thus far. Finally, the finished stoves are moved to storage and shipping, where they are packaged and shipped to the appropriate locations.

At present, the pattern shop and the maintenance department are located in the same room. One large counter is used by both maintenance personnel to get tools and parts and by sand molders that need various patterns for the molding operation. Peter Nawler and Bob Bryan, who work behind the counter, are able to service a total of 10 people per hour (or about 5 per hour each). On the average, 4 people from maintenance and 3 people from the molding department arrive at the counter per hour. People from the molding department and from maintenance arrive randomly, and to be served they form a single line. Pete and Bob have always had a policy of first come, first served. Because of the location of the pattern shop and maintenance department, it takes about 3 minutes for a person from the maintenance department to walk to the pattern and maintenance room, and it takes about 1 minute for a person to walk from the molding department to the pattern and maintenance room.

FIGURE 13.3 Overview of Factory



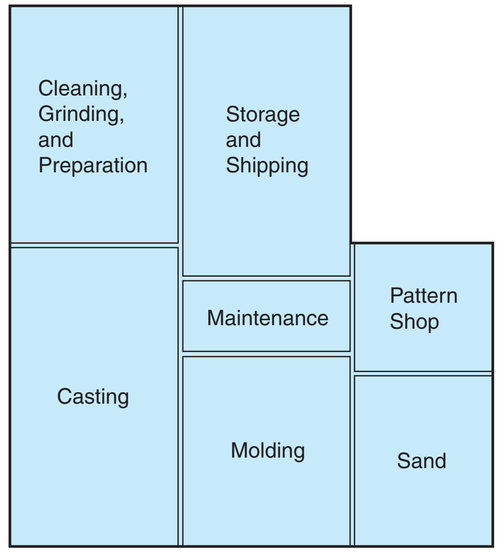
After observing the operation of the pattern shop and maintenance room for several weeks, George decided to make some changes to the layout of the factory. An overview of these changes is shown in Figure 13.4.

Separating the maintenance shop from the pattern shop had a number of advantages. It would take people from the maintenance department only 1 minute instead of 3 to get to the new maintenance department. Using time and motion studies, George was also able to determine that improving the layout of the maintenance department would allow Bob to serve 6 people from the maintenance department per hour, and improving the layout of the pattern department would allow Pete to serve 7 people from the molding shop per hour.

Discussion Questions

1. How much time would the new layout save?
2. If maintenance personnel were paid $9.50 per hour and molding personnel were paid $11.75 per hour, how much could be saved per hour with the new factory layout?

FIGURE 13.4 Overview of Factory After Changes



Case Study

Winter Park Hotel

Donna Shader, manager of the Winter Park Hotel, is considering how to restructure the front desk to reach an optimum level of staff efficiency and guest service. At present, the hotel has five clerks on duty, each with a separate waiting line, during the peak check-in time of 3:00 P.M. to 5:00 P.M. Observation of arrivals during this time show that an average of 90 guests arrive each hour (although there is no upward limit on the number that could arrive at any given time). It takes an average of 3 minutes for the front-desk clerk to register each guest.

Donna is considering three plans for improving guest service by reducing the length of time guests spend waiting in line. The first proposal would designate one employee as a quick-service clerk for guests registering under corporate accounts, a market segment that fills about 30% of all occupied rooms. Because corporate guests are preregistered, their registration takes just 2 minutes. With these guests separated from the rest of the clientele, the average time for registering a typical guest would climb to 3.4 minutes. Under plan 1, noncorporate guests would choose any of the remaining four lines.

The second plan is to implement a single-line system. All guests could form a single waiting line to be served by whichever of five clerks became available. This option would require sufficient lobby space for what could be a substantial queue.

The third proposal involves using an automatic teller machine (ATM) for check-ins. This ATM would provide approximately the same service rate as a clerk would. Given that initial use of this technology might be minimal, Shader estimated that 20% of customers, primarily frequent guests, would be willing to use the machines. (This might be a conservative estimate if the guests perceive direct benefits from using the ATM, as bank customers do. Citibank reports that some 95% of its Manhattan customers use its ATMs.) Donna would set up a single queue for customers who prefer human check-in clerks. This would be served by the five clerks, although Donna is hopeful that the machine will allow a reduction to four.

Discussion Questions

1. Determine the average amount of time that a guest spends checking in. How would this change under each of the stated options?
2. Which option do you recommend?

Internet Case Study

See our Internet home page, at **www.pearsonhighered.com/render**, for this additional case study: Pantry Shopper. This case involves providing better service in a grocery store.

Bibliography

Baldwin, Rusty O., Nathaniel J. Davis IV, Scott F. Midkiff, and John E. Kobza. "Queueing Network Analysis: Concepts, Terminology, and Methods," *Journal of Systems & Software* 66, 2 (2003): 99–118.

Barron, K. "Hurry Up and Wait," *Forbes* (October 16, 2000): 158–164.

Cayirli, Tugba, and Emre Veral. "Outpatient Scheduling in Health Care: A Review of Literature," *Production & Operations Management* 12, 4 (2003): 519–549.

Cooper, R. B. *Introduction to Queuing Theory*, 2nd ed. New York: Elsevier—North Holland, 1980.

de Bruin, Arnoud M., A. C. van Rossum, M. C. Visser, and G. M. Koole. "Modeling the Emergency Cardiac Inpatient Flow: An Application of Queuing Theory," *Health Care Management Science* 10, 2 (2007): 125–137.

Derbala, Ali. "Priority Queuing in an Operating System," *Computers & Operations Research* 32, 2(2005): 229–238.

Grassmann, Winfried K. "Finding the Right Number of Servers in Real-World Queuing Systems," *Interfaces* 18, 2 (March–April 1988): 94–104.

Janic, Milan. "Modeling Airport Congestion Charges," *Transportation Planning & Technology* 28, 1 (2005): 1–26.

Katz, K., B. Larson, and R. Larson. "Prescription for the Waiting-in-Line Blues," *Sloan Management Review* (Winter 1991): 44–53.

Koizumi, Naoru, Eri Kuno, and Tony E. Smith. "Modeling Patient Flows Using a Queuing Network with Blocking," *Health Care Management Science* 8, 1 (2005): 49–60.

Larson, Richard C. "Perspectives on Queues: Social Justice and the Psychology of Queuing," *Operations Research* 35, 6 (November–December 1987): 895–905.

Murtojärvi, Mika, et al. "Determining the Proper Number and Price of Software Licenses," *IEEE Transactions on Software Engineering* 33, 5 (2007): 305–315.

Prabhu, N. U. *Foundations of Queuing Theory*. Norwell, MA: Kluewer Academic Publishers, 1997.

Regattieri, A., R. Gamberini, F. Lolli, and R. Manzini. "Designing Production and Service Systems Using Queuing Theory: Principles and Application to an Airport Passenger Security Screening System," *International Journal of Services and Operations Management* 6, 2 (2010): 206–225.

Tarko, A. P. "Random Queues in Signalized Road Networks," *Transportation Science* 34, 4 (November 2000): 415–425.

Appendix 13.1 Using QM for Windows

For all these problems, from the Module menu, select *Waiting Lines* and then select *New* to enter a new problem. Then select the type of model you want to use from the ones that appear.

This appendix illustrates the ease of use of the QM for Windows in solving queuing problems. Program 13.5 represents the Arnold's Muffler Shop analysis with 2 servers. The only required inputs are selection of the proper model, a title, whether to include costs, the time units being used for arrival and service rates (hours in this example), the arrival rate (2 cars per hour), the service rate (3 cars per hour), and the number of servers (2). Because the time units are specified as hours, W and W_q are given in hours, but they are also converted into minutes and seconds, as seen in Program 13.5.

Program 13.6 reflects a constant service time model, illustrated in the chapter by Garcia-Golding Recycling, Inc. The other queuing models can also be solved by QM for Windows, which additionally provides cost/economic analysis.

PROGRAM 13.5

Using QM for Windows to Solve a Multichannel Queuing Model (Arnold Muffler Shop Data)

Cost analysis: (•) No costs ( ) Use Costs — Time unit (arrival, service rate): hours

Waiting Lines Results

(untitled) Solution

Parameter	Value	Parameter	Value	Minutes	Seconds
M/M/s		Average server utilization	0.3333		
Arrival rate(lambda)	2.	Average number in the queue(Lq)	0.0833		
Service rate(mu)	3.	Average number in the system(Ls)	0.75		
Number of servers	2.	Average time in the queue(Wq)	0.0417	2.5	150.
		Average time in the system(Ws)	0.375	22.5	1,350.

PROGRAM 13.6

Using QM for Windows to Solve a Constant Service Time Model (Garcia-Golding Data)

Cost analysis: (•) No costs ( ) Use Costs — Time unit (arrival, service rate): hours

Waiting Lines Results

Garcia-Golding Recycling, Inc. Solution

Parameter	Value	Parameter	Value	Minutes	Seconds
Constant service times		Average server utilization	0.6667		
Arrival rate(lambda)	8.	Average number in the queue(Lq)	0.6667		
Service rate(mu)	12.	Average number in the system(Ls)	1.3333		
Number of servers	1.	Average time in the queue(Wq)	0.0833	5.	300.
		Average time in the system(Ws)	0.1667	10.	600.