CHAPTER 15

Waiting Line Models

CONTENTS

Recall the last time that you had to wait at a supermarket checkout counter, for a teller at your local bank, or to be served at a fast-food restaurant. In these and many other waiting line situations, the time spent waiting is undesirable. Adding more checkout clerks, bank tellers, or servers is not always the most economical strategy for improving service, so businesses need to determine ways to keep waiting times within tolerable limits.

Models have been developed to help managers understand and make better decisions concerning the operation of waiting lines. In management science terminology, a waiting line is also known as a **queue**, and the body of knowledge dealing with waiting lines is known as **queueing theory**. In the early 1900s, A. K. Erlang, a Danish telephone engineer, began a study of the congestion and waiting times occurring in the completion of telephone calls. Since then, queueing theory has grown far more sophisticated, with applications in a wide variety of waiting line situations.

Waiting line models consist of mathematical formulas and relationships that can be used to determine the **operating characteristics** (performance measures) for a waiting line. Operating characteristics of interest include these:

1. The probability that no units are in the system (i.e., the system is idle)
2. The average number of units in the waiting line
3. The average number of units in the system (the number of units in the waiting line plus the number of units being served)
4. The average time a unit spends in the waiting line
5. The average time a unit spends in the system (the waiting time plus the service time)
6. The probability that an arriving unit has to wait for service

Managers who have such information are better able to make decisions that balance desirable service levels against the cost of providing the service.

The Q.M. in Action, ATM Waiting Times at Citibank, describes how a waiting line model was used to help determine the number of automatic teller machines to place at New York City banking centers. A waiting line model prompted the creation of a new kind of line and a chief line director to implement first-come, first-served queue discipline at Whole Foods Market in the Chelsea neighborhood of New York City. In addition, a waiting line model helped the New Haven, Connecticut, fire department develop policies to improve response time for both fire and medical emergencies.

The waiting line model used at Citibank is discussed in Section 15.3

Q.M. *in* ACTION

*ATM WAITING TIMES AT CITIBANK**

The New York City franchise of U.S. Citibanking operates more than 250 banking centers. Each center provides one or more automatic teller machines (ATMs) capable of performing a variety of banking transactions. At each center, a waiting line is formed by randomly arriving customers who seek service at one of the ATMs.

In order to make decisions on the number of ATMs to have at selected banking center locations, management needed information about potential waiting times and general customer service. Waiting line operating characteristics such as average number of customers in the waiting line, average time a customer spends waiting, and the probability that an arriving customer has to wait would help management determine the number of ATMs to recommend at each banking center.

For example, one busy Midtown Manhattan center had a peak arrival rate of 172 customers per hour. A multiple-server waiting line model with

*Based on information provided by Stacey Karter of Citibank.

(*continued*)

six ATMs showed that 88% of the customers would have to wait, with an average wait time between six and seven minutes. This level of service was judged unacceptable. Expansion to seven ATMs was recommended for this location based on the waiting line model's projection of acceptable waiting times. Use of the waiting line model provided guidelines for making incremental ATM decisions at each banking center location.

15.1 Structure of a Waiting Line System

To illustrate the basic features of a waiting line model, we consider the waiting line at the Burger Dome fast-food restaurant. Burger Dome sells hamburgers, cheeseburgers, french fries, soft drinks, and milk shakes, as well as a limited number of specialty items and dessert selections. Although Burger Dome would like to serve each customer immediately, at times more customers arrive than can be handled by the Burger Dome food service staff. Thus, customers wait in line to place and receive their orders.

Burger Dome is concerned that the methods currently used to serve customers are resulting in excessive waiting times and a possible loss of sales. Management wants to conduct a waiting line study to help determine the best approach to reduce waiting times and improve service.

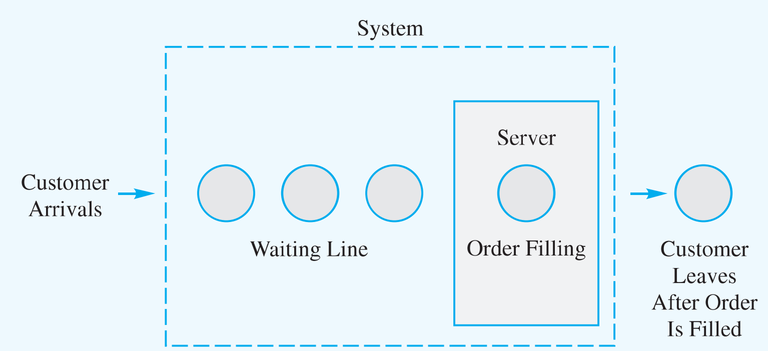
Single-Server Waiting Line

In the current Burger Dome operation, an employee takes a customer's order, determines the total cost of the order, receives payment from the customer, and then fills the order. Once the first customer's order is filled, the employee takes the order of the next customer waiting for service. This operation is an example of a **single-server waiting line**. Each customer entering the Burger Dome restaurant is served by a single order-filling station that handles order placement, bill payment, and food delivery. When more customers arrive than can be served immediately, they form a waiting line and wait for the order-filling station to become available. A diagram of the Burger Dome single-server waiting line is shown in Figure 15.1.

Distribution of Arrivals

Defining the arrival process for a waiting line involves determining the probability distribution for the number of arrivals in a given period of time. For many waiting line situations,

FIGURE 15.1 THE BURGER DOME SINGLE-SERVER WAITING LINE



the arrivals occur *randomly and independently* of other arrivals, and we cannot predict when an arrival will occur. In such cases, analysts have found that the **Poisson probability distribution** provides a good description of the arrival pattern.

The Poisson probability function provides the probability of x arrivals in a specific time period. The probability function is as follows:[1]

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x = 0, 1, 2, \ldots \tag{15.1}$$

where

x = the number of arrivals in the time period

λ = the *mean* number of arrivals per time period

e = 2.71828

The mean number of arrivals per time period, λ, is called the **arrival rate**. Values of $e^{-\lambda}$ can be found using a calculator or by using Appendix E.

Suppose that Burger Dome analyzed data on customer arrivals and concluded that the arrival rate is 45 customers per hour. For a one-minute period, the arrival rate would be λ = 45 customers/60 minutes = 0.75 customers per minute. Thus, we can use the following Poisson probability function to compute the probability of x customer arrivals during a one-minute period:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{0.75^x e^{-0.75}}{x!} \tag{15.2}$$

Thus, the probabilities of 0, 1, and 2 customer arrivals during a one-minute period are

$$P(0) = \frac{(0.75)^0 e^{-0.75}}{0!} = e^{-0.75} = 0.4724$$

$$P(1) = \frac{(0.75)^1 e^{-0.75}}{1!} = 0.75e^{-0.75} = 0.75(0.4724) = 0.3543$$

$$P(2) = \frac{(0.75)^2 e^{-0.75}}{2!} = \frac{(0.5625)(0.4724)}{2} = 0.1329$$

The probability of no customers in a one-minute period is 0.4724, the probability of one customer in a one-minute period is 0.3543, and the probability of two customers in a one-minute period is 0.1329. Table 15.1 shows the Poisson probabilities for customer arrivals during a one-minute period.

The waiting line models that will be presented in Sections 15.2 and 15.3 use the Poisson probability distribution to describe the customer arrivals at Burger Dome. In practice, you should record the actual number of arrivals per time period for several days or weeks and compare the frequency distribution of the observed number of arrivals to the Poisson probability distribution to determine whether the Poisson probability distribution provides a reasonable approximation of the arrival distribution.

[1]The term $x!$, x *factorial*, is defined as $x! = x(x-1)(x-2)\ldots(2)(1)$. For example, $4! = (4)(3)(2)(1) = 24$. For the special case of $x = 0$, $0! = 1$ by definition.

TABLE 15.1 POISSON PROBABILITIES FOR THE NUMBER OF CUSTOMER ARRIVALS AT A BURGER DOME RESTAURANT DURING A ONE-MINUTE PERIOD ($\lambda = 0.75$)

Number of Arrivals	Probability
0	0.4724
1	0.3543
2	0.1329
3	0.0332
4	0.0062
5 or more	0.0010

Distribution of Service Times

The service time is the time a customer spends at the service facility once the service has started. At Burger Dome, the service time starts when a customer begins to place the order with the employee and continues until the customer receives the order. Service times are rarely constant. At Burger Dome, the number of items ordered and the mix of items ordered vary considerably from one customer to the next. Small orders can be handled in a matter of seconds, but large orders may require more than two minutes.

If the probability distribution for the service time can be assumed to follow an **exponential probability distribution**, formulas are available for providing useful information about the operation of the waiting line. Using an exponential probability distribution, the probability that the service time will be less than or equal to a time of length t is

$$P(\text{service time} \leq t) = 1 - e^{-\mu t} \tag{15.3}$$

where

μ = the *mean* number of units that can be served per time period

$e = 2.71828$

The mean number of units that can be served per time period, μ, is called the **service rate**.

A property of the exponential probability distribution is that there is a 0.6321 probability that the random variable takes on a value less than its mean. In waiting line applications, the exponential probability distribution indicates that approximately 63% of the service times are less than the mean service time and approximately 37% of the service times are greater than the mean service time.

Suppose that Burger Dome studied the order-filling process and found that a single employee can process an average of 60 customer orders per hour. On a one-minute basis, the service rate would be $\mu = 60$ customers/60 minutes $= 1$ customer per minute. For example, with $\mu = 1$, we can use equation (15.3) to compute probabilities such as the probability that an order can be processed in $^1/_2$ minute or less, 1 minute or less, and 2 minutes or less. These computations are

$$P(\text{service time} \leq 0.5 \text{ min.}) = 1 - e^{-1(0.5)} = 1 - 0.6065 = 0.3935$$
$$P(\text{service time} \leq 1.0 \text{ min.}) = 1 - e^{-1(1.0)} = 1 - 0.3679 = 0.6321$$
$$P(\text{service time} \leq 2.0 \text{ min.}) = 1 - e^{-1(2.0)} = 1 - 0.1353 = 0.8647$$

Thus, we would conclude that there is a 0.3935 probability that an order can be processed in $^1/_2$ minute or less, a 0.6321 probability that it can be processed in 1 minute or less, and a 0.8647 probability that it can be processed in 2 minutes or less.

In several waiting line models presented in this chapter, we assume that the probability distribution for the service time follows an exponential probability distribution. In practice, you should collect data on actual service times to determine whether the exponential probability distribution is a reasonable approximation of the service times for your application.

Queue Discipline

In describing a waiting line system, we must define the manner in which the waiting units are arranged for service. For the Burger Dome waiting line, and in general for most customer-oriented waiting lines, the units waiting for service are arranged on a **first-come, first-served** basis; this approach is referred to as an **FCFS** queue discipline. However, some situations call for different queue disciplines. For example, when people board an airplane, the last passengers to board are typically the first to deplane since many airlines have the passengers with seat assignments in the back of the plane board first. On the other hand, it does not seem prudent for hospital emergency rooms to operate under either of these queue disciplines, and so we have other types of queue disciplines that assign priorities to the waiting units and then serve the unit with the highest priority first. In this chapter we consider only waiting lines based on a first-come, first-served queue discipline. The Q.M. in Action, The Serpentine Line and an FCFS Queue Discipline at Whole Foods Market, describes how an FCFS queue discipline is used at a supermarket.

Steady-State Operation

When the Burger Dome restaurant opens in the morning, no customers are in the restaurant, and the characteristics of the waiting line system fluctuate depending on realized arrival and service times. Gradually, activity builds up to a normal or steady state. The beginning or startup period is referred to as the **transient period**. The transient period ends when the system reaches the normal or **steady-state operation**. Waiting line models describe the steady-state operating characteristics of a waiting line.

Q.M. *in* ACTION

*THE SERPENTINE LINE AND AN FCFS QUEUE DISCIPLINE AT WHOLE FOODS MARKET**

The Whole Foods Market in the Chelsea neighborhood of New York City employs a chief line director to implement a FCFS queue discipline. Companies such as Wendy's, American Airlines, and Chemical Bank were among the first to employ serpentine lines to implement an FCFS queue discipline. Such lines are commonplace today. We see them at banks, amusement parks, and fast-food outlets. The line is called *serpentine* because of the way it winds around. When a customer gets to the front of the line, the customer then goes to the first available server. People like serpentine lines because they prevent people who join the line later from being served ahead of an earlier arrival.

As popular as serpentine lines have become, supermarkets have not employed them because of a lack of space. At the typical supermarket, a separate line forms at each checkout counter. When ready to check out, a person picks one of the checkout counters and stays in that line until receiving service. Sometimes a person joining another checkout line later will receive service first, which tends to upset people. Manhattan's Whole Foods Market solved this problem by creating a new kind of line and employing a chief line director to direct the first person in line to the next available checkout counter.

The waiting line at the Whole Foods Market is actually three parallel lines. Customers join the shortest line and follow a rotation when they reach the front of the line. For instance, if the first customer in line 1 is sent to a checkout counter, the next customer sent to a checkout counter is the first person in line 2, then the first person in line 3, and so on. This way an FCFS queue discipline is implemented without a long, winding serpentine line.

The Whole Foods Market's customers seem to really like the system, and the line director, Bill Jones, has become something of a celebrity. Children point to him on the street and customers invite him over for dinner.

*Based on I. Parker, "Mr. Next," *The New Yorker* (January 13, 2003).

15.2 Single-Server Waiting Line Model with Poisson Arrivals and Exponential Service Times

In this section we present formulas that can be used to determine the steady-state operating characteristics for a single-server waiting line. The formulas are applicable if the arrivals follow a Poisson probability distribution and the service times follow an exponential probability distribution. As these assumptions apply to the Burger Dome waiting line problem introduced in Section 15.1, we show how formulas can be used to determine Burger Dome's operating characteristics and thus provide management with helpful decision-making information.

Waiting line models are often based on assumptions such as Poisson arrivals and exponential service times. When applying any waiting line model, data should be collected on the actual system to ensure that the assumptions of the model are reasonable.

The mathematical methodology used to derive the formulas for the operating characteristics of waiting lines is rather complex. However, our purpose in this chapter is not to provide the theoretical development of waiting line models, but rather to show how the formulas that have been developed can provide information about operating characteristics of the waiting line. Readers interested in the mathematical development of the formulas can consult the specialized texts listed in Appendix F at the end of the text.

Operating Characteristics

The following formulas can be used to compute the steady-state operating characteristics for a single-server waiting line with Poisson arrivals and exponential service times, where

λ = the mean number of arrivals per time period (the arrival rate)

μ = the mean number of services per time period (the service rate)

Equations (15.4) through (15.10) do not provide formulas for optimal conditions. Rather, these equations provide information about the steady-state operating characteristics of a waiting line.

1. The probability that no units are in the system:

$$P_0 = 1 - \frac{\lambda}{\mu} \tag{15.4}$$

2. The average number of units in the waiting line:

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{15.5}$$

3. The average number of units in the system:

$$L = L_q + \frac{\lambda}{\mu} \tag{15.6}$$

4. The average time a unit spends in the waiting line:

$$W_q = \frac{L_q}{\lambda} \tag{15.7}$$

5. The average time a unit spends in the system:

$$W = W_q + \frac{1}{\mu} \qquad \textbf{(15.8)}$$

6. The probability that an arriving unit has to wait for service:

$$P_w = \frac{\lambda}{\mu} \qquad \textbf{(15.9)}$$

7. The probability of n units in the system:

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 \qquad \textbf{(15.10)}$$

The values of the arrival rate λ and the service rate μ are clearly important components in determining the operating characteristics. Equation (15.9) shows that the ratio of the arrival rate to the service rate, λ/μ, provides the probability that an arriving unit has to wait because the service facility is in use. Hence, λ/μ is referred to as the *utilization factor* for the service facility.

The operating characteristics presented in equations (15.4) through (15.10) are applicable only when the service rate μ is *greater than* the arrival rate λ—in other words, when $\lambda/\mu < 1$. If this condition does not exist, the waiting line will continue to grow without limit because the service facility does not have sufficient capacity to handle the arriving units. Thus, in using equations (15.4) through (15.10), we must have $\mu > \lambda$.

Operating Characteristics for the Burger Dome Problem

Recall that for the Burger Dome problem we had an arrival rate of $\lambda = 0.75$ customers per minute and a service rate of $\mu = 1$ customer per minute. Thus, with $\mu > \lambda$, equations (15.4) through (15.10) can be used to provide operating characteristics for the Burger Dome single-server waiting line:

Problem 5 asks you to compute the operating characteristics for a single-server waiting line application.

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{0.75}{1} = 0.25$$

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{0.75^2}{1(1 - 0.75)} = 2.25 \text{ customers}$$

$$L = L_q + \frac{\lambda}{\mu} = 2.25 + \frac{0.75}{1} = 3 \text{ customers}$$

$$W_q = \frac{L_q}{\lambda} = \frac{2.25}{0.75} = 3 \text{ minutes}$$

$$W = W_q + \frac{1}{\mu} = 3 + \frac{1}{1} = 4 \text{ minutes}$$

$$P_w = \frac{\lambda}{\mu} = \frac{0.75}{1} = 0.75$$

Equation (15.10) can be used to determine the probability of any number of customers in the system. Applying this equation provides the probability information in Table 15.2.

TABLE 15.2 THE PROBABILITY OF *n* CUSTOMERS IN THE SYSTEM FOR THE BURGER DOME WAITING LINE PROBLEM

Number of Customers	Probability
0	0.2500
1	0.1875
2	0.1406
3	0.1055
4	0.0791
5	0.0593
6	0.0445
7 or more	0.1335

Managers' Use of Waiting Line Models

The results of the single-server waiting line for Burger Dome show several important things about the operation of the waiting line. In particular, customers wait an average of three minutes before beginning to place an order, which appears somewhat long for a business based on fast service. In addition, the facts that the average number of customers waiting in line is 2.25 and that 75% of the arriving customers have to wait for service are indicators that something should be done to improve the waiting line operation. Table 15.2 shows a 0.1335 probability that seven or more customers are in the Burger Dome system at one time. This condition indicates a fairly high probability that Burger Dome will experience some long waiting lines if it continues to use the single-server operation.

If the operating characteristics are unsatisfactory in terms of meeting company standards for service, Burger Dome's management should consider alternative designs or plans for improving the waiting line operation.

Improving the Waiting Line Operation

Waiting line models often indicate when improvements in operating characteristics are desirable. However, the decision of how to modify the waiting line configuration to improve the operating characteristics must be based on the insights and creativity of the analyst.

After reviewing the operating characteristics provided by the waiting line model, Burger Dome's management concluded that improvements designed to reduce waiting times were desirable. To make improvements in the waiting line operation, analysts often focus on ways to improve the service rate. Generally, service rate improvements are obtained by making either or both of the following changes:

1. Increase the service rate by making a creative design change or by using new technology.
2. Add one or more servers so that more customers can be served simultaneously.

Assume that in considering alternative 1, Burger Dome's management decides to employ a design change that allows the customer to fill out and submit a paper order form directly to the kitchen while they are waiting in line. This allows the customer's food to be ready by the time the employee collects payment from the customer. With this design, Burger Dome's management estimates the service rate can be increased from the current

TABLE 15.3 OPERATING CHARACTERISTICS FOR THE BURGER DOME SYSTEM WITH THE SERVICE RATE INCREASED TO $\mu = 1.25$ CUSTOMERS PER MINUTE

Probability of no customers in the system	0.400
Average number of customers in the waiting line	0.900
Average number of customers in the system	1.500
Average time in the waiting line	1.200 minutes
Average time in the system	2.000 minutes
Probability that an arriving customer has to wait	0.600
Probability that seven or more customers are in the system	0.028

60 customers per hour to 75 customers per hour. Thus, the service rate for the revised system is $\mu = 75$ customers/60 minutes = 1.25 customers per minute. For $\lambda = 0.75$ customers per minute and $\mu = 1.25$ customers per minute, equations (15.4) through (15.10) can be used to provide the new operating characteristics for the Burger Dome waiting line. These operating characteristics are summarized in Table 15.3.

The information in Table 15.3 indicates that all operating characteristics have improved because of the increased service rate. In particular, the average time a customer spends in the waiting line has been reduced from 3 to 1.2 minutes, and the average time a customer spends in the system has been reduced from 4 to 2 minutes. Are any other alternatives available that Burger Dome can use to increase the service rate? If so, and if the mean service rate μ can be identified for each alternative, equations (15.4) through (15.10) can be used to determine the revised operating characteristics and any improvements in the waiting line system. The added cost of any proposed change can be compared to the corresponding service improvements to help the manager determine whether the proposed service improvements are worthwhile.

Problem 11 asks you to determine whether a change in the service rate will meet the company's service guideline for its customers.

As mentioned previously in alternative 2, another option often available is to add one or more servers so that orders for multiple customers can be filled simultaneously. The extension of the single-server waiting line model to the multiple-server waiting line model is the topic of the next section.

Excel Solution of Waiting Line Model

Waiting line models are easily implemented with the aid of spreadsheets. The Excel worksheet for the Burger Dome single-server waiting line is shown in Figure 15.2. The arrival rate and the service rate are entered in cells B7 and B8. The formulas for the waiting line's operating characteristics are placed in cells C13 to C18. The worksheet computes the same values for the operating characteristics that we obtained earlier. Modifications in the waiting line design can be evaluated by entering different arrival rates and/or service rates into cells B7 and B8. The new operating characteristics of the waiting line will be shown immediately.

The Excel worksheet in Figure 15.2 is a template that can be used with any single-server waiting line model with Poisson arrivals and exponential service times. This worksheet and similar Excel worksheets for the other waiting line models presented in this chapter are available at the WEBfiles link on the website that accompanies this text.

FIGURE 15.2 FORMULA WORKSHEET FOR THE BURGER DOME SINGLE-SERVER WAITING LINE

	A	B	C
1	**Single-Server Waiting Line Model**		
2			
3	**Assumptions**		
4	**Poisson Arrivals**		
5	**Exponential Service Times**		
6			
7	Arrival Rate	0.75	
8	Service Rate	1	
9			
10			
11	**Operating Characteristics**		
12			
13	Probability that no customers are in the system, P_0		=1-(B7/B8)
14	Average number of customers in the waiting line, L_q		=(B7^2)/(B8*(B8-B7))
15	Average number of customers in the system, L		=C14+(B7/B8)
16	Average time a customer spends in the waiting line, W_q		=C14/B7
17	Average time a customer spends in the system, W		=C16+(1/B8)
18	Probability an arriving customer has to wait, P_w		=B7/B8
19			

WEB file

Single

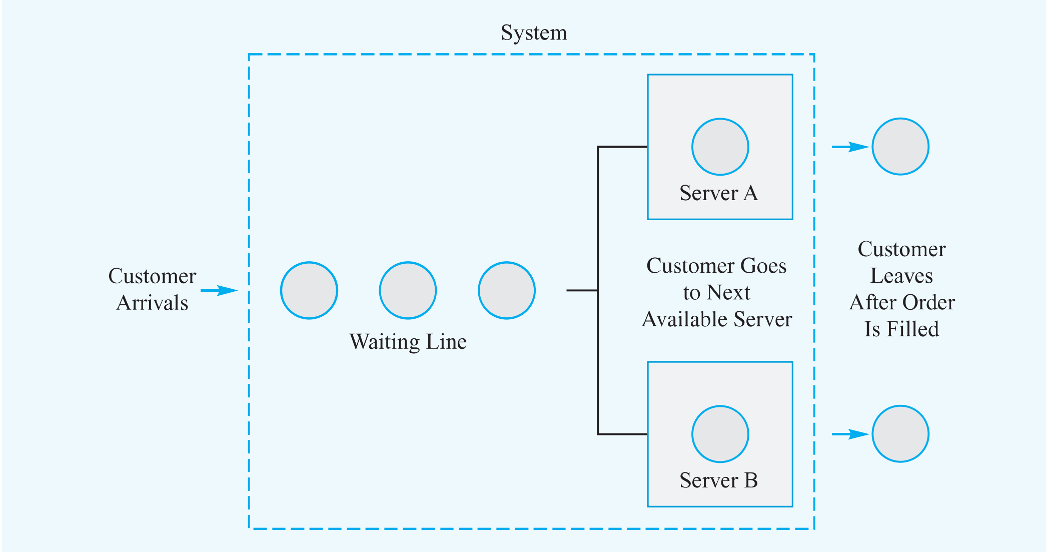
NOTES AND COMMENTS

1. The assumption that arrivals follow a Poisson probability distribution is equivalent to the assumption that the time between arrivals has an exponential probability distribution. For example, if the arrivals for a waiting line follow a Poisson probability distribution with a mean of 20 arrivals per hour, the time between arrivals will follow an exponential probability distribution, with a mean time between arrivals of $^1/_{20}$ or 0.05 hour.
2. Many individuals believe that whenever the service rate μ is greater than the arrival rate λ, the system should be able to handle or serve all arrivals without any customer waiting for service. This would be true if the time between customer arrivals was constant and the service time was constant. However, as the Burger Dome example shows, the variability of arrival times and service times may result in long waiting times even when the service rate exceeds the arrival rate. A contribution of waiting line models is that they can point out undesirable waiting line operating characteristics even when the $\mu > \lambda$ condition appears satisfactory.

15.3 Multiple-Server Waiting Line Model with Poisson Arrivals and Exponential Service Times

A **multiple-server waiting line** consists of two or more servers that are assumed to be identical in terms of service capability. For multiple-server systems, there are two typical queueing possibilities: (1) arriving customers wait in a single waiting line (called a "pooled" or "shared" queue) and then move to the first available server for processing, or (2) each server

FIGURE 15.3 THE BURGER DOME TWO-SERVER WAITING LINE



has a "dedicated" queue and an arriving customer selects one of these lines to join (and typically is not allowed to switch lines). In this chapter, we focus on the system design with a single shared waiting line for all servers. Operating characteristics for a multiple-server system are typically better when a single shared queue, rather than multiple dedicated waiting lines, is used. The single-server Burger Dome operation can be expanded to a two-server system by opening a second server. Figure 15.3 shows a diagram of the Burger Dome two-server waiting line.

In this section we present formulas that can be used to determine the steady-state operating characteristics for a multiple-server waiting line. These formulas are applicable if the following conditions exist:

1. The arrivals follow a Poisson probability distribution.
2. The service time for each server follows an exponential probability distribution.
3. The service rate μ is the same for each server.
4. The arrivals wait in a single waiting line and then move to the first open server for service.

Operating Characteristics

The following formulas can be used to compute the steady-state operating characteristics for multiple-server waiting lines, where

λ = the arrival rate for the system
μ = the service rate for *each* server
k = the number of servers

1. The probability that no units are in the system:

$$P_0 = \frac{1}{\sum_{n=0}^{k-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^k}{k!}\left(\frac{k\mu}{k\mu - \lambda}\right)} \quad \textbf{(15.11)}$$

2. The average number of units in the waiting line:

$$L_q = \frac{(\lambda/\mu)^k \lambda\mu}{(k-1)!(k\mu - \lambda)^2} P_0 \tag{15.12}$$

3. The average number of units in the system:

$$L = L_q + \frac{\lambda}{\mu} \tag{15.13}$$

4. The average time a unit spends in the waiting line:

$$W_q = \frac{L_q}{\lambda} \tag{15.14}$$

5. The average time a unit spends in the system:

$$W = W_q + \frac{1}{\mu} \tag{15.15}$$

6. The probability that an arriving unit has to wait for service:

$$P_w = \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \left(\frac{k\mu}{k\mu - \lambda}\right) P_0 \tag{15.16}$$

7. The probability of n units in the system:

$$P_n = \frac{(\lambda/\mu)^n}{n!} P_0 \qquad \text{for } n \leq k \tag{15.17}$$

$$P_n = \frac{(\lambda/\mu)^n}{k!k^{(n-k)}} P_0 \qquad \text{for } n > k \tag{15.18}$$

Because μ is the service rate for each server, $k\mu$ is the service rate for the multiple-server system. As was true for the single-server waiting line model, the formulas for the operating characteristics of multiple-server waiting lines can be applied only in situations where the service rate for the system exceeds the arrival rate for the system; in other words, the formulas are applicable only if $k\mu$ is greater than λ.

Some expressions for the operating characteristics of multiple-server waiting lines are more complex than their single-server counterparts. However, equations (15.11) through (15.18) provide the same information as provided by the single-server model. To help simplify the use of the multiple-server equations, Table 15.4 contains values of P_0 for selected values of λ/μ and k. The values provided in the table correspond to cases where $k\mu > \lambda$, and hence the service rate is sufficient to process all arrivals.

TABLE 15.4 VALUES OF P_0 FOR MULTIPLE-SERVER WAITING LINES WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES

	Number of Servers (k)			
Ratio λ/μ	2	3	4	5
0.15	0.8605	0.8607	0.8607	0.8607
0.20	0.8182	0.8187	0.8187	0.8187
0.25	0.7778	0.7788	0.7788	0.7788
0.30	0.7391	0.7407	0.7408	0.7408
0.35	0.7021	0.7046	0.7047	0.7047
0.40	0.6667	0.6701	0.6703	0.6703
0.45	0.6327	0.6373	0.6376	0.6376
0.50	0.6000	0.6061	0.6065	0.6065
0.55	0.5686	0.5763	0.5769	0.5769
0.60	0.5385	0.5479	0.5487	0.5488
0.65	0.5094	0.5209	0.5219	0.5220
0.70	0.4815	0.4952	0.4965	0.4966
0.75	0.4545	0.4706	0.4722	0.4724
0.80	0.4286	0.4472	0.4491	0.4493
0.85	0.4035	0.4248	0.4271	0.4274
0.90	0.3793	0.4035	0.4062	0.4065
0.95	0.3559	0.3831	0.3863	0.3867
1.00	0.3333	0.3636	0.3673	0.3678
1.20	0.2500	0.2941	0.3002	0.3011
1.40	0.1765	0.2360	0.2449	0.2463
1.60	0.1111	0.1872	0.1993	0.2014
1.80	0.0526	0.1460	0.1616	0.1646
2.00		0.1111	0.1304	0.1343
2.20		0.0815	0.1046	0.1094
2.40		0.0562	0.0831	0.0889
2.60		0.0345	0.0651	0.0721
2.80		0.0160	0.0521	0.0581
3.00			0.0377	0.0466
3.20			0.0273	0.0372
3.40			0.0186	0.0293
3.60			0.0113	0.0228
3.80			0.0051	0.0174
4.00				0.0130
4.20				0.0093
4.40				0.0063
4.60				0.0038
4.80				0.0017

Operating Characteristics for the Burger Dome Problem

To illustrate the multiple-server waiting line model, we return to the Burger Dome fast-food restaurant waiting line problem. Suppose that management wants to evaluate the desirability of opening a second order-processing station so that two customers can be served simultaneously. Assume a single waiting line with the first customer in line moving to the first available server. Let us evaluate the operating characteristics for this two-server system.

We use equations (15.11) through (15.18) for the $k = 2$-server system. For an arrival rate of $\lambda = 0.75$ customers per minute and a service rate of $\mu = 1$ customer per minute for each server, we obtain the operating characteristics:

WEB file

Multiple

$$P_0 = 0.4545 \quad \text{(from Table 15.4 with } \lambda/\mu = 0.75)$$

$$L_q = \frac{(0.75/1)^2(0.75)(1)}{(2-1)![2(1)-0.75]^2}(0.4545) = 0.1227 \text{ customer}$$

$$L = L_q + \frac{\lambda}{\mu} = 0.1227 + \frac{0.75}{1} = 0.8727 \text{ customer}$$

$$W_q = \frac{L_q}{\lambda} = \frac{0.1227}{0.75} = 0.1636 \text{ minute}$$

$$W = W_q + \frac{1}{\mu} = 0.1636 + \frac{1}{1} = 1.1636 \text{ minutes}$$

$$P_w = \frac{1}{2!}\left(\frac{0.75}{1}\right)^2\left[\frac{2(1)}{2(1)-0.75}\right](0.4545) = 0.2045$$

Try Problem 18 for practice in determining the operating characteristics for a two-server waiting line.

Using equations (15.17) and (15.18), we can compute the probabilities of n customers in the system. The results from these computations are summarized in Table 15.5.

We can now compare the steady-state operating characteristics of the two-server system to the operating characteristics of the original single-server system discussed in Section 15.2.

1. The average time a customer spends in the system (waiting time plus service time) is reduced from $W = 4$ minutes to $W = 1.1636$ minutes.
2. The average number of customers in the waiting line is reduced from $L_q = 2.25$ customers to $L_q = 0.1227$ customers.
3. The average time a customer spends in the waiting line is reduced from $W_q = 3$ minutes to $W_q = 0.1636$ minutes.
4. The probability that a customer has to wait for service is reduced from $P_w = 0.75$ to $P_w = 0.2045$.

Clearly the two-server system will substantially improve the operating characteristics of the waiting line. The waiting line study provides the operating characteristics that can be anticipated under three configurations: the original single-server system, a single-server system with the design change involving direct submission of paper order form to kitchen, and a two-server system composed of two order-filling employees. After considering these results, what action would you recommend? In this case, Burger Dome adopted the

TABLE 15.5 THE PROBABILITY OF n CUSTOMERS IN THE SYSTEM FOR THE BURGER DOME TWO-SERVER WAITING LINE

Number of Customers	Probability
0	0.4545
1	0.3409
2	0.1278
3	0.0479
4	0.0180
5 or more	0.0109

following policy statement: For periods when customer arrivals are expected to average 45 customers per hour, Burger Dome will open two order-processing servers with one employee assigned to each.

By changing the arrival rate λ to reflect arrival rates at different times of the day and then computing the operating characteristics, Burger Dome's management can establish guidelines and policies that tell the store managers when to schedule service operations with a single server, two servers, or perhaps even three or more servers.

NOTES AND COMMENTS

1. The multiple-server waiting line model is based on a single waiting line. You may have also encountered situations where each of the k servers has its own waiting line. Analysts have shown that the operating characteristics of multiple-server systems are better if a single waiting line is used. Also, people tend to like them better; no one who comes in after you can be served ahead of you, and so they appeal to one's sense of fairness. Thus, when possible, banks, airline reservation counters, airport security systems, food-service establishments, and other businesses frequently use a single waiting line for a multiple-server system.

15.4 Some General Relationships for Waiting Line Models

In Sections 15.2 and 15.3 we presented formulas for computing the operating characteristics for single-server and multiple-server waiting lines with Poisson arrivals and exponential service times. The operating characteristics of interest included

L_q = the average number of units in the waiting line
L = the average number of units in the system
W_q = the average time a unit spends in the waiting line
W = the average time a unit spends in the system

John D. C. Little showed that several relationships exist among these four characteristics and that these relationships apply to a variety of different waiting line systems. Two of the relationships, referred to as *Little's flow equations*, are

$$L = \lambda W \quad (15.19)$$

$$L_q = \lambda W_q \quad (15.20)$$

Equation (15.19) shows that the average number of units in the system, L, can be found by multiplying the arrival rate, λ, by the average time a unit spends in the system, W. Equation (15.20) shows that the same relationship holds between the average number of units in the waiting line, L_q, and the average time a unit spends in the waiting line, W_q.

Using equation (15.20) and solving for W_q, we obtain

$$W_q = \frac{L_q}{\lambda} \quad (15.21)$$

Equation (15.21) follows directly from Little's second flow equation. We used it for the single-server waiting line model in Section 15.2 and the multiple-server waiting line model in Section 15.3 [see equations (15.7) and (15.14)]. Once L_q is computed for either of these models, equation (15.21) can then be used to compute W_q.

Another general expression that applies to waiting line models is that the average time in the system, W, is equal to the average time in the waiting line, W_q, plus the average service time. For a system with a service rate μ, the mean service time is $1/\mu$. Thus, we have the general relationship

$$W = W_q + \frac{1}{\mu} \qquad \textbf{(15.22)}$$

Recall that we used equation (15.22) to provide the average time in the system for both the single- and multiple-server waiting line models [see equations (15.8) and (15.15)].

The importance of Little's flow equations is that they apply to *any waiting line model* regardless of whether arrivals follow the Poisson probability distribution and regardless of whether service times follow the exponential probability distribution. For example, in a study of the grocery checkout counters at Murphy's Foodliner, an analyst concluded that arrivals follow the Poisson probability distribution with an arrival rate of 24 customers per hour, or $\lambda = 24/60 = 0.40$ customers per minute. However, the analyst found that service times follow a normal probability distribution rather than an exponential probability distribution. The service rate was found to be 30 customers per hour, or $\mu = 30/60 = 0.50$ customers per minute. A time study of actual customer waiting times showed that, on average, a customer spends 4.5 minutes in the system (waiting time plus checkout time); that is, $W = 4.5$. Using the waiting line relationships discussed in this section, we can now compute other operating characteristics for this waiting line.

The advantage of Little's flow equations is that they show how operating characteristics L, L_q, W, and W_q are related in any waiting line system. Arrivals and service times do not have to follow specific probability distributions for the flow equations to be applicable.

First, using equation (15.22) and solving for W_q, we have

$$W_q = W - \frac{1}{\mu} = 4.5 - \frac{1}{0.50} = 2.5 \text{ minutes}$$

With both W and W_q known, we can use Little's flow equations, (15.19) and (15.20), to compute

$$L = \lambda W = 0.40(4.5) = 1.8 \text{ customers}$$
$$L_q = \lambda W_q = 0.40(2.5) = 1 \text{ customer}$$

The application of Little's flow equations is demonstrated in Problem 24.

The manager of Murphy's Foodliner can now review these operating characteristics to see whether action should be taken to improve the service and to reduce the waiting time and the length of the waiting line.

NOTES AND COMMENTS

1. In waiting line systems where the length of the waiting line is limited (e.g., a small waiting area), some arriving units will be blocked from joining the waiting line and will be lost. In this case, the blocked or lost arrivals will make the mean number of units entering the system something less than the arrival rate. In other instances, arrivals will decide the line is too long and will leave. By defining λ as the mean number of units *joining the system,* rather than the arrival rate, the relationships discussed in this section can be used to determine W, L, W_q, and L_q.

15.5 Economic Analysis of Waiting Lines

Frequently, decisions involving the design of waiting lines will be based on a subjective evaluation of the operating characteristics of the waiting line. For example, a manager may decide that an average waiting time of one minute or less and an average of two customers or fewer in the system are reasonable goals. The waiting line models presented in the preceding sections can be used to determine the number of servers that will meet the manager's waiting line performance goals.

On the other hand, a manager may want to identify the cost of operating the waiting line system and then base the decision regarding system design on a minimum hourly or daily operating cost. Before an economic analysis of a waiting line can be conducted, a total cost model, which includes the cost of waiting and the cost of service, must be developed.

Waiting cost is based on average number of units in the system. It includes the time spent waiting in line plus the time spent being served.

To develop a total cost model for a waiting line, we begin by defining the notation to be used:

$$
\begin{aligned}
c_w &= \text{the waiting cost per time period for each unit} \\
L &= \text{the average number of units in the system} \\
c_s &= \text{the service cost per time period for each server} \\
k &= \text{the number of servers} \\
TC &= \text{the total cost per time period}
\end{aligned}
$$

The total cost is the sum of the waiting cost and the service cost; that is,

$$TC = c_w L + c_s k \tag{15.23}$$

Adding more servers always improves the operating characteristics of the waiting line and reduces the waiting cost. However, additional servers increase the service cost. An economic analysis of waiting lines attempts to find the number of servers that will minimize total cost by balancing the waiting cost and the service cost.

To conduct an economic analysis of a waiting line, we must obtain reasonable estimates of the waiting cost and the service cost. Of these two costs, the waiting cost is usually the more difficult to evaluate. In the Burger Dome restaurant problem, the waiting cost would be the cost per minute for a customer waiting for service. This cost is not a direct cost to Burger Dome. However, if Burger Dome ignores this cost and allows long waiting lines, customers ultimately will take their business elsewhere. Thus, Burger Dome will experience lost sales and, in effect, incur a cost.

The service cost is generally easier to determine as it relates to any cost associated with establishing each server operation. In the Burger Dome problem, this cost would include the server's wages, benefits, and any other direct costs associated with establishing a server. At Burger Dome, this cost is estimated to be \$10 per hour.

To demonstrate the use of equation (15.23), we assume that Burger Dome is willing to assign a cost of \$15 per hour for customer waiting time. We use the average number of units in the system, L, as computed in Sections 15.2 and 15.3 to obtain the total hourly cost for the single-server and two-server systems:

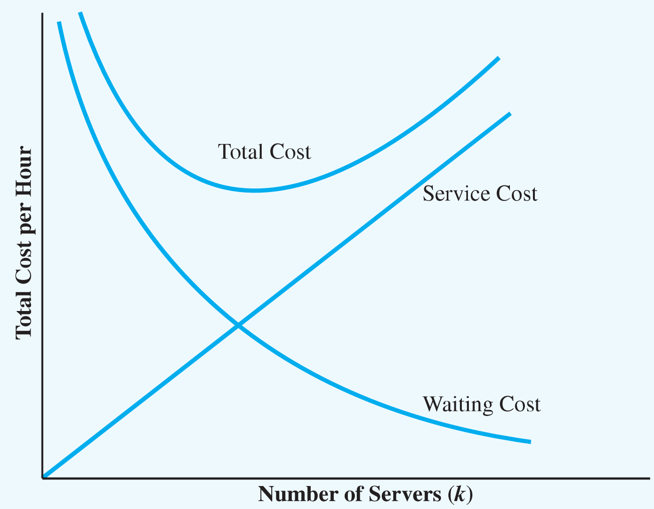
Single-server system ($L = 3$ customers):

$$
\begin{aligned}
TC &= c_w L + c_s k \\
&= 15(3) + 10(1) = \$55.00 \text{ per hour}
\end{aligned}
$$

Two-server system ($L = 0.8727$ customer):

$$
\begin{aligned}
TC &= c_w L + c_s k \\
&= 15(0.8727) + 10(2) = \$33.09 \text{ per hour}
\end{aligned}
$$

FIGURE 15.4 THE GENERAL SHAPE OF WAITING COST, SERVICE COST, AND TOTAL COST CURVES IN WAITING LINE MODELS



Thus, based on the cost data provided by Burger Dome, the two-server system provides the more economical operation. Note that when the cost of serving a customer c_s exceeds the cost of customer waiting time c_w by a sufficient amount, the single-server system will be more economical for Burger Dome.

Problem 21 tests your ability to conduct an economic analysis of proposed single-server and two-server waiting line systems.

Figure 15.4 shows the general shape of the cost curves in the economic analysis of waiting lines. The service cost increases as the number of servers is increased. However, with more servers, the service is better. As a result, waiting time and cost decrease as the number of servers is increased. The number of servers that will provide a good approximation of the minimum total cost design can be found by evaluating the total cost for several design alternatives.

NOTES AND COMMENTS

1. In dealing with government agencies and utility companies, customers may not be able to take their business elsewhere. In these situations, no lost business occurs when long waiting times are encountered. This condition is one reason that service in such organizations may be poor and that customers in such situations may experience long waiting times.
2. In some instances, the organization providing the service also employs the units waiting for the service. For example, consider the case of a company that owns and operates the trucks used to deliver goods to and from its manufacturing plant. In addition to the costs associated with the trucks waiting to be loaded or unloaded, the firm also pays the wages of the truck loaders and unloaders who are effectively operating as servers. In this case, the cost of having the trucks wait and the cost of operating the servers are direct expenses to the firm. An economic analysis of the waiting line system is highly recommended for these types of situations.

15.6 Other Waiting Line Models

D. G. Kendall suggested a notation that is helpful in classifying the wide variety of different waiting line models that have been developed. The three-symbol Kendall notation is as follows:

$$A/B/k$$

where

A denotes the probability distribution for the arrivals

B denotes the probability distribution for the service time

k denotes the number of servers

Depending on the letter appearing in the *A* or *B* position, a variety of waiting line systems can be described. The letters that are commonly used are as follows:

M designates a Poisson probability distribution for the arrivals or an exponential probability distribution for service time

D designates that the arrivals or the service times are deterministic or constant

G designates that the arrivals or the service times have a general probability distribution with a known mean and variance

Using the Kendall notation, the single-server waiting line model with Poisson arrivals and exponential service times is classified as an $M/M/1$ model. The two-server waiting line model with Poisson arrivals and exponential service times presented in Section 15.3 would be classified as an $M/M/2$ model.

NOTES AND COMMENTS

1. In some cases, the Kendall notation is extended to five symbols. The fourth symbol indicates the largest number of units that can be in the system, and the fifth symbol indicates the size of the population. The fourth symbol is used in situations where the waiting line can hold a finite or maximum number of units, and the fifth symbol is necessary when the population of arriving units or customers is finite. When the fourth and fifth symbols of the Kendall notation are omitted, the waiting line system is assumed to have infinite capacity, and the population is assumed to be infinite.

15.7 Single-Server Waiting Line Model with Poisson Arrivals and Arbitrary Service Times

Let us return to the single-server waiting line model where arrivals are described by a Poisson probability distribution. However, we now assume that the probability distribution for the service times is not an exponential probability distribution. Thus, using the Kendall notation, the waiting line model that is appropriate is an $M/G/1$ model, where *G* denotes a general or unspecified probability distribution.

When providing input to the M/G/1 model, be consistent in terms of the time period. For example, if λ and μ are expressed in terms of the number of units per hour, the standard deviation of the service time should be expressed in hours. The example that follows uses minutes as the time period for the arrival and service data.

Operating Characteristics for the *M/G/*1 Model

The notation used to describe the operating characteristics for the *M/G/*1 model is

λ = the arrival rate
μ = the service rate
σ = the standard deviation of the service time

Some of the steady-state operating characteristics of the *M/G/*1 waiting line model are as follows:

1. The probability that no units are in the system:

$$P_0 = 1 - \frac{\lambda}{\mu} \qquad \textbf{(15.24)}$$

2. The average number of units in the waiting line:

$$L_q = \frac{\lambda^2\sigma^2 + (\lambda/\mu)^2}{2(1 - \lambda/\mu)} \qquad \textbf{(15.25)}$$

3. The average number of units in the system:

$$L = L_q + \frac{\lambda}{\mu} \qquad \textbf{(15.26)}$$

4. The average time a unit spends in the waiting line:

$$W_q = \frac{L_q}{\lambda} \qquad \textbf{(15.27)}$$

5. The average time a unit spends in the system:

$$W = W_q + \frac{1}{\mu} \qquad \textbf{(15.28)}$$

6. The probability that an arriving unit has to wait for service:

$$P_w = \frac{\lambda}{\mu} \qquad \textbf{(15.29)}$$

Note that the relationships for L, W_q, and W are the same as the relationships used for the waiting line models in Sections 15.2 and 15.3. They are given by Little's flow equations.

Problem 27 provides another application of a single-server waiting line with Poisson arrivals and arbitrary service times.

An Example Retail sales at Hartlage's Seafood Supply are handled by one clerk. Customer arrivals follow a Poisson distribution, and the arrival rate is 21 customers per hour or $\lambda = 21/60 = 0.35$ customers per minute. A study of the service process shows that the service time is a random variable with an average of 2 minutes per customer and a standard deviation of $\sigma = 1.2$ minutes. The mean time of 2 minutes per customer shows that the clerk has a service rate of $\mu = 1/2 = 0.50$ customers per minute. The operating characteristics of this *M/G/*1 waiting line system are as follows:

WEB file

Single-Arbitrary

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{0.35}{0.50} = 0.30$$

$$L_q = \frac{(0.35)^2(1.2)^2 + (0.35/0.50)^2}{2(1 - 0.35/0.50)} = 1.1107 \text{ customers}$$

$$L = L_q + \frac{\lambda}{\mu} = 1.1107 + \frac{0.35}{0.50} = 1.8107 \text{ customers}$$

$$W_q = \frac{L_q}{\lambda} = \frac{1.1107}{0.35} = 3.1733 \text{ minutes}$$

$$W = W_q + \frac{1}{\mu} = 3.1733 + \frac{1}{0.50} = 5.1733 \text{ minutes}$$

$$P_w = \frac{\lambda}{\mu} = \frac{0.35}{0.50} = 0.70$$

Hartlage's manager can review these operating characteristics to determine whether scheduling a second clerk appears to be worthwhile.

Constant Service Times

We want to comment briefly on the single-server waiting line model that assumes random arrivals but constant service times. Such a waiting line can occur in production and manufacturing environments where machine-controlled service times are generally constant. This waiting line is described by the *M/D/*1 model, with the *D* referring to the deterministic service times. With the *M/D/*1 model, the average number of units in the waiting line, L_q, can be found by using equation (15.25) with the condition that the standard deviation of the constant service time is $\sigma = 0$. Thus, the expression for the average number of units in the waiting line for the *M/D/*1 waiting line becomes

$$L_q = \frac{(\lambda/\mu)^2}{2(1 - \lambda/\mu)} \qquad \textbf{(15.30)}$$

The other expressions presented earlier in this section can be used to determine additional operating characteristics of the *M/D/*1 system.

NOTES AND COMMENTS

1. Whenever the operating characteristics of a waiting line are unacceptable, managers often try to improve service by increasing the service rate μ. This approach is good, but equation (15.25) shows that the variation in the service times also affects the operating characteristics of the waiting line. Because the standard deviation of service times, σ, appears in the numerator of equation (15.25), a larger variation in service times results in a larger average number of units in the waiting line. Hence, another alternative for improving the service capabilities of a waiting line is to reduce the variation in the service times. Thus, even when the service rate of the service facility cannot be increased, a reduction in σ will reduce the average number of units in the waiting line and improve the operating characteristics of the system.

15.8 Multiple-Server Model with Poisson Arrivals, Arbitrary Service Times, and No Waiting Line

An interesting variation of the waiting line models discussed so far involves a system in which no waiting is allowed. Arriving units or customers seek service from one of several servers. If all servers are busy, arriving units are denied access to the system. In waiting line terminology, arrivals occurring when the system is full are **blocked** and are cleared from the system. Such customers may be lost or may attempt a return to the system later.

The specific model considered in this section is based on the following assumptions:

1. The system has k servers.
2. The arrivals follow a Poisson probability distribution, with arrival rate λ.
3. The service times for each server may have any probability distribution.
4. The service rate μ is the same for each server.
5. An arrival enters the system only if at least one server is available. An arrival occurring when all servers are busy is blocked—that is, denied service and not allowed to enter the system.

With G denoting a general or unspecified probability distribution for service times, the appropriate model for this situation is referred to as an $M/G/k$ model with "blocked customers cleared." The question addressed in this type of situation is, How many servers should be used?

A primary application of this model involves the design of telephone and other communication systems where the arrivals are the calls and the servers are the number of telephone or communication lines available. In such a system, the calls are made to one telephone number, with each call automatically switched to an open server if possible. When all servers are busy, additional calls receive a busy signal and are denied access to the system.

Operating Characteristics for the $M/G/k$ Model with Blocked Customers Cleared

With no waiting allowed, operating characteristics L_q and W_q considered in previous waiting line models are automatically zero regardless of the number of servers. In this situation, the more important design consideration involves determining how the percentage of blocked customers is affected by the number of servers.

We approach the problem of selecting the best number of servers by computing the steady-state probabilities that j of the k servers will be busy. These probabilities are

$$P_j = \frac{(\lambda/\mu)^j/j!}{\sum_{i=0}^{k} (\lambda/\mu)^i/i!} \qquad \textbf{(15.31)}$$

where

λ = the arrival rate
μ = the service rate for each server
k = the number of servers
P_j = the probability that j of the k servers are busy for $j = 0, 1, 2, \ldots, k$

The most important probability value is P_k, which is the probability that all k servers are busy. Thus, P_k also indicates the percentage of arrivals that are blocked and denied access to the system.

Another operating characteristic of interest is the average number of units in the system; note that this number is equivalent to the average number of servers in use. Letting L denote the average number of units in the system, we have

$$L = \frac{\lambda}{\mu}(1 - P_k) \tag{15.32}$$

where L will certainly be less than k.

An Example Microdata Software, Inc., uses a telephone ordering system for its computer software products. Callers place orders with Microdata by using the company's 800 telephone number. Assume that calls to this telephone number arrive at a rate of $\lambda = 12$ calls per hour. The time required to process a telephone order varies considerably from order to order. However, each Microdata sales representative can be expected to handle $\mu = 6$ calls per hour. Currently, the Microdata 800 telephone number has three internal lines or servers, each operated by a separate sales representative. Calls received on the 800 number are automatically transferred to an open line or server if available.

Whenever all three lines are busy, callers receive a busy signal. In the past, Microdata's management assumed that callers receiving a busy signal would call back later. However, recent research on telephone ordering showed that a substantial number of callers who are denied access do not call back later. These lost calls represent lost revenues for the firm, so Microdata's management requested an analysis of the telephone ordering system. Specifically, management wanted to know the percentage of callers who get busy signals and are blocked from the system. If management's goal is to provide sufficient capacity to handle 90% of the callers, how many telephone lines and sales representatives should Microdata use?

We can demonstrate the use of equation (15.31) by computing P_3, the probability that all three of the currently available telephone lines will be in use and additional callers will be blocked:

WEB file

No Waiting

$$P_3 = \frac{(^{12}/_6)^3/3!}{(^{12}/_6)^0/0! + (^{12}/_6)^1/1! + (^{12}/_6)^2/2! + (^{12}/_6)^3/3!} = \frac{1.3333}{6.3333} = 0.2105$$

With $P_3 = 0.2105$, approximately 21% of the calls, or slightly more than one in five calls, are being blocked. Only 79% of the calls are being handled immediately by the three-line system.

Let us assume that Microdata expands to a four-line system. Then, the probability that all four servers will be in use and that callers will be blocked is

$$P_4 = \frac{(^{12}/_6)^4/4!}{(^{12}/_6)^0/0! + (^{12}/_6)^1/1! + (^{12}/_6)^2/2! + (^{12}/_6)^3/3! + (^{12}/_6)^4/4!} = \frac{0.667}{7} = 0.0952$$

Problem 30 provides practice in calculating probabilities for multiple-server systems with no waiting line.

With only 9.52% of the callers blocked, 90.48% of the callers will reach the Microdata sales representatives. Thus, Microdata should expand its order-processing operation to four lines to meet management's goal of providing sufficient capacity to handle at least 90% of the callers. The average number of calls in the four-line system and thus the average number of lines and sales representatives that will be busy is

$$L = \frac{\lambda}{\mu}(1 - P_4) = \frac{12}{6}(1 - 0.0952) = 1.8095$$

TABLE 15.6 PROBABILITIES OF BUSY LINES FOR THE MICRODATA FOUR-LINE SYSTEM

Number of Busy Lines	Probability
0	0.1429
1	0.2857
2	0.2857
3	0.1905
4	0.0952

Although an average of fewer than two lines will be busy, the four-line system is necessary to provide the capacity to handle at least 90% of the callers. We used equation (15.31) to calculate the probability that 0, 1, 2, 3, or 4 lines will be busy. These probabilities are summarized in Table 15.6.

As discussed in Section 15.5, an economic analysis of waiting lines can be used to guide system design decisions. In the Microdata system, the cost of the additional line and additional sales representative should be relatively easy to establish. This cost can be balanced against the cost of the blocked calls. With 9.52% of the calls blocked and $\lambda = 12$ calls per hour, an eight-hour day will have an average of $8(12)(0.0952) = 9.1$ blocked calls. If Microdata can estimate the cost of possible lost sales, the cost of these blocked calls can be established. The economic analysis based on the service cost and the blocked-call cost can assist in determining the optimal number of lines for the system.

15.9 Waiting Line Models with Finite Calling Populations

In previous waiting line models, the arrival rate was constant and independent of the number of units in the system. With a finite calling population, the arrival rate decreases as the number of units in the system increases because, with more units in the system, fewer units are available for arrivals.

For the waiting line models introduced so far, the population of units or customers arriving for service has been considered to be unlimited. In technical terms, when no limit is placed on how many units may seek service, the model is said to have an **infinite calling population**. Under this assumption, the arrival rate λ remains constant regardless of how many units are in the waiting line system. This assumption of an infinite calling population is made in most waiting line models.

In other cases, the maximum number of units or customers that may seek service is assumed to be finite. In this situation the arrival rate for the system changes, depending on the number of units in the waiting line, and the waiting line model is said to have a **finite calling population**. The formulas for the operating characteristics of the previous waiting line models must be modified to account for the effect of the finite calling population.

The finite calling population model discussed in this section is based on the following assumptions:

1. The arrivals for *each unit* follow a Poisson probability distribution, with arrival rate λ.
2. The service times follow an exponential probability distribution, with service rate μ.
3. The population of units that may seek service is finite.

With a single server, the waiting line model is referred to as an *M*/*M*/1 model with a finite calling population.

The arrival rate λ is defined differently for the finite calling population model. Specifically, λ is defined in terms of the arrival rate for each unit.

The arrival rate for the $M/M/1$ model with a finite calling population is defined in terms of how often each unit arrives or seeks service. This situation differs from that for previous waiting line models, in which λ denoted the arrival rate for the system. With a finite calling population, the arrival rate for the system varies depending on the number of units in the system. Instead of adjusting for the changing system arrival rate, in the finite calling population model λ indicates the arrival rate for each unit.

Operating Characteristics for the *M/M/*1 Model with a Finite Calling Population

The following formulas are used to determine the steady-state operating characteristics for an $M/M/1$ model with a finite calling population, where

$$\lambda = \text{the arrival rate for each unit}$$
$$\mu = \text{the service rate}$$
$$N = \text{the size of the population}$$

1. The probability that no units are in the system:

$$P_0 = \frac{1}{\displaystyle\sum_{n=0}^{N} \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n} \qquad \textbf{(15.33)}$$

2. The average number of units in the waiting line:

$$L_q = N - \frac{\lambda + \mu}{\lambda}(1 - P_0) \qquad \textbf{(15.34)}$$

3. The average number of units in the system:

$$L = L_q + (1 - P_0) \qquad \textbf{(15.35)}$$

4. The average time a unit spends in the waiting line:

$$W_q = \frac{L_q}{(N - L)\lambda} \qquad \textbf{(15.36)}$$

5. The average time a unit spends in the system:

$$W = W_q + \frac{1}{\mu} \qquad \textbf{(15.37)}$$

6. The probability an arriving unit has to wait for service:

$$P_w = 1 - P_0 \quad \textbf{(15.38)}$$

7. The probability of n units in the system:

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_0 \quad \text{for } n = 0, 1, \ldots, N \quad \textbf{(15.39)}$$

One of the primary applications of the $M/M/1$ model with a finite calling population is referred to as the *machine repair problem.* In this problem, a group of machines is considered to be the finite population of "customers" that may request repair service. Whenever a machine breaks down, an arrival occurs in the sense that a new repair request is initiated. If another machine breaks down before the repair work has been completed on the first machine, the second machine begins to form a "waiting line" for repair service. Additional breakdowns by other machines will add to the length of the waiting line. The assumption of first-come, first-served indicates that machines are repaired in the order they break down. The $M/M/1$ model shows that one person or one server is available to perform the repair service. To return the machine to operation, each machine with a breakdown must be repaired by the single-server operation. This model is often applied by computer maintenance departments of various organizations.

An Example The Kolkmeyer Manufacturing Company uses a group of six identical machines, each of which operates an average of 20 hours between breakdowns. Thus, the arrival rate or request for repair service for each machine is $\lambda = 1/20 = 0.05$ per hour. With randomly occurring breakdowns, the Poisson probability distribution is used to describe the machine breakdown arrival process. One person from the maintenance department provides the single-server repair service for the six machines. The exponentially distributed service times have a mean of two hours per machine, or a service rate of $\mu = 1/2 = 0.50$ machines per hour.

With $\lambda = 0.05$ and $\mu = 0.50$, we use equations (15.33) through (15.38) to compute the operating characteristics for this system. Note that the use of equation (15.33) makes the computations involved somewhat cumbersome. Confirm for yourself that equation (15.33) provides the value of $P_0 = 0.4845$. The computations for the other operating characteristics are

$$L_q = 6 - \left(\frac{0.05 + 0.50}{0.05}\right)(1 - 0.4845) = 0.3297 \text{ machines}$$

$$L = 0.3295 + (1 - 0.4845) = 0.8451 \text{ machines}$$

$$W_q = \frac{0.3295}{(6 - 0.845)0.50} = 1.279 \text{ hours}$$

$$W = 1.279 + \frac{1}{0.50} = 3.279 \text{ hours}$$

$$P_w = 1 - P_0 = 1 - 0.4845 = 0.5155$$

FIGURE 15.5 WORKSHEET FOR THE KOLKMEYER TWO-SERVER MACHINE REPAIR PROBLEM

WEB file

Finite

	A	B	C	D
1	**Waiting Line Model with a Finite Calling Population**			
2				
3	**Assumptions**			
4	**Poisson Arrivals**			
5	**Exponential Service Times**			
6	**Finite Calling Population**			
7				
8	Number of Servers	2		
9	Arrival Rate For Each Unit	0.05		
10	Service Rate For Each Server	0.5		
11	Population Size	6		
12				
13				
14	**Operating Characteristics**			
15				
16	Probability that no customers are in the system, P_0		0.5602	
17	Average number of customers in the waiting line, L_q		0.0227	
18	Average number of customers in the system, L		0.5661	
19	Average time a customer spends in the waiting line, W_q		0.0834	
20	Average time a customer spends in the system, W		2.0834	
21	Probability an arriving customer has to wait, P_w		0.1036	
22				

Operating characteristics of an M/M/1 waiting line with a finite calling population are considered in Problem 34.

Finally, equation (15.39) can be used to compute the probabilities of any number of machines being in the repair system.

As with other waiting line models, the operating characteristics provide the manager with useful information about the operation of the waiting line. In this case, the fact that a machine breakdown waits an average of $W_q = 1.279$ hours before maintenance begins and the fact that more than 50% of the machine breakdowns must wait for service, $P_w = 0.5155$, indicates that a two-server system may be needed to improve the machine repair service.

An Excel worksheet template at the Web files link on the website that accompanies this text may be used to analyze the multiple-server finite calling population model.

Computations of the operating characteristics of a multiple-server finite calling population waiting line are more complex than those for the single-server model. A computer solution is virtually mandatory in this case. The Excel worksheet for the Kolkmeyer two-server machine repair system is shown in Figure 15.5. With two repair personnel, the average machine breakdown waiting time is reduced to $W_q = 0.0834$ hours, or 5 minutes, and only 10%, $P_w = 0.1036$, of the machine breakdowns wait for service. Thus, the two-server system significantly improves the machine repair service operation. Ultimately, by considering the cost of machine downtime and the cost of the repair personnel, management can determine whether the improved service of the two-server system is cost effective.

Summary

In this chapter we presented a variety of waiting line models that have been developed to help managers make better decisions concerning the operation of waiting lines. For each model we presented formulas that could be used to develop operating characteristics or performance measures for the system being studied. The operating characteristics presented include the following:

1. Probability that no units are in the system
2. Average number of units in the waiting line
3. Average number of units in the system
4. Average time a unit spends in the waiting line
5. Average time a unit spends in the system
6. Probability that arriving units will have to wait for service

We also showed how an economic analysis of the waiting line could be conducted by developing a total cost model that includes the cost associated with units waiting for service and the cost required to operate the service facility.

As many of the examples in this chapter show, the most obvious applications of waiting line models are situations in which customers arrive for service such as at a grocery checkout counter, bank, or restaurant. However, with a little creativity, waiting line models can be applied to many different situations, such as telephone calls waiting for connections, mail orders waiting for processing, machines waiting for repairs, manufacturing jobs waiting to be processed, and money waiting to be spent or invested. The Q.M. in Action, Allocating Voting Machines to Polling Locations, describes an application in which a waiting line model helped decrease the waiting times voters experience on Election Day.

The complexity and diversity of waiting line systems found in practice often prevent an analyst from finding an existing waiting line model that fits the specific application being studied. Simulation, the topic discussed in Chapter 16, provides an approach to determining the operating characteristics of such waiting line systems.

Q.M. *in* ACTION

*ALLOCATING VOTING MACHINES TO POLLING LOCATIONS**

In the 2004 U.S. presidential election, many voters waited more than ten hours to cast their ballots. Similar problems with long lines at the voting booth were reported in the United States in the 2006 and 2008 elections as well as in the 2010 elections in the United Kingdom. Long lines at a voting booth can result in a voter leaving without casting her/his ballot. Queueing models can be used to diagnose why these long lines are occurring and offer improvements.

Many of the problems occurring in the U.S. elections can be linked to the implementation of direct-recording electronic (DRE) voting machines (better known as touch-screen systems). Because these systems are quite expensive, many election boards had only a relatively few DRE voting machines to allocate to polling locations. Voters' unfamiliarity with the systems also resulted in increased voting times when using DRE machines. Most election boards initially allocated voting machines to polling locations without considering queueing effects.

Starting in 2008, the Board of Elections in Franklin County, Ohio (the location of the state capital, Columbus) has used queueing models to help determine the optimal allocation of voting machines to polling locations. Voting machines can be considered as servers in this context and the voters can be thought of as customers. Queueing models were used to predict voter waiting times based on expected voter turnout, number of registered voters, and ballot lengths. The use of queueing models was credited with greatly reducing the waiting times for Franklin County voters in the 2010 presidential election, even though voter turnout was at a record high.

*Based on work done by Ted Allen (The Ohio State University), Mike Fry and David Kelton (University of Cincinnati), and Muer Yang (University of St. Thomas).

Glossary

Queue A waiting line.
Queueing theory The body of knowledge dealing with waiting lines.
Operating characteristics The performance measures for a waiting line, including the probability that no units are in the system, the average number of units in the waiting line, the average waiting time, and so on.
Single-server waiting line A waiting line with only one service facility.
Poisson probability distribution A probability distribution used to describe the arrival pattern for some waiting line models.
Arrival rate The mean number of customers or units arriving in a given period of time.
Exponential probability distribution A probability distribution used to describe the service time for some waiting line models.
Service rate The mean number of customers or units that can be served by one service facility in a given period of time.
First-come, first-served (FCFS) The queue discipline that serves waiting units on a first-come, first-served basis.
Transient period The startup period for a waiting line, occurring before the waiting line reaches a normal or steady-state operation.
Steady-state operation The normal operation of the waiting line after it has gone through a startup or transient period. The operating characteristics of waiting lines are computed for steady-state conditions.
Multiple-server waiting line A waiting line with two or more parallel service facilities.
Blocked When arriving units cannot enter the waiting line because the system is full. Blocked units can occur when waiting lines are not allowed or when waiting lines have a finite capacity.
Infinite calling population The population of customers or units that may seek service has no specified upper limit.
Finite calling population The population of customers or units that may seek service has a fixed and finite value.

Problems

1. Willow Brook National Bank operates a drive-up teller window that allows customers to complete bank transactions without getting out of their cars. On weekday mornings, arrivals to the drive-up teller window occur at random, with an arrival rate of 24 customers per hour or 0.4 customers per minute.
   a. What is the mean or expected number of customers that will arrive in a five-minute period?
   b. Assume that the Poisson probability distribution can be used to describe the arrival process. Use the arrival rate in part (a) and compute the probabilities that exactly 0, 1, 2, and 3 customers will arrive during a five-minute period.
   c. Delays are expected if more than three customers arrive during any five-minute period. What is the probability that delays will occur?
2. In the Willow Brook National Bank waiting line system (see Problem 1), assume that the service times for the drive-up teller follow an exponential probability distribution with a

service rate of 36 customers per hour, or 0.6 customers per minute. Use the exponential probability distribution to answer the following questions:

a. What is the probability that the service time is one minute or less?
b. What is the probability that the service time is two minutes or less?
c. What is the probability that the service time is more than two minutes?

3. Use the single-server drive-up bank teller operation referred to in Problems 1 and 2 to determine the following operating characteristics for the system:
   a. The probability that no customers are in the system
   b. The average number of customers waiting
   c. The average number of customers in the system
   d. The average time a customer spends waiting
   e. The average time a customer spends in the system
   f. The probability that arriving customers will have to wait for service

4. Use the single-server drive-up bank teller operation referred to in Problems 1–3 to determine the probabilities of 0, 1, 2, and 3 customers in the system. What is the probability that more than three customers will be in the drive-up teller system at the same time?

SELF test

5. The reference desk of a university library receives requests for assistance. Assume that a Poisson probability distribution with an arrival rate of 10 requests per hour can be used to describe the arrival pattern and that service times follow an exponential probability distribution with a service rate of 12 requests per hour.
   a. What is the probability that no requests for assistance are in the system?
   b. What is the average number of requests that will be waiting for service?
   c. What is the average waiting time in minutes before service begins?
   d. What is the average time at the reference desk in minutes (waiting time plus service time)?
   e. What is the probability that a new arrival has to wait for service?

6. Movies Tonight is a typical video and DVD movie rental outlet for home-viewing customers. During the weeknight evenings, customers arrive at Movies Tonight with an arrival rate of 1.25 customers per minute. The checkout clerk has a service rate of 2 customers per minute. Assume Poisson arrivals and exponential service times.
   a. What is the probability that no customers are in the system?
   b. What is the average number of customers waiting for service?
   c. What is the average time a customer waits for service to begin?
   d. What is the probability that an arriving customer will have to wait for service?
   e. Do the operating characteristics indicate that the one-clerk checkout system provides an acceptable level of service?

7. Speedy Oil provides a single-server automobile oil change and lubrication service. Customers provide an arrival rate of 2.5 cars per hour. The service rate is 5 cars per hour. Assume that arrivals follow a Poisson probability distribution and that service times follow an exponential probability distribution.
   a. What is the average number of cars in the system?
   b. What is the average time that a car waits for the oil and lubrication service to begin?
   c. What is the average time a car spends in the system?
   d. What is the probability that an arrival has to wait for service?

8. For the Burger Dome single-server waiting line in Section 15.2, assume that the arrival rate is increased to 1 customer per minute and that the service rate is increased to 1.25 customers per minute. Compute the following operating characteristics for the new system: P_0, L_q, L, W_q, W, and P_w. Does this system provide better or poorer service compared to the original system? Discuss any differences and the reason for these differences.

9. Marty's Barber Shop has one barber. Customers have an arrival rate of 2.2 customers per hour, and haircuts are given with a service rate of 5 per hour. Use the Poisson arrivals and exponential service times model to answer the following questions:
   a. What is the probability that no units are in the system?
   b. What is the probability that one customer is receiving a haircut and no one is waiting?
   c. What is the probability that one customer is receiving a haircut and one customer is waiting?
   d. What is the probability that one customer is receiving a haircut and two customers are waiting?
   e. What is the probability that more than two customers are waiting?
   f. What is the average time a customer waits for service?
10. Trosper Tire Company decided to hire a new mechanic to handle all tire changes for customers ordering a new set of tires. Two mechanics applied for the job. One mechanic has limited experience, can be hired for $14 per hour, and can service an average of three customers per hour. The other mechanic has several years of experience, can service an average of four customers per hour, but must be paid $20 per hour. Assume that customers arrive at the Trosper garage at the rate of two customers per hour.
    a. What are the waiting line operating characteristics using each mechanic, assuming Poisson arrivals and exponential service times?
    b. If the company assigns a customer waiting cost of $30 per hour, which mechanic provides the lower operating cost?

SELF test

11. Agan Interior Design provides home and office decorating assistance to its customers. In normal operation, an average of 2.5 customers arrive each hour. One design consultant is available to answer customer questions and make product recommendations. The consultant averages 10 minutes with each customer.
    a. Compute the operating characteristics of the customer waiting line, assuming Poisson arrivals and exponential service times.
    b. Service goals dictate that an arriving customer should not wait for service more than an average of 5 minutes. Is this goal being met? If not, what action do you recommend?
    c. If the consultant can reduce the average time spent per customer to 8 minutes, what is the mean service rate? Will the service goal be met?
12. Pete's Market is a small local grocery store with only one checkout counter. Assume that shoppers arrive at the checkout lane according to a Poisson probability distribution, with an arrival rate of 15 customers per hour. The checkout service times follow an exponential probability distribution, with a service rate of 20 customers per hour.
    a. Compute the operating characteristics for this waiting line.
    b. If the manager's service goal is to limit the waiting time prior to beginning the checkout process to no more than five minutes, what recommendations would you provide regarding the current checkout system?
13. After reviewing the waiting line analysis of Problem 12, the manager of Pete's Market wants to consider one of the following alternatives for improving service. What alternative would you recommend? Justify your recommendation.
    a. Hire a second person to bag the groceries while the cash register operator is entering the cost data and collecting money from the customer. With this improved single-server operation, the service rate could be increased to 30 customers per hour.
    b. Hire a second person to operate a second checkout counter. The two-server operation would have a service rate of 20 customers per hour for each server.
14. Ocala Software Systems operates a technical support center for its software customers. If customers have installation or use problems with Ocala software products, they may telephone the technical support center and obtain free consultation. Currently, Ocala operates

its support center with one consultant. If the consultant is busy when a new customer call arrives, the customer hears a recorded message stating that all consultants are currently busy with other customers. The customer is then asked to hold and is told that a consultant will provide assistance as soon as possible. The customer calls follow a Poisson probability distribution, with an arrival rate of five calls per hour. On average, it takes 7.5 minutes for a consultant to answer a customer's questions. The service time follows an exponential probability distribution.

a. What is the service rate in terms of customers per hour?
b. What is the probability that no customers are in the system and the consultant is idle?
c. What is the average number of customers waiting for a consultant?
d. What is the average time a customer waits for a consultant?
e. What is the probability that a customer will have to wait for a consultant?
f. Ocala's customer service department recently received several letters from customers complaining about the difficulty in obtaining technical support. If Ocala's customer service guidelines state that no more than 35% of all customers should have to wait for technical support and that the average waiting time should be two minutes or less, does your waiting line analysis indicate that Ocala is or is not meeting its customer service guidelines? What action, if any, would you recommend?

15. To improve customer service, Ocala Software Systems (see Problem 14) wants to investigate the effect of using a second consultant at its technical support center. What effect would the additional consultant have on customer service? Would two technical consultants enable Ocala to meet its service guidelines (no more than 35% of all customers having to wait for technical support and an average customer waiting time of two minutes or less)? Discuss.

16. The new Fore and Aft Marina is to be located on the Ohio River near Madison, Indiana. Assume that Fore and Aft decides to build a docking facility where one boat at a time can stop for gas and servicing. Assume that arrivals follow a Poisson probability distribution, with an arrival rate of 5 boats per hour, and that service times follow an exponential probability distribution, with a service rate of 10 boats per hour. Answer the following questions:

a. What is the probability that no boats are in the system?
b. What is the average number of boats that will be waiting for service?
c. What is the average time a boat will spend waiting for service?
d. What is the average time a boat will spend at the dock?
e. If you were the manager of Fore and Aft Marina, would you be satisfied with the service level your system will be providing? Why or why not?

17. The manager of the Fore and Aft Marina in Problem 16 wants to investigate the possibility of enlarging the docking facility so that two boats can stop for gas and servicing simultaneously. Assume that the arrival rate is 5 boats per hour and that the service rate for each server is 10 boats per hour.

a. What is the probability that the boat dock will be idle?
b. What is the average number of boats that will be waiting for service?
c. What is the average time a boat will spend waiting for service?
d. What is the average time a boat will spend at the dock?
e. If you were the manager of Fore and Aft Marina, would you be satisfied with the service level your system will be providing? Why or why not?

SELF test

18. All airplane passengers at the Lake City Regional Airport must pass through a security screening area before proceeding to the boarding area. The airport has three screening stations available, and the facility manager must decide how many to have open at any particular time. The service rate for processing passengers at each screening station is 3 passengers per minute. On Monday morning the arrival rate is 5.4 passengers per minute.

Assume that processing times at each screening station follow an exponential distribution and that arrivals follow a Poisson distribution.

a. Suppose two of the three screening stations are open on Monday morning. Compute the operating characteristics for the screening facility.
b. Because of space considerations, the facility manager's goal is to limit the average number of passengers waiting in line to 10 or fewer. Will the two-screening-station system be able to meet the manager's goal?
c. What is the average time required for a passenger to pass through security screening?

19. Refer again to the Lake City Regional Airport described in Problem 18. When the security level is raised to high, the service rate for processing passengers is reduced to 2 passengers per minute at each screening station. Suppose the security level is raised to high on Monday morning. The arrival rate is 5.4 passengers per minute.
    a. The facility manager's goal is to limit the average number of passengers waiting in line to 10 or fewer. How many screening stations must be open in order to satisfy the manager's goal?
    b. What is the average time required for a passenger to pass through security screening?

20. A Florida coastal community experiences a population increase during the winter months, with seasonal residents arriving from northern states and Canada. Staffing at a local post office is often in a state of change due to the relatively low volume of customers in the summer months and the relatively high volume of customers in the winter months. The service rate of a postal clerk is 0.75 customers per minute. The post office counter has a maximum of three workstations. The target maximum time a customer waits in the system is five minutes.
    a. For a particular Monday morning in November, the anticipated arrival rate is 1.2 customers per minute. What is the recommended staffing for this Monday morning? Show the operating characteristics of the waiting line.
    b. A new population growth study suggests that over the next two years the arrival rate at the postal office during the busy winter months can be expected to be 2.1 customers per minute. Use a waiting line analysis to make a recommendation to the post office manager.

SELF test

21. Refer to the Agan Interior Design situation in Problem 11. Agan's management would like to evaluate two alternatives:
    - Use one consultant with an average service time of 8 minutes per customer.
    - Expand to two consultants, each of whom has an average service time of 10 minutes per customer.

    If the consultants are paid \$16 per hour and the customer waiting time is valued at \$25 per hour for waiting time prior to service, should Agan expand to the two-consultant system? Explain.

22. A fast-food franchise is considering operating a drive-up window food-service operation. Assume that customer arrivals follow a Poisson probability distribution, with an arrival rate of 24 cars per hour, and that service times follow an exponential probability distribution. Arriving customers place orders at an intercom station at the back of the parking lot and then drive to the service window to pay for and receive their orders. The following three service alternatives are being considered:
    - A single-server operation in which one employee fills the order and takes the money from the customer. The average service time for this alternative is 2 minutes.
    - A single-server operation in which one employee fills the order while a second employee takes the money from the customer. The average service time for this alternative is 1.25 minutes.
    - A two-server operation with two service windows and two employees. The employee stationed at each window fills the order and takes the money for customers arriving at the window. The average service time for this alternative is 2 minutes for each server.

Answer the following questions and recommend one of the design options.

a. What is the probability that no cars are in the system?
b. What is the average number of cars waiting for service?
c. What is the average number of cars in the system?
d. What is the average time a car waits for service?
e. What is the average time in the system?
f. What is the probability that an arriving car will have to wait for service?

23. The following cost information is available for the fast-food franchise in Problem 22:
- Customer waiting time is valued at $25 per hour to reflect the fact that waiting time is costly to the fast-food business.
- The cost of each employee is $6.50 per hour.
- To account for equipment and space, an additional cost of $20 per hour is attributable to each server.

What is the lowest-cost design for the fast-food business?

SELF test

24. A study of the multiple-server food-service operation at the Red Birds baseball park shows that the average time between the arrival of a customer at the food-service counter and his or her departure with a filled order is 10 minutes. During the game, customers arrive at the rate of four per minute. The food-service operation requires an average of 2 minutes per customer order.
a. What is the service rate per server in terms of customers per minute?
b. What is the average waiting time in the line prior to placing an order?
c. On average, how many customers are in the food-service system?

25. To understand how a multiple-server waiting line system with a shared queue compares to a multiple-server waiting line system with a dedicated queue for each server, reconsider the Burger Dome example. Suppose Burger Dome establishes two servers but arranges the restaurant layout so that an arriving customer must decide which server's queue to join. Assume that this system equally splits the customer arrivals so that each server sees half of the customers. How does this system compare with the two-server waiting line system with a shared queue from Section 15.3? Compare the average number of customers waiting, average number of customers in the system, average waiting time, and average time in the system.

26. Manning Autos operates an automotive service. To complete their repair work, Manning mechanics often need to retrieve parts from the company's parts department counter. Mechanics arrive at the parts counter at rate of four per hour. The parts coordinator spends an average of six minutes with each mechanic, discussing the parts the mechanic needs and retrieving the parts from inventory.
a. Currently, Manning has one parts coordinator. On average, each mechanic waits four minutes before the parts coordinator is available to answer questions or retrieve parts from inventory. Find L_q, W, and L for this single-server parts operation.
b. A trial period with a second parts coordinator showed that, on average, each mechanic waited only one minute before a parts coordinator was available. Find L_q, W, and L for this two-server parts operation.
c. If the cost of each mechanic is $20 per hour and the cost of each parts coordinator is $12 per hour, is the one-server or the two-server system more economical?

SELF test

27. Gubser Welding, Inc., operates a welding service for construction and automotive repair jobs. Assume that the arrival of jobs at the company's office can be described by a Poisson probability distribution with an arrival rate of two jobs per 8-hour day. The time required to complete the jobs follows a normal probability distribution, with a mean time of 3.2 hours and a standard deviation of 2 hours. Answer the following questions, assuming that Gubser uses one welder to complete all jobs:
a. What is the mean arrival rate in jobs per hour?
b. What is the mean service rate in jobs per hour?

c. What is the average number of jobs waiting for service?
d. What is the average time a job waits before the welder can begin working on it?
e. What is the average number of hours between when a job is received and when it is completed?
f. What percentage of the time is Gubser's welder busy?

28. Jobs arrive randomly at a particular assembly plant; assume that the arrival rate is five jobs per hour. Service times (in minutes per job) do not follow the exponential probability distribution. Two proposed designs for the plant's assembly operation are shown.

	Service Time	
Design	**Mean**	**Standard Deviation**
A	6.0	3.0
B	6.25	0.6

a. What is the service rate in jobs per hour for each design?
b. For the service rates in part (a), what design appears to provide the best or fastest service rate?
c. What are the standard deviations of the service times in hours?
d. Use the *M*/*G*/1 model to compute the operating characteristics for each design.
e. Which design provides the best operating characteristics? Why?

29. The Robotics Manufacturing Company operates an equipment repair business where emergency jobs arrive randomly at the rate of three jobs per 8-hour day. The company's repair facility is a single-server system operated by a repair technician. The service time varies, with a mean repair time of 2 hours and a standard deviation of 1.5 hours. The company's cost of the repair operation is $28 per hour. In the economic analysis of the waiting line system, Robotics uses $35 per hour cost for customers waiting during the repair process.
a. What are the arrival rate and service rate in jobs per hour?
b. Show the operating characteristics, including the total cost per hour.
c. The company is considering purchasing a computer-based equipment repair system that would enable a constant repair time of 2 hours. For practical purposes, the standard deviation is 0. Because of the computer-based system, the company's cost of the new operation would be $32 per hour. The firm's director of operations rejected the request for the new system because the hourly cost is $4 higher and the mean repair time is the same. Do you agree? What effect will the new system have on the waiting line characteristics of the repair service?
d. Does paying for the computer-based system to reduce the variation in service time make economic sense? How much will the new system save the company during a 40-hour workweek?

SELF test

30. A large insurance company maintains a central computing system that contains a variety of information about customer accounts. Insurance agents in a six-state area use telephone lines to access the customer information database. Currently, the company's central computer system allows three users to access the central computer simultaneously. Agents who attempt to use the system when it is full are denied access; no waiting is allowed. Management realizes that with its expanding business, more requests will be made to the central information system. Being denied access to the system is inefficient as well as annoying for agents. Access requests follow a Poisson probability distribution, with a mean of 42 calls per hour. The service rate per line is 20 calls per hour.
a. What is the probability that 0, 1, 2, and 3 access lines will be in use?
b. What is the probability that an agent will be denied access to the system?
c. What is the average number of access lines in use?

d. In planning for the future, management wants to be able to handle $\lambda = 50$ calls per hour; in addition, the probability that an agent will be denied access to the system should be no greater than the value computed in part (b). How many access lines should this system have?

31. Mid-West Publishing Company publishes college textbooks. The company operates an 800 telephone number whereby potential adopters can ask questions about forthcoming texts, request examination copies of texts, and place orders. Currently, two extension lines are used, with two representatives handling the telephone inquiries. Calls occurring when both extension lines are being used receive a busy signal; no waiting is allowed. Each representative can accommodate an average of 12 calls per hour. The arrival rate is 20 calls per hour.
    a. How many extension lines should be used if the company wants to handle 90% of the calls immediately?
    b. What is the average number of extension lines that will be busy if your recommendation in part (a) is used?
    c. What percentage of calls receive a busy signal for the current telephone system with two extension lines?

32. City Cab, Inc., uses two dispatchers to handle requests for service and to dispatch the cabs. The telephone calls that are made to City Cab use a common telephone number. When both dispatchers are busy, the caller hears a busy signal; no waiting is allowed. Callers who receive a busy signal can call back later or call another cab service. Assume that the arrival of calls follows a Poisson probability distribution, with a mean of 40 calls per hour, and that each dispatcher can handle a mean of 30 calls per hour.
    a. What percentage of time are both dispatchers idle?
    b. What percentage of time are both dispatchers busy?
    c. What is the probability that callers will receive a busy signal if two, three, or four dispatchers are used?
    d. If management wants no more than 12% of the callers to receive a busy signal, how many dispatchers should be used?

33. Kolkmeyer Manufacturing Company (see Section 15.9) is considering adding two machines to its manufacturing operation. This addition will bring the number of machines to eight. The president of Kolkmeyer asked for a study of the need to add a second employee to the repair operation. The arrival rate is 0.05 machines per hour for each machine, and the service rate for each individual assigned to the repair operation is 0.50 machines per hour.
    a. Compute the operating characteristics if the company retains the single-employee repair operation.
    b. Compute the operating characteristics if a second employee is added to the machine repair operation.
    c. Each employee is paid \$20 per hour. Machine downtime is valued at \$80 per hour. From an economic point of view, should one or two employees handle the machine repair operation? Explain.

SELF test

34. Five administrative assistants use an office copier. The average time between arrivals for each assistant is 40 minutes, which is equivalent to an arrival rate of $1/40 = 0.025$ arrivals per minute. The mean time each assistant spends at the copier is 5 minutes, which is equivalent to a service rate of $1/5 = 0.20$ per minute. Use the *M*/*M*/1 model with a finite calling population to determine the following:
    a. The probability that the copier is idle
    b. The average number of administrative assistants in the waiting line
    c. The average number of administrative assistants at the copier
    d. The average time an assistant spends waiting for the copier
    e. The average time an assistant spends at the copier

f. During an 8-hour day, how many minutes does an assistant spend at the copier? How much of this time is waiting time?
g. Should management consider purchasing a second copier? Explain.

35. Schips Department Store operates a fleet of 10 trucks. The trucks arrive at random times throughout the day at the store's truck dock to be loaded with new deliveries or to have incoming shipments from the regional warehouse unloaded. Each truck returns to the truck dock for service two times per 8-hour day. Thus, the arrival rate per truck is 0.25 trucks per hour. The service rate is 4 trucks per hour. Using the Poisson arrivals and exponential service times model with a finite calling population of 10 trucks, determine the following operating characteristics:
a. The probability that no trucks are at the truck dock
b. The average number of trucks waiting for loading/unloading
c. The average number of trucks in the truck dock area
d. The average waiting time before loading/unloading begins
e. The average waiting time in the system
f. What is the hourly cost of operation if the cost is $50 per hour for each truck and $30 per hour for the truck dock?
g. Consider a two-server truck dock operation where the second server could be operated for an additional $30 per hour. How much would the average number of trucks waiting for loading/unloading have to be reduced to make the two-server truck dock economically feasible?
h. Should the company consider expanding to the two-server truck dock? Explain.

Case Problem 1 Regional Airlines

Regional Airlines is establishing a new telephone system for handling flight reservations. During the 10:00 A.M. to 11:00 A.M. time period, calls to the reservation agent occur randomly at an average of one call every 3.75 minutes. Historical service time data show that a reservation agent spends an average of 3 minutes with each customer. The waiting line model assumptions of Poisson arrivals and exponential service times appear reasonable for the telephone reservation system.

Regional Airlines' management believes that offering an efficient telephone reservation system is an important part of establishing an image as a service-oriented airline. If the system is properly implemented, Regional Airlines will establish good customer relations, which in the long run will increase business. However, if the telephone reservation system is frequently overloaded and customers have difficulty contacting an agent, a negative customer reaction may lead to an eventual loss of business. The cost of a ticket reservation agent is $20 per hour. Thus, management wants to provide good service, but it does not want to incur the cost of overstaffing the telephone reservation operation by using more agents than necessary.

At a planning meeting, Regional's management team agreed that an acceptable customer service goal is to answer at least 85% of the incoming calls immediately. During the planning meeting, Regional's vice president of administration pointed out that the data show that the average service rate for an agent is faster than the average arrival rate of the telephone calls. The vice president's conclusion was that personnel costs could be minimized by using one agent and that the single agent should be able to handle the telephone reservations and still have some idle time. The vice president of marketing restated the importance of customer service and expressed support for at least two reservation agents.

The current telephone reservation system design does not allow callers to wait. Callers who attempt to reach a reservation agent when all agents are occupied receive a busy signal and are blocked from the system. A representative from the telephone company suggested that Regional Airlines consider an expanded system that accommodates waiting. In the expanded system, when a customer calls and all agents are busy, a recorded message tells the customer that the call is being held in the order received and that an agent will be available shortly. The customer can stay on the line and listen to background music while waiting for an agent. Regional's management will need more information before switching to the expanded system.

Managerial Report

Prepare a managerial report for Regional Airlines analyzing the telephone reservation system. Evaluate both the system that does not allow waiting and the expanded system that allows waiting. Include the following information in your report:

1. A detailed analysis of the operating characteristics of the reservation system with one agent as proposed by the vice president of administration. What is your recommendation concerning a single-agent system?
2. A detailed analysis of the operating characteristics of the reservation system based on your recommendation regarding the number of agents Regional should use.
3. A detailed analysis of the advantages or disadvantages of the expanded system. Discuss the number of waiting callers the expanded system would need to accommodate.
4. This report represents a pilot study of the reservation system for the 10:00 A.M. to 11:00 A.M. time period during which an average of one call arrives every 3.75 minutes; however, the arrival rate of incoming calls is expected to change from hour to hour. Describe how your waiting line analysis could be used to develop a ticket agent staffing plan that would enable the company to provide different levels of staffing for the ticket reservation system at different times during the day. Indicate the information that you would need to develop this staffing plan.

Case Problem 2 Office Equipment, Inc.

Office Equipment, Inc. (OEI) leases automatic mailing machines to business customers in Fort Wayne, Indiana. The company built its success on a reputation of providing timely maintenance and repair service. Each OEI service contract states that a service technician will arrive at a customer's business site within an average of three hours from the time that the customer notifies OEI of an equipment problem.

Currently, OEI has 10 customers with service contracts. One service technician is responsible for handling all service calls. A statistical analysis of historical service records indicates that a customer requests a service call at an average rate of one call per 50 hours of operation. If the service technician is available when a customer calls for service, it takes the technician an average of 1 hour of travel time to reach the customer's office and an average of 1.5 hours to complete the repair service. However, if the service technician is busy with another customer when a new customer calls for service, the technician completes the current service call and any other waiting service calls before responding to the new service call. In such cases, once the technician is free from all existing service commitments, the technician takes an average of 1 hour of travel time to reach the new customer's office and an average of 1.5 hours to complete the repair service. The cost of the service technician is $80 per hour. The downtime cost (wait time and service time) for customers is $100 per hour.

OEI is planning to expand its business. Within one year, OEI projects that it will have 20 customers, and within two years, OEI projects that it will have 30 customers. Although OEI is satisfied that one service technician can handle the 10 existing customers, management is concerned about the ability of one technician to meet the average three-hour service call guarantee when the OEI customer base expands. In a recent planning meeting, the marketing manager made a proposal to add a second service technician when OEI reaches 20 customers and to add a third service technician when OEI reaches 30 customers. Before making a final decision, management would like an analysis of OEI service capabilities. OEI is particularly interested in meeting the average three-hour waiting time guarantee at the lowest possible total cost.

Managerial Report

Develop a managerial report summarizing your analysis of the OEI service capabilities. Make recommendations regarding the number of technicians to be used when OEI reaches 20 customers and when OEI reaches 30 customers. Include a discussion of the following issues in your report:

1. What is the arrival rate for each customer per hour?
2. What is the service rate in terms of the number of customers per hour? Note that the average travel time of 1 hour becomes part of the service time because the time that the service technician is busy handling a service call includes the travel time plus the time required to complete the repair.
3. Waiting line models generally assume that the arriving customers are in the same location as the service facility. Discuss the OEI situation in light of the fact that a service technician travels an average of 1 hour to reach each customer. How should the travel time and the waiting time predicted by the waiting line model be combined to determine the total customer waiting time?
4. OEI is satisfied that one service technician can handle the 10 existing customers. Use a waiting line model to determine the following information:
   - Probability that no customers are in the system
   - Average number of customers in the waiting line
   - Average number of customers in the system
   - Average time a customer waits until the service technician arrives
   - Average time a customer waits until the machine is back in operation
   - Probability that a customer will have to wait more than one hour for the service technician to arrive
   - The total cost per hour for the service operation

   Do you agree with OEI management that one technician can meet the average three-hour service call guarantee? Explain.
5. What is your recommendation for the number of service technicians to hire when OEI expands to 20 customers? Use the information that you developed in part (4) to justify your answer.
6. What is your recommendation for the number of service technicians to hire when OEI expands to 30 customers? Use the information that you developed in part (4) to justify your answer.
7. What are the annual savings of your recommendation in part (6) compared to the planning committee's proposal that 30 customers will require three service technicians? Assume 250 days of operation per year.