CHAPTER 11: WAITING LINE MODELS

CHAPTER 11: WAITING LINE MODELS

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.

A waiting line is also known as a **queue**, and the body of knowledge dealing with waiting lines is known as **queueing theory.**

WAITING LINE MODELS

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:

1. The probability that no units are in the system
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

STRUCTURE OF A WAITING LINE SYSTEM
(1 OF 9)

Consider the waiting line at the Burger Dome fast-food restaurant. Burger Dome sells cheeseburgers, French fries, soft drinks, and other items.

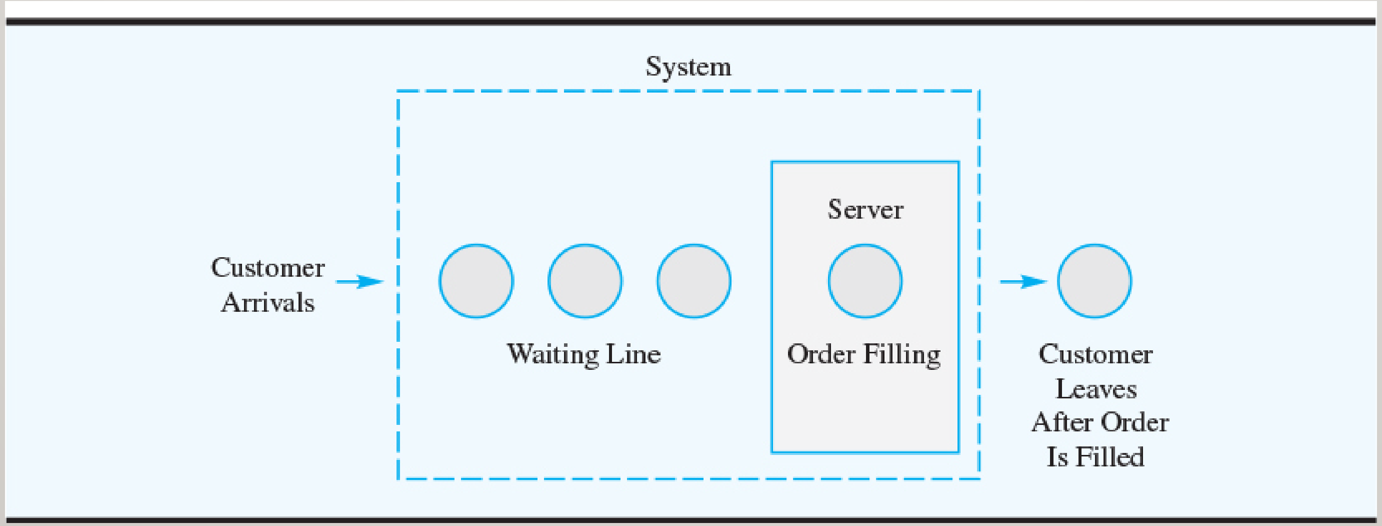
Although Burger Dome would like to serve each customer immediately, at times more customers arrive than can be handled by the Burger Dome food service staff. Thus, customers wait in line to place and receive their orders.

Burger Dome is concerned that the methods currently used to serve customers are resulting in excessive waiting times and a possible loss of sales.

STRUCTURE OF A WAITING LINE SYSTEM

(2 OF 9)

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**.



STRUCTURE OF A WAITING LINE SYSTEM (3 OF 9)

For many waiting line situations, the arrivals occur *randomly and independently* of other arrivals, and we cannot predict when an arrival will occur. In such cases, 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

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

x = the number of arrivals in the time period

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

e = 2.71828

STRUCTURE OF A WAITING LINE SYSTEM
(4 OF 9)

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.

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

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

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

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

STRUCTURE OF A WAITING LINE SYSTEM
(5 OF 9)

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.

STRUCTURE OF A WAITING LINE SYSTEM (6 OF 9)

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

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

where

μ = the *mean* number of units that can be served per time period, also known as the **service rate**.

e = 2.71828

STRUCTURE OF A WAITING LINE SYSTEM

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 μ = 60 customers ÷ 60 minutes = 1 customer per minute.

For example:

$$P(\text{service time} \leq 0.5 \text{ min.}) = 1 - e^{-1(0.5)} = 1 - 0.6065 = 0.3935$$

$$P(\text{service time} \leq 1.0 \text{ min.}) = 1 - e^{-1(1.0)} = 1 - 0.3679 = 0.6321$$

$$P(\text{service time} \leq 2.0 \text{ min.}) = 1 - e^{-1(2.0)} = 1 - 0.1353 = 0.8647$$

STRUCTURE OF A WAITING LINE SYSTEM
(8 OF 9)

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.

- Boarding an airplane
- Emergency rooms

This chapter considers only a FCFS que discipline.

STRUCTURE OF A WAITING LINE SYSTEM

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.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (1 OF 11)

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.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (2 OF 11)

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)

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (3 OF 11)

1. The probability that no units are in the system

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

2. The average number of units in the waiting line

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

3. The average number of units in the system

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

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (4 OF 11)

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

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

5. The average time a unit spends in the system

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

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (5 OF 11)

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

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

7. The probability of *n* units in the system

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

These equations are applicable only when the service rate μ is *greater than* the arrival rate λ (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.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (6 OF 11)

Recall that for the Burger Dome problem we had an arrival rate of

λ = 0.75 customers per minute and a service rate of

μ = 1 customer per minute. Therefore:

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

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{0.75^2}{1(1 - .075)} = 2.25 \text{ customers}$$

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

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

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

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

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (7 OF 11)

The results:

- Customers wait an average of three minutes before beginning to place an order, which appears somewhat long for a business based on fast service.
- The average number of customers waiting in line is 2.25
- 75% of the arriving customers have to wait for service.
- There is a 0.1335 probability that seven or more customers are in the Burger Dome system at one time.

These are indicators that something should be done to improve the waiting line operation. 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.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (8 OF 11)

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.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (9 OF 11)

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 that the service rate can be increased from the current 60 customers per hour to 75 customers per hour.

Thus, the service rate for the revised system is

μ = 75 customers ÷ 60 minutes = 1.25 customers per minute.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (10 OF 11)

The new operating characteristics are:

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 min
Average time in the system	2.000 min
Probability that an arriving customer has to wait	0.600
Probability that seven or more customers are in the system	0.028

- The average time a customer spends in the waiting line has been reduced from 3 to 1.2 minutes.
- The average time a customer spends in the system has been reduced from 4 to 2 minutes.

SINGLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (11 OF 11)

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.

Another option often available is to add one or more servers so that orders for multiple customers can be filled simultaneously.

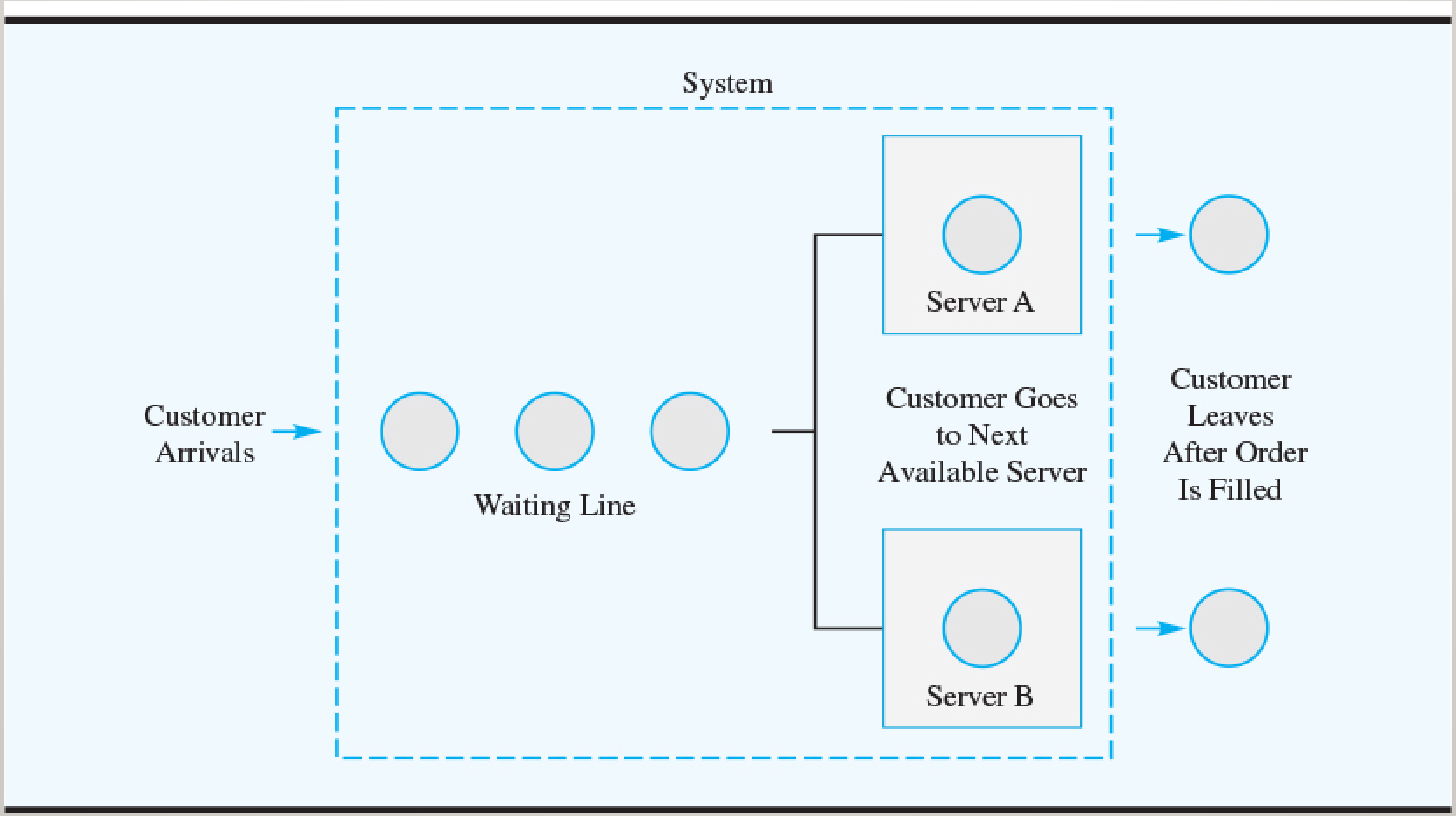
MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (1 OF 12)

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 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.

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (2 OF 12)



MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (3 OF 12)

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.

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (4 OF 12)

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

$$\lambda = \text{the arrival rate for the system}$$

$$\mu = \text{the service rate for } \textit{each} \text{ server}$$

$$k = \text{the number of servers}$$

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (5 OF 12)

1. The probability that no units are in the system

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

2. The average number of units in the waiting line

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

3. The average number of units in the system

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

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (6 OF 12)

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

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

5. The average time a unit spends in the system

$$W = W_q = \frac{1}{\mu}$$

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (7 OF 12)

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

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

7. The probability of *n* units in the system

$$P_w = \left(\frac{\lambda/\mu}{n!}\right)^n P_0 \qquad \text{for } n \le k$$

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

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (8 OF 12)

Because μ is the service rate for each server, k_{μ} 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_{μ} is greater than λ.

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (9 OF 12)

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 the previous equations for the k = 2-server system using an arrival rate of λ = 0.75 customers per minute and a service rate of μ = 1 customer per minute for each server.

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (10 OF 12)

The results:

$$P_0 = 0.4545 \quad \text{(from Table 11.4 with } \lambda/\mu = 0.75\text{)}$$

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

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

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

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

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

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (11 OF 12)

Compared to the single-server system:

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 $P_w = 0.2045$.

MULTIPLE-SERVER WAITING LINE MODEL WITH POISSON ARRIVALS AND EXPONENTIAL SERVICE TIMES (12 OF 12)

Clearly the two-server system will substantially improve the operating characteristics of the waiting line.

Burger Dome adopted the 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.

SOME GENERAL RELATIONSHIPS FOR WAITING LINE MODELS (1 OF 2)

John D. C. Little showed that several relationships exist among these 4 characteristics, regardless of the waiting line model:

1. The average number of units in the waiting line
2. The average number of units in the system
3. The average time a unit spends in the waiting line
4. The average time a unit spends in the system

SOME GENERAL RELATIONSHIPS FOR WAITING LINE MODELS
(2 OF 2)

Little's flow equations:

1. 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*.

$$L = \lambda W$$

$$L_q = \lambda W_q$$

2. The average time in the system, *W*, is equal to the average time in the wait line, W_q, plus the average service time.

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

ECONOMIC ANALYSIS OF WAITING LINES

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.

OTHER WAITING LINE MODELS (1 OF 2)

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.

OTHER WAITING LINE MODELS (2 OF 2)

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 would be classified as an *M*/*M*/2 model.