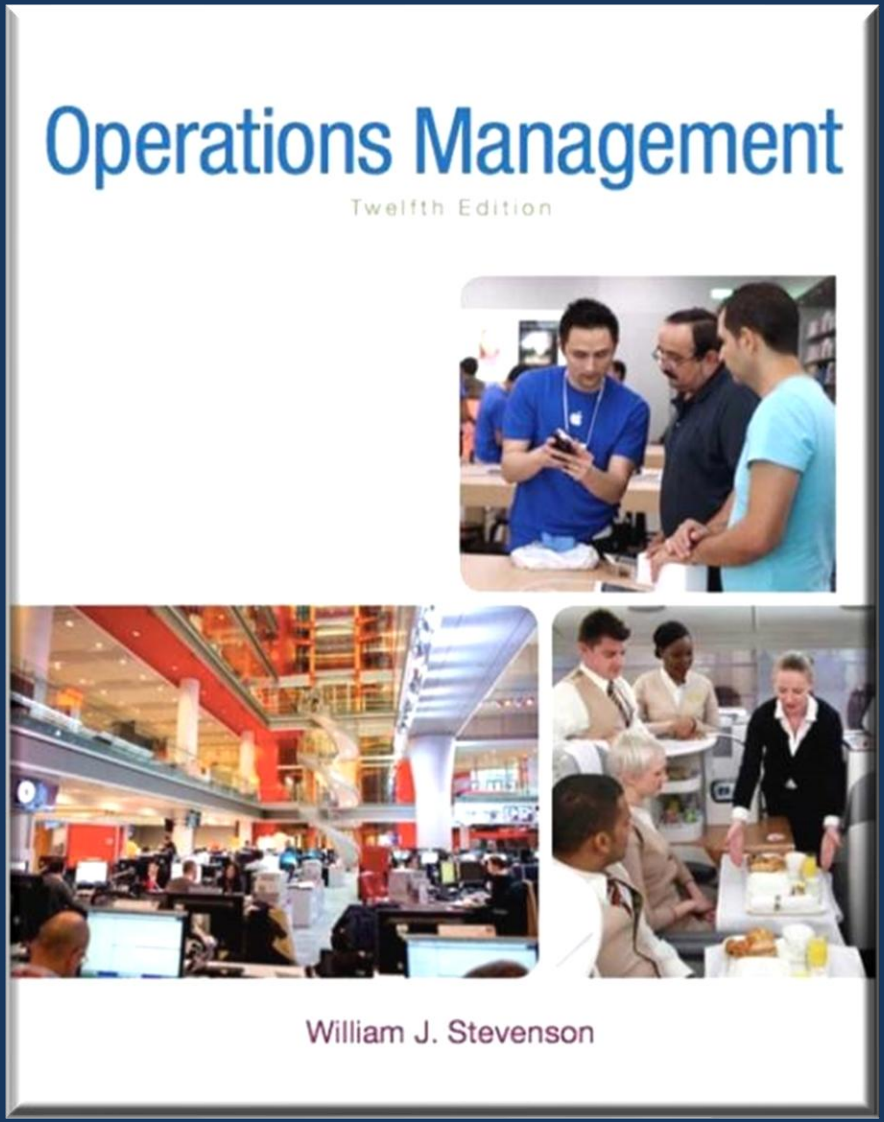


Chapter 18

Management of Waiting Lines

Waiting Lines

- **Waiting lines occur in all sorts of service systems**
- **Wait time is non-value added**
  - Wait time ranges from the acceptable to the emergent
    - Short waits in a drive-thru
    - Sitting in an airport waiting for a delayed flight
    - Waiting for emergency service personnel
  - Waiting time costs
    - Lower productivity
    - Reduced competitiveness
    - Wasted resources
    - Diminished quality of life

LO 18.1

Why Is There Waiting?

- **Waiting lines tend to form even when a system is not fully loaded**
  - Variability
    - Arrival and service rates are variable
  - Services cannot be completed ahead of time and stored for later use

LO 18.2

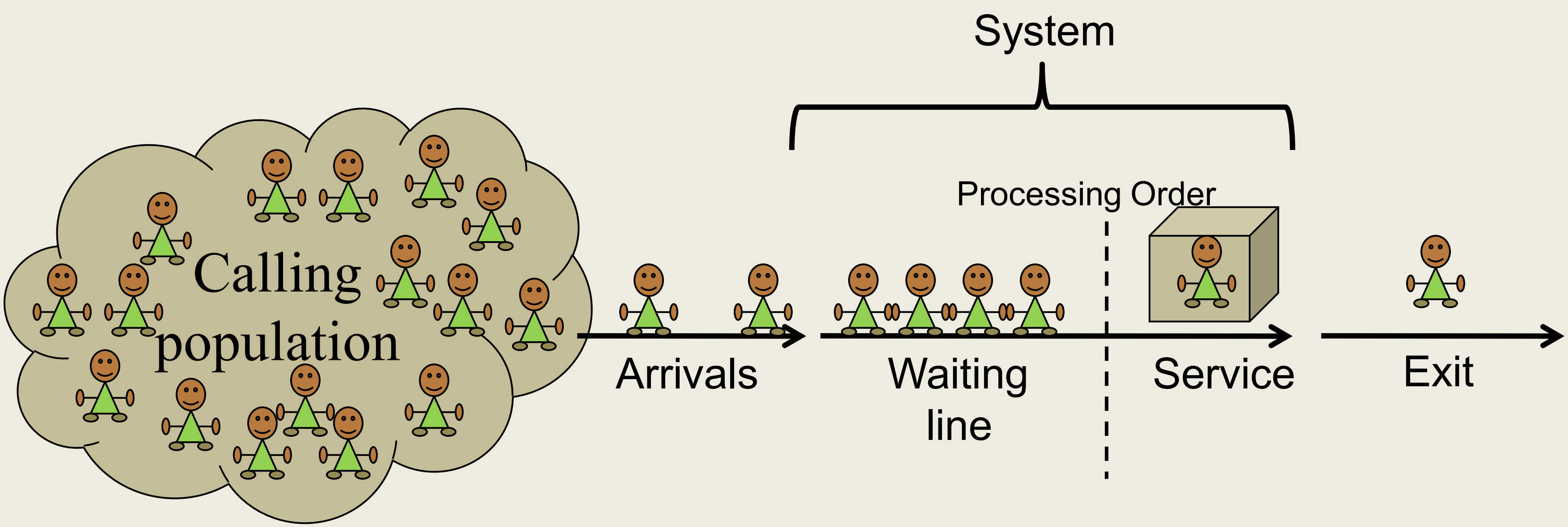
Managerial Implications of Waiting lines

- The cost to provide waiting space
- A possible loss of customers.
- A possible loss of goodwill.
- A possible reduction in customer satisfaction.
- Disruption in other business operation.

Characteristics of Waiting lines

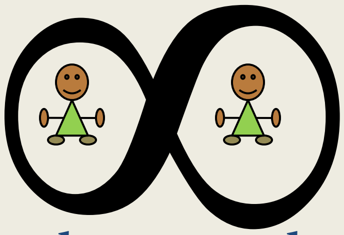
- Population source.
- Number of Servers.
- Arrival and service patterns.
- Queue discipline

Simple Queuing System

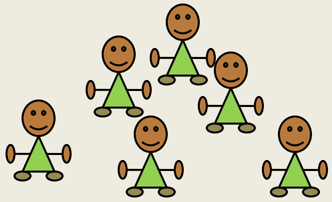


Population Source

- **Infinite source**
  - Customer arrivals are unrestricted
  - The number of potential customers greatly exceeds system capacity



- **Finite source**
  - The number of potential customers is limited



Channels and Phases

- **Channel**
  - A server in a service system
  - It is assumed that each channel can handle one customer at a time
- **Phases**
  - The number of steps in a queuing system

Arrival and Service Patterns

- **Arrival pattern**
  - Most commonly used models assume the *arrival rate* can be described by the Poisson distribution
    - Arrivals per unit of time
  - Equivalently, *interarrival* times are assumed to follow the negative exponential distribution
    - The time between arrivals
- **Service pattern**
  - Service times are frequently assumed to follow a negative exponential distribution

Poisson Distribution

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \text{for } x = 0, 1, 2, 3, 4, \ldots$$

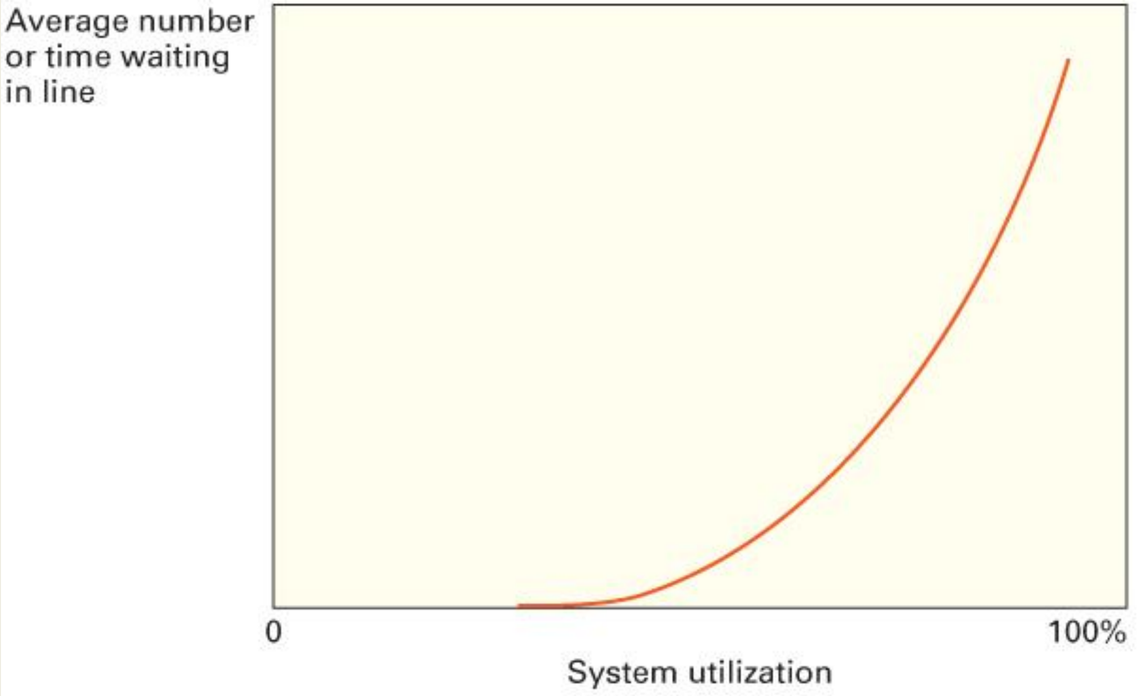
where	$P(x)$	=	probability of x arrivals
	x	=	number of arrivals per unit of time
	λ	=	average arrival rate
	e	=	2.7183 (which is the base of the natural logarithms)

Waiting Line Metrics

- **Managers typically consider five measures when evaluating waiting line performance:**
  1. The average number of customers waiting (in line or in the system)
  2. The average time customers wait (in line or in the system)
  3. System utilization
  4. The implied cost of a given level of capacity and its related waiting line
  5. The probability that an arrival will have to wait for service

LO 18.3

Waiting Line Performance



The average number waiting in line and the average time customers wait in line increase exponentially as the system utilization increases

LO 18.3

Queuing Models: Infinite Source

- **Four basic infinite source models**
  - All assume a Poisson arrival rate
    1. Single server, exponential service time
    2. Single server, constant service time
    3. Multiple servers, exponential service time
    4. Multiple priority service, exponential service time

Queuing Models: Infinite Source

Symbol	Represents
λ	Customer arrival rate
μ	Service rate per server
L_q	The average number of customers waiting for service
L_s	The average number of customers in the system (waiting and/or being served)
r	The average number of customers being served
ρ	The system utilization
W_q	The average time customers wait in line
W_s	The average time customers spend in the system (waiting in line and service time)
$1/\mu$	Service time
P_0	The probability of zero units in the system
P_n	The probability of *n* units in the system
M	The number of servers
L_{max}	The maximum expected number waiting in line

Basic relationship

- **System utilization:**

$$\rho = \frac{\lambda}{M\mu}$$

λ	Customer arrival rate
M	The number of servers
μ	Service rate per server

- **Average number of customers being served:**

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

Basic relationship

- **The average number of customers:**

$$L_s = L_q + r$$

$$L_s = \lambda W_s$$

$$L_q = \lambda W_q$$

W_q	The average time customers wait in line
W_s	The average time customers spend in the system

- **The average time customers are**

$$\text{Waiting in line:} \quad W_q = \frac{L_q}{\lambda}$$

$$\text{In the system:} \quad W_s = W_q + \frac{1}{\mu} = \frac{L_s}{\lambda}$$

Example 1

Customers arrive at a bakery at an average rate of 18 per hour on weekday mornings. The arrival distribution can be described by a Poisson distribution with a mean of 18. Each clerk can serve a customer in an average of three minutes; this time can be described by an exponential distribution with a mean of 3.0 minutes.

a. What are the arrival and service *rates?*

b. Compute the average number of customers being served at any time.

c. Suppose it has been determined that the average number of customers waiting in line is 8.1. Compute the average number of customers in the system (i.e., waiting in line or being served), the average time customers wait in line, and the average time in the system.

d. Determine the system utilization for M 1, 2, and 3 servers.

a. The arrival rate is given in the problem: λ = 18 customers per hour. Change the service *time* to a comparable hourly rate. Thus,

60 minutes per hour/3 minutes per customer = μ = 20 customers per hour

b. $r = \frac{\lambda}{\mu} = \frac{18}{20} = .90$ customer.

c. Given: $L_q = 8.1$ customers.

$$L_s = L_q + r = 8.1 + .90 = 9.0 \text{ customers}$$

$$W_q = \frac{L_q}{\lambda} = \frac{8.1}{18} = .45 \text{ hour}$$

$$W_s = \text{Waiting in line plus service}$$

$$= W_q + \frac{1}{\mu} = .45 + \frac{1}{20} = .50 \text{ hour}$$

d. System utilization is $\rho = \frac{\lambda}{M\mu}$.

$$\text{For } M = 1,\ \rho = \frac{18}{1(20)} = .90$$

$$\text{For } M = 2,\ \rho = \frac{18}{2(20)} = .45$$

$$\text{For } M = 3,\ \rho = \frac{18}{3(20)} = .30$$

Single Server, Exponential Service Time

- **M/M/1**

Performance Measure	Equation
Average number in line	$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Probability of zero units in the system	$P_0 = 1 - \left(\frac{\lambda}{\mu}\right)$
Probability of *n* units in the system	$P_n = P_0\left(\frac{\lambda}{\mu}\right)^n$
Probability of less than *n* units in the system	$P_{<n} = 1 - \left(\frac{\lambda}{\mu}\right)^n$

Example 2

An airline is planning to open a satellite ticket desk in a new shopping plaza, staffed by one ticket agent. It is estimated that requests for tickets and information will average 15 per hour, and requests will have a Poisson distribution. Service time is assumed to be exponentially distributed. Previous experience with similar satellite operations suggests that mean service time should average about three minutes per request. Determine each of the following:

a. System utilization.

b. Percentage of time the server (agent) will be idle.

c. The expected number of customers waiting to be served.

d. The average time customers will spend in the system.

e. The probability of zero customers in the system and the probability of four customers in the system.

$$\lambda = 15 \text{ customers per hour}$$

$$\mu = \frac{1}{\text{Service time}} = \frac{1 \text{ customer}}{3 \text{ minutes}} \times 60 \text{ minutes per hour}$$
$$= 20 \text{ customers per hour}$$

a. $\rho = \dfrac{\lambda}{M\mu} = \dfrac{15}{1(20)} = .75$

b. Percentage idle time $= 1 - \rho = 1 - .75 = .25$, or 25 percent

c. $L_q = \dfrac{\lambda^2}{\mu(\mu - \lambda)} = \dfrac{15^2}{20(20 - 15)} = 2.25$ customers

d. $W_s = \dfrac{L_q}{\lambda} + \dfrac{1}{\mu} = \dfrac{2.25}{15} + \dfrac{1}{20} = .20$ hour, or 12 minutes

e. $P_0 = 1 - \dfrac{\lambda}{\mu} = 1 - \dfrac{15}{20} = .25$ and $P_4 = P_0\left(\dfrac{\lambda}{\mu}\right)^4 = .25\left(\dfrac{15}{20}\right)^4 = .079$

Single Server, Constant Service Time

- **M/D/1**
  - If a system can reduce variability, it can shorten waiting lines noticeably
  - For, example, by making service time constant, the average number of customers waiting in line can be cut in half

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

  - Average time customers spend waiting in line is also cut by half.
  - Similar improvements can be made by smoothing arrival rates (such as by use of appointments)

LO 18.4

Example 3

Wanda's Car Wash & Dry is an automatic, five-minute operation with a single bay. On a typical Saturday morning, cars arrive at a mean rate of eight per hour, with arrivals tending to follow a Poisson distribution. Find

a. The average number of cars in line.

b. The average time cars spend in line and service.

Example 3

Wanda's Car Wash & Dry is an automatic, five-minute operation with a single bay. On a typical Saturday morning, cars arrive at a mean rate of eight per hour, with arrivals tending to follow a Poisson distribution. Find

a. The average number of cars in line.

b. The average time cars spend in line and service.

λ = 8 cars per hour

μ = 1 per 5 minutes, or 12 per hour

Psychology of Waiting

- If those waiting in line have nothing else to occupy their thoughts, they often tend to focus on the fact they are waiting in line
  - They will usually perceive the waiting time to be longer than the actual waiting time
  - Steps can be taken to make waiting more acceptable to customers
    - Occupy them while they wait
      - In-flight snack
      - Have them fill out forms while they wait
    - Make the waiting environment more comfortable
    - Provide customers information concerning their wait

Operations Strategy

- **Managers must carefully weigh the costs and benefits of service system capacity alternatives**
- **Options for reducing wait times:**
  - Work to increase processing rates, instead of increasing the number of servers
  - Use new processing equipment and/or methods
  - Reduce processing time variability through standardization
  - Shift demand