# Chapter 13

# Waiting Lines and Queuing Theory Models

# *Chapter Outline*

**13.1 Introduction**

**13.2 Waiting Line Costs**

**13.3 Characteristics of a Queuing System**

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

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

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

# Introduction

- ***Queuing theory*** **is the study of** ***waiting lines.***
- **It is one of the oldest and most widely used quantitative analysis techniques.**
- **The three basic components of a queuing process are arrivals, service facilities, and the actual waiting line.**
- **Analytical models of waiting lines can help managers evaluate the cost and effectiveness of service systems.**

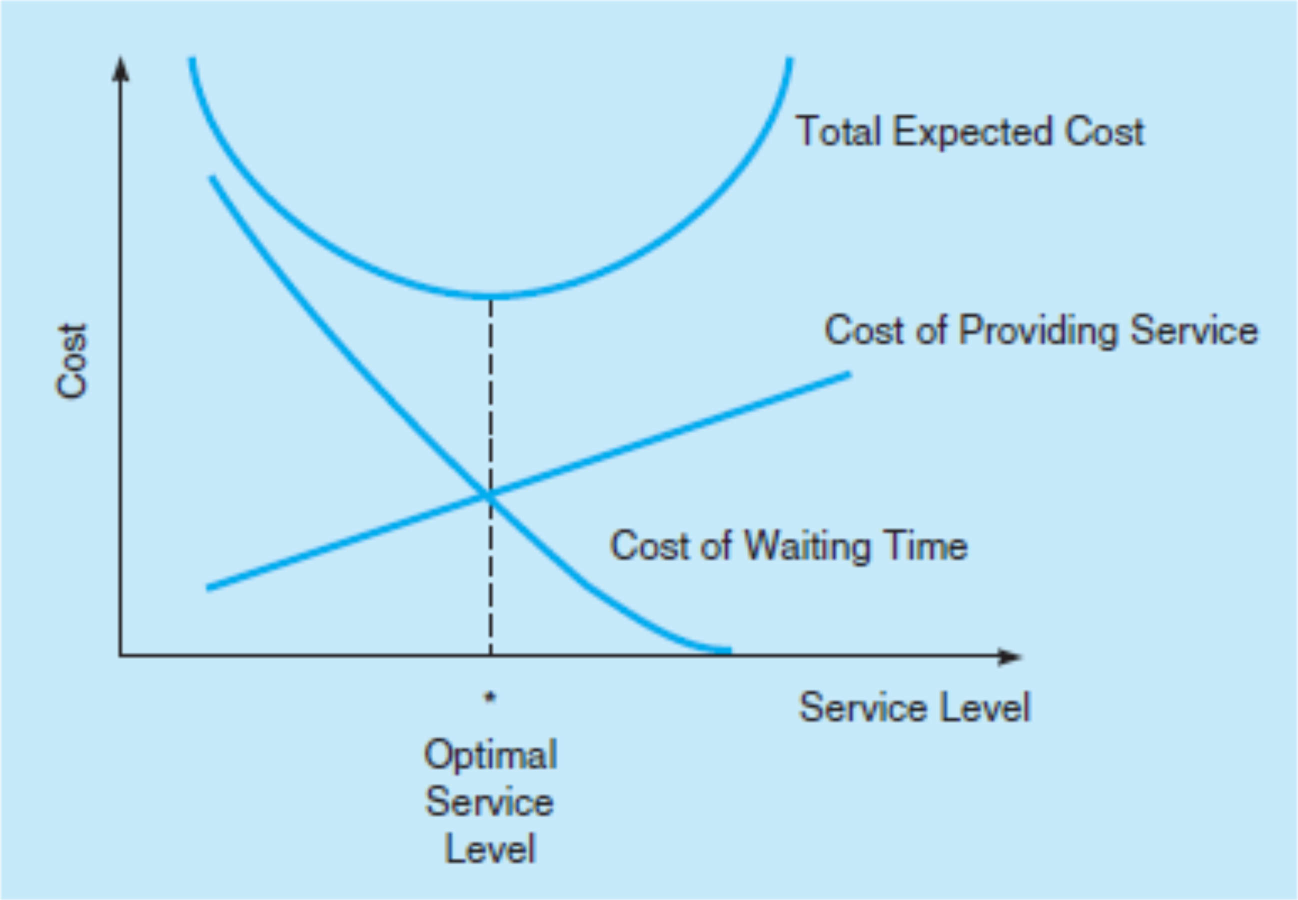
# *Waiting Line Costs*

- Most waiting line problems are focused on finding the ideal level of service a firm should provide.
- In most cases, this service level is something management can control.
- When an organization ***does*** have control, they often try to find the balance between two extremes.

# *Waiting Line Costs*

- There is generally a trade-off between cost of providing service and cost of waiting time.
  - A *large staff* and *many* service facilities generally results in high levels of service but have high costs.
  - Having the *minimum* number of service facilities keeps *service cost* down but may result in dissatisfied customers.
- Service facilities are evaluated on their *total expected cost* which is the sum of *service costs* and *waiting costs.*
- Organizations typically want to find the service level that minimizes the total expected cost.

# Queuing Costs and Service Levels



# *Three Rivers Shipping Company*

- Three Rivers Shipping operates a docking facility on the Ohio River.
- An average of 5 ships arrive to unload their cargos each shift.
- Idle ships are expensive.
- More staff can be hired to unload the ships, but that is expensive as well.
- Three Rivers Shipping Company wants to determine the optimal number of teams of stevedores to employ each shift to obtain the minimum total expected cost.

# Three Rivers Shipping Company Waiting Line Cost Analysis

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

*Optimal cost*

Table 13.1

# *Characteristics of a Queuing System*

- There are three parts to a queuing system:
  1. The arrivals or inputs to the system (sometimes referred to as the ***calling population***).
  2. The queue or waiting line itself.
  3. The service facility.
- These components have their own characteristics that must be examined before mathematical models can be developed.

# *Characteristics of a Queuing System*

**Arrival Characteristics have three major characteristics: *size*, *pattern*, and *behavior.***

- **The size of the calling population can be either unlimited (essentially *infinite*) or limited (*finite*). *An example of a finite population is a shop with only eight machines that might break down and require service.***
- **Most cases consider unlimited size**
- **The pattern of arrivals can arrive according to a known pattern(Ex: Scheduled) or can arrive *randomly.***
  - **Random arrivals generally follow a *Poisson distribution.* (predicted by this model)**

# *Characteristics of a Queuing System*

Behavior of arrivals

- Most queuing models assume customers are patient and will wait in the queue until they are served and do not switch lines.
- ***Balking*** refers to customers who refuse to join the queue.
- ***Reneging*** customers enter the queue but become impatient and leave without receiving their service.
- That these behaviors exist is a strong argument for the use of queuing theory to managing waiting lines.

# *Characteristics of a Queuing System*

## Waiting Line Characteristics

- Waiting lines can be either ***limited*** (Ex, Restaurant with limited seating)or ***unlimited.***
- Queue discipline refers to the rule by which customers in the line receive service.
  - The most common rule is ***first-in, first-out*** (***FIFO***).
  - Other rules are possible and may be based on other important characteristics.
- Other rules can be applied to select which customers enter which queue, but may apply FIFO once they are in the queue.(EX: emergencies in hospitals)

# *Characteristics of a Queuing System*

## Service Facility Characteristics

- Basic queuing system configurations:
  - Service systems are classified in terms of the number of channels, or servers, and the number of phases, or service stops.
  - A ***single-channel system*** with one server is quite common.
  - ***Multichannel systems*** exist when multiple servers are fed by one common waiting line.
  - In a ***single-phase system,*** the customer receives service form just one server.
  - In a ***multiphase system,*** the customer has to go through more than one server.

# Four basic queuing system configurations

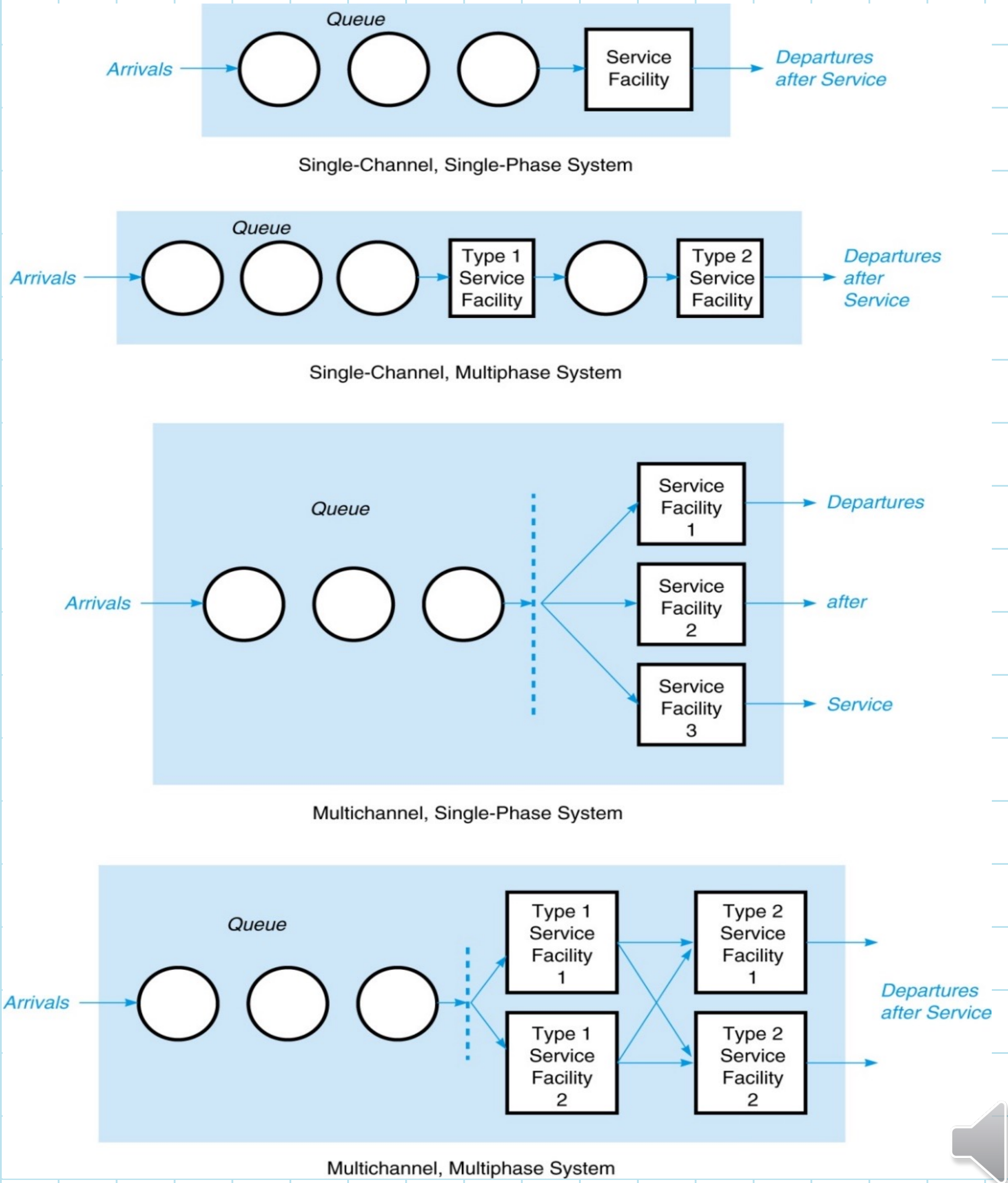
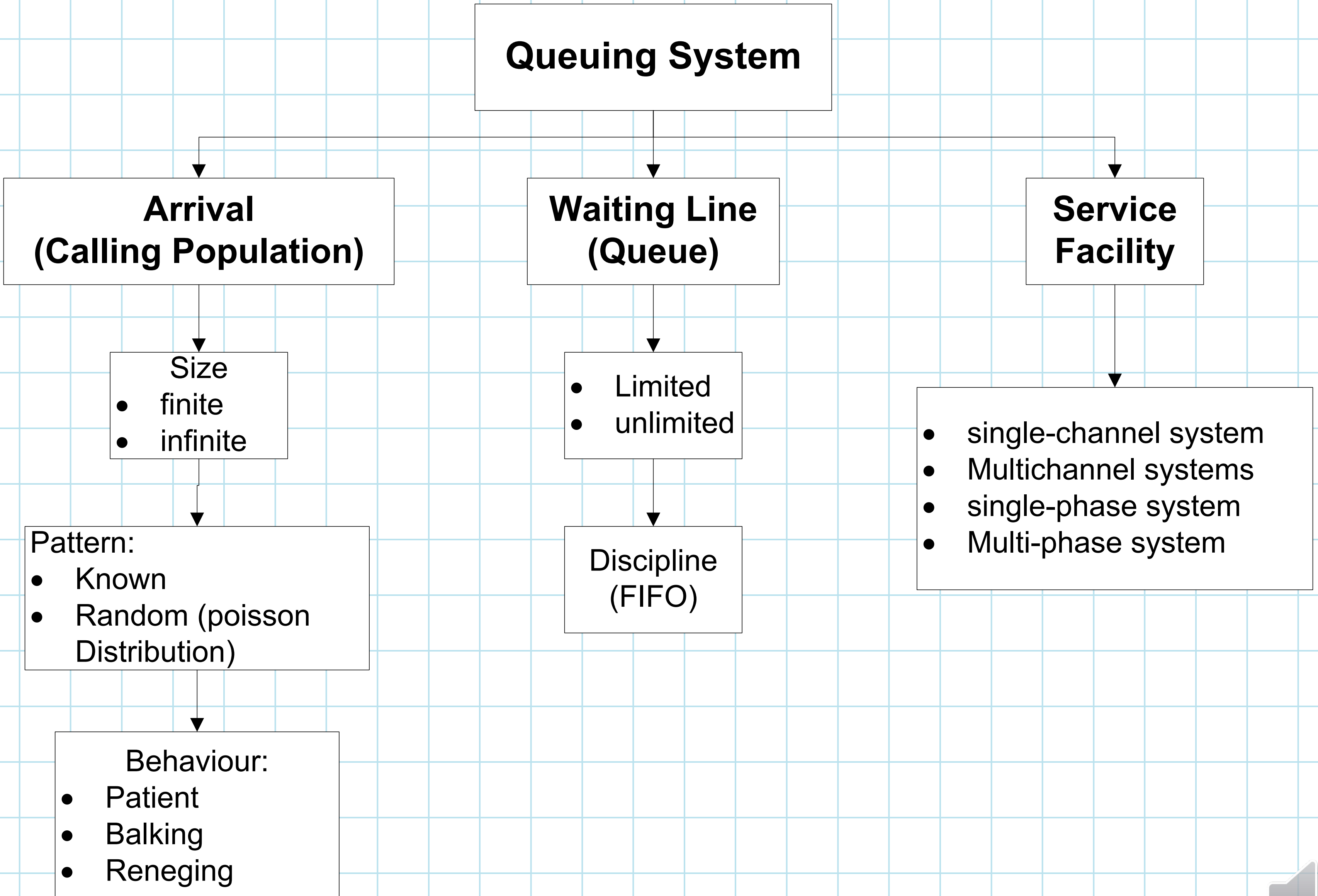



Figure 13.2

# Summary of Queueing System characteristics



# *Characteristics of a Queuing System*

**Service time distribution**

- **Service patterns can be either constant or random.**
- **Constant service times are often machine controlled.**
- **More often, service times are randomly distributed according to a *negative exponential probability distribution.***
- **Analysts should observe, collect, and plot service time data to ensure that the observations fit the assumed distributions when applying these models.**

# Identifying Models Using Kendall Notation

- D. G. Kendall developed a notation for queuing models that specifies the pattern of arrival, the service time distribution, and the number of channels.
- Notation takes the form:

$$\text{Arrival distribution} \,/\, \text{Service time distribution} \,/\, \text{Number of service channels open}$$

- Specific letters are used to represent probability distributions.

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

*D* = constant (deterministic) rate

# *Identifying Models Using Kendall Notation*

- A single-channel model with Poisson arrivals and exponential service times would be represented by:

$$M/M/1$$

- If a second channel is added the notation would read:

$$M/M/2$$

- A three-channel system with Poisson arrivals and constant service time would be

$$M/D/3$$

# *Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)*

**Assumptions of the model:**

- **Arrivals are served on a FIFO basis.**
- **There is no balking or reneging.**
- **Arrivals are independent of each other but the arrival rate is constant over time.**
- **Arrivals follow a Poisson distribution.**
- **Service times are variable and independent but the average is known.**
- **Service times follow a negative exponential distribution.**
- **Average service rate is greater than the average arrival rate.**

## *Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)*

- When these assumptions are met, we can develop a series of equations that define the queue's ***operating characteristics.***
- Queuing Equations:

  Let

  $\lambda$ = mean number of arrivals per time period

  $\mu$ = mean number of customers or units served per time period

  The arrival rate and the service rate must be defined for the same time period.

# Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)

1. The average number of customers or units in the system, $L$:

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

2. The average time a customer spends in the system, $W$:

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

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

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

# *Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)*

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

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

5. The ***utilization factor*** for the system, $\rho$, the probability the service facility is being used:

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

# Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)

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

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

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

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

# *Arnold's Muffler Shop*

- Arnold's mechanic can install mufflers at a rate of 3 per hour.
- Customers arrive at a rate of 2 per hour.
- So:

  $\lambda$ = 2 cars arriving per hour

  $\mu$ = 3 cars serviced per hour

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

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

# *Arnold's Muffler Shop*

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

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

$$\rho = \frac{\lambda}{\mu} = \frac{2}{3} = 0.67$$ **= percentage of time mechanic is busy**

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

# Arnold's Muffler Shop

## Probability of more than $k$ cars in the system

| k | $P_{n>k} = (^2/_3)^{k+1}$ | |
|---|---|---|
| 0 | 0.667 | Note that this is equal to $1 - P_0 = 1 - 0.33 = 0.667$ |
| 1 | 0.444 | |
| 2 | 0.296 | |
| 3 | 0.198 | Implies that there is a 19.8% chance that more than 3 cars are in the system |
| 4 | 0.132 | |
| 5 | 0.088 | |
| 6 | 0.058 | |
| 7 | 0.039 | |

# *Arnold's Muffler Shop*

- Introducing costs into the model:
  - Arnold wants to do an economic analysis of the queuing system and determine the waiting cost and service cost.
  - The total service cost is:

$$\text{Total service cost} = (\text{Number of channels}) \times (\text{Cost per channel})$$

$$\text{Total service cost} = mC_s$$

# *Arnold's Muffler Shop*

Waiting cost when the cost is based on time in the system:

Total waiting cost = (Total time spent waiting by all arrivals) x (Cost of waiting)

= (Number of arrivals) x (Average wait per arrival)$C_w$

Total waiting cost = $(\lambda W)C_w$

If waiting time cost is based on time in the queue:

Total waiting cost = $(\lambda W_q)C_w$

# *Arnold's Muffler Shop*

So the total cost of the queuing system when based on time in the system is:

Total cost = Total service cost + Total waiting cost

Total cost = $mC_s + \lambda W C_w$

And when based on time in the queue:

Total cost = $mC_s + \lambda W_q C_w$

# *Arnold's Muffler Shop*

- Arnold estimates the cost of customer ***waiting*** time in line is $50 per hour.

Total daily waiting cost = (8 hours per day)$\lambda W_q C_w$

= (8)(2)($^2/_3$)($50) = $533.33

- Arnold has identified the mechanics wage $15 per hour as the ***service*** cost.

Total daily service cost = (8 hours per day)$mC_s$

= (8)(1)($15) = $120

- So the **total cost** of the system is:

Total daily cost of the queuing system = $533.33 + $120 = $653.33

# *Arnold's Muffler Shop*

- Arnold is thinking about hiring a different mechanic who can install mufflers at a faster rate.
- The new operating characteristics would be:

  $\lambda$ = 2 cars arriving per hour

  $\mu$ = 4 cars serviced per hour

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

$$W = \frac{1}{\mu - \lambda} = \frac{1}{4-2} = 1/2 \text{ hour that an average car spends in the system}$$

# *Arnold's Muffler Shop*

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

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

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

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{2}{4} = 0.5$$ = probability that there are 0 cars in the system

# *Arnold's Muffler Shop*

**Probability of more than $k$ cars in the system**

| k | $P_{n>k} = (^2/_4)^{k+1}$ |
|---|---|
| 0 | 0.500 |
| 1 | 0.250 |
| 2 | 0.125 |
| 3 | 0.062 |
| 4 | 0.031 |
| 5 | 0.016 |
| 6 | 0.008 |
| 7 | 0.004 |

# *Arnold's Muffler Shop Case*

- The customer waiting cost is the same \$50 per hour:

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

  $= (8)(2)(^1/_4)(\$50) = \$200.00$

- The new mechanic is more expensive at \$20 per hour:

  Total daily service cost $= (8 \text{ hours per day}) m C_s$

  $= (8)(1)(\$20) = \$160$

- So the total cost of the system is:

  Total daily cost of the queuing system $= \$200 + \$160 = \$360$