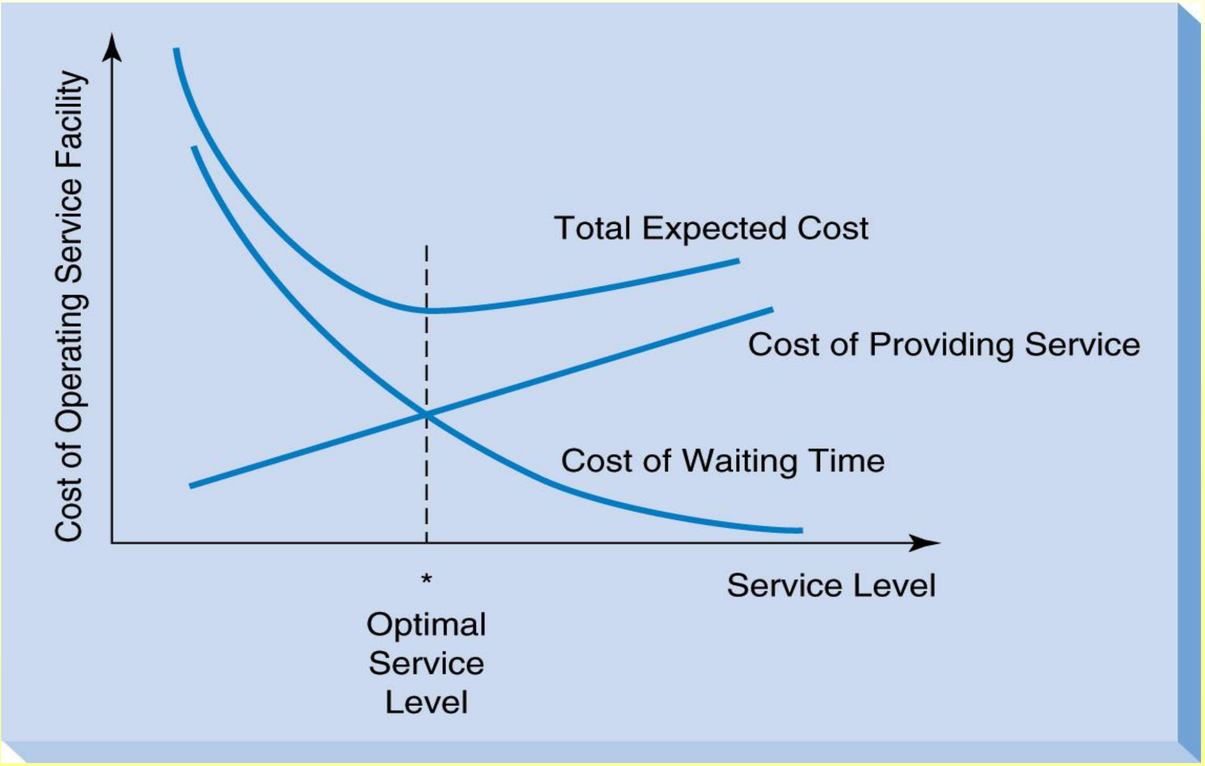
Chapter 9:
Queuing Models

Queuing or Waiting Line Analysis

- Queues (waiting lines) affect people everyday
- A primary goal is finding the best level of service
- Analytical modeling (using formulas) can be used for many queues
- For more complex situations, computer simulation is needed

Queuing System Costs

1. Cost of providing service
2. Cost of <u>not</u> providing service (waiting time)



Characteristics of a Queuing System

The queuing system is determined by:

- Arrival characteristics
- Queue characteristics
- Service facility characteristics

Arrival Characteristics

- Size of the arrival population – either infinite or limited
- Arrival distribution:
  - Either fixed or random
  - Either measured by time between consecutive arrivals, or arrival rate
  - The **Poisson distribution** is often used for random arrivals

Poisson Distribution

- Average arrival rate is known
- Average arrival rate is constant for some number of time periods
- Number of arrivals in each time period is independent
- As the time interval approaches 0, the average number of arrivals approaches 0

Poisson Distribution

λ = the average arrival rate per time unit

P(x) = the probability of exactly x arrivals occurring during one time period

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

Behavior of Arrivals

- Most queuing formulas assume that all arrivals stay until service is completed
- **Balking** refers to customers who do not join the queue
- **Reneging** refers to customers who join the queue but give up and leave before completing service

Queue Characteristics

- Queue length (max possible queue length) – either limited or unlimited
- Service discipline – usually FIFO (First In First Out)

Service Facility Characteristics

1. Configuration of service facility
   - Number of servers (or channels)
   - Number of phases (or service stops)
2. Service distribution
   - The time it takes to serve 1 arrival
   - Can be fixed or random
   - Exponential distribution is often used

Exponential Distribution

μ = average service time

t = the length of service time ($t \geq 0$)

$P(t)$ = probability that service time will be greater than t

$$P(t) = e^{-\mu t}$$

Measuring Queue Performance

- ρ = utilization factor (probability of all servers being busy)
- L_q = average number in the queue
- L = average number in the *system*
- W_q = average waiting time
- W = average time in the *system*
- P_0 = probability of 0 customers in system
- P_n = probability of exactly n customers in system

Kendall's Notation

A / B / s

A = Arrival distribution

(M for Poisson, D for deterministic, and G for general)

B = Service time distribution

(M for exponential, D for deterministic, and G for general)

S = number of servers

The Queuing Models Covered Here All Assume

1. Arrivals follow the Poisson distribution
2. FIFO service
3. Single phase
4. Unlimited queue length
5. Steady state conditions

We will look at 5 of the most commonly used queuing systems.

Name (Kendall Notation)	Example
Simple system (M / M / 1)	Customer service desk in a store
Multiple server (M / M / s)	Airline ticket counter
Constant service (M / D / 1)	Automated car wash
General service (M / G / 1)	Auto repair shop
Limited population (M / M / s / ∞ / N)	An operation with only 12 machines that might break

Single Server Queuing System (M/M/1)

- Poisson arrivals
- Arrival population is unlimited
- Exponential service times
- All arrivals wait to be served
- λ is constant
- $\mu > \lambda$ (average service rate > average arrival rate)

Operating Characteristics for M/M/1 Queue

1. Average server utilization

$$\rho = \lambda / \mu$$

2. Average number of customers waiting

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

3. Average number in system

$$L = L_q + \lambda / \mu$$

4. Average waiting time

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

5. Average time in the system

$$W = W_q + 1/\mu$$

6. Probability of 0 customers in system

$$P_0 = 1 - \lambda/\mu$$

7. Probability of exactly n customers in system

$$P_n = (\lambda/\mu)^n P_0$$

Total Cost of Queuing System

$$\text{Total Cost} = C_w \times L + C_s \times s$$

C_w = cost of customer waiting time per time period

L = average number customers in system

C_s = cost of servers per time period

s = number of servers