MTM4501-Operations Research

Gökhan Göksu, PhD

Week 14

Course Content

- Definition of OR and Its History
- Decision Theory and Models
- Network Analysis
- Inventory Management Models
- Queue Models
  - Waiting Line Models
  - Queuing Theory

Queue Models

Consider the following examples:

- Customers waiting for hair cutting at a barber shop
- Customers waiting for bank service at a bank teller
- Customers waiting for bar service at a cafeteria
- Customers waiting to pay at a supermarket cash desk
- Cars waiting to pay at a highway exit cash desk
- Cars waiting at traffic lights
- Trucks waiting to load or unload at a dock
- Airplanes waiting to take off at a runway
- Items waiting to be processed by a machine
- Machines waiting to be repaired for maintenance
- Items waiting to be inspected at a quality control desk
- Jobs waiting to be executed by a computer
- Documents waiting to be signed in an office
- Bills waiting to be processed at a legislative system

Queue Models

- All above examples may be given as examples of queues (or waiting lines)
- Customers wait for a service as the service capacity is not sufficient to supply the service at once.
- The objective of queuing analysis is to offer a reasonably satisfactory service to waiting customers.

Queue Models

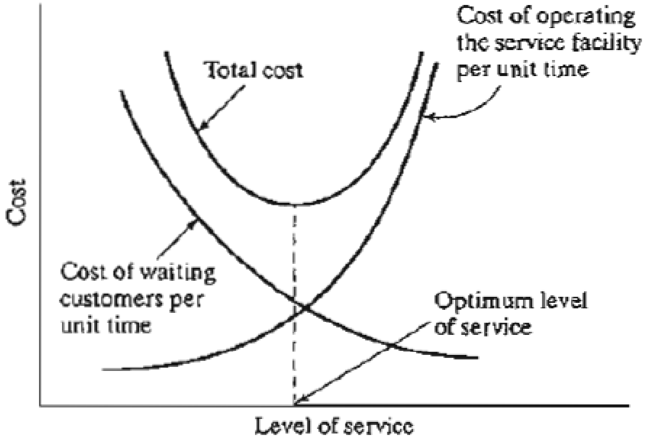


Figure: Cost-based queuing decision model

Fundamentals of Queue Models

- **Customers:** Independent entities that arrive to a service provider at random times and wait for some type of service, then leave.
- **Queue:** Customers that arrived to the server/service provider and are waiting in line for their service to start in the queue.
- **Server (Tur: Hizmet Sağlayıcı/Sunucu):** Able to serve only one customer at a time; An entity that serves customers on a first-in, first-out (FIFO) basis, with the length of service delivery time dependent on the type of service.
- **Arrival Rate (Tur: Geliş Oranı):** The average number of customers per unit time (customers have arrived with the aim of getting service). It is represented by λ. λ is assumed to be described by normal distribution.
- **Service Rate (Tur: Hizmet Oranı):** The average number of customers served per unit time. It is represented by μ.

Remark: $\mu > \lambda$: A queue is formed when customers arrive faster than they can get served.

Queue Models

- **Examples:**
  - If the Service Time is 10 minutes and a customer arrives every 15 minutes, there will be no queue at all!!!
  - If the Service Time is 15 minutes and a customer arrives every 10 minutes, the queue will extend indefinitely!!!
- **Service Discipline:** Represents the order in which customers are selected from a queue. Considering the first-come, first-served (FIFO) discipline is the most common.
- **Arrival Source:** The source where customers are generated can be either **infinite** or **finite**. A limited resource constrains the incoming customers for service (e.g., machines requesting service from a mechanic). An example of an infinite resource could be calls coming to a call center.
- **Number of Customers Waiting in the Queue (Queue Length):** The expected number of waiting customers for a service. Represented by L_q.

Queue Models

- **Number of Customers in the System:** The total of customers waiting for service and those being serviced. Represented by L_s.
- **Waiting Time in the Queue:** The total waiting time in the queue per customer. Represented by W_q.
- **Total Waiting Time in the System:** The sum of waiting time in the queue per customer and the total service time. Represented by W_s.

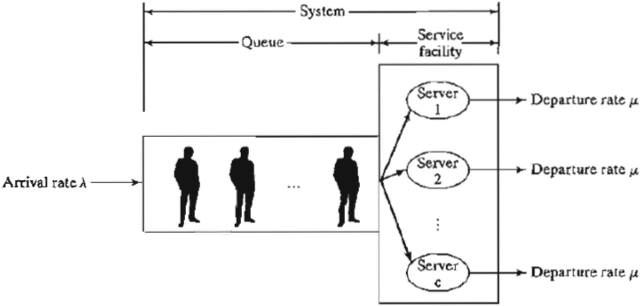
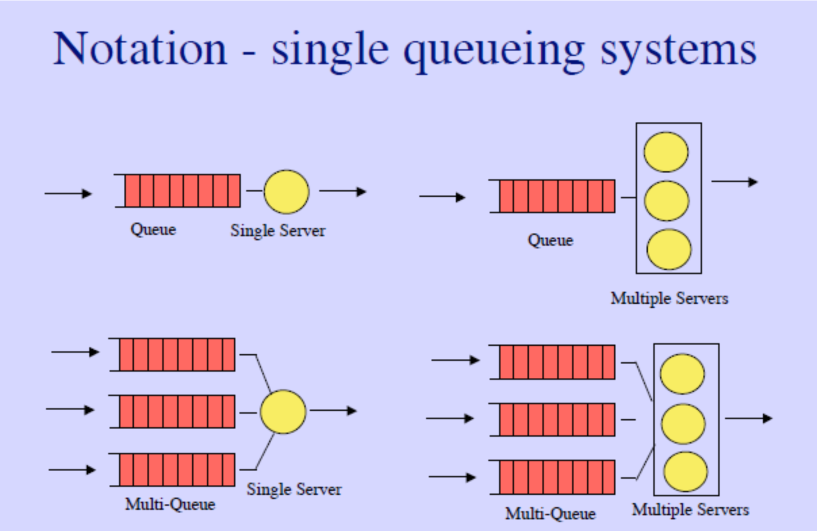
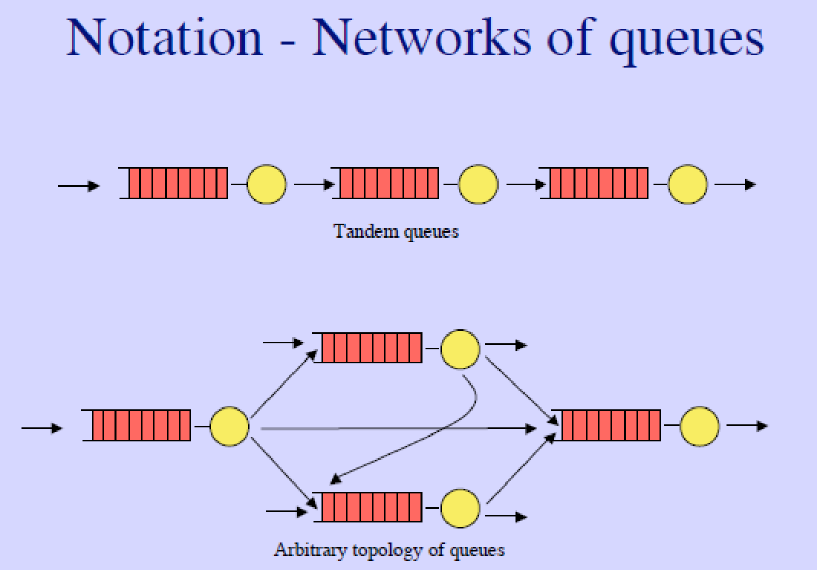


Figure: Schematic representation of a queue system with c parallel servers

Queue Models



Queue Models



Model 1: Single-Server Queue Model with Infinite Arrival Source

- P_n: Probability of having n customers in the system
- n: Number of customers in the system (in the queue and being served)

This model derives P_n as a function of λ and μ. These probabilities are then used to determine performance measures such as the average queue length, average waiting time, and the average utilization of the facility. The probabilities P_n are determined using the transition rate diagram shown below.

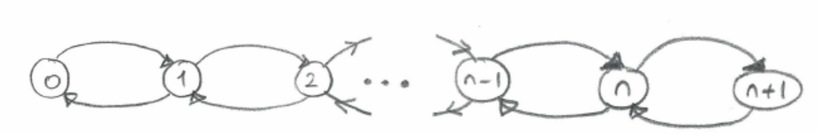


Figure: Transition rate diagram

The queue system is in state n when the number of customers in the system is n.

- λ: Arrival rate
- μ: Service rate

Model 1: Single-Server Queue Model with Infinite Arrival Source

When the system is in state n, three possible events can occur:

- When a departure occurs at a rate of μ, the system is in state $n-1$.
- When an arrival occurs at a rate of λ, the system is in state $n+1$.
- When there is no arrival or departure, the system remains in state n.

These are the last three nodes of the transition diagram. Note that state 0 can transition to state 1 only if there is an arrival at a rate of λ. Also, note that μ is undefined at state 0 since no departure can occur if the system is empty. Based on the fact that the expected flow rates entering and leaving state n must be equal, considering that state n can only transition to states $n-1$ and $n+1$, the following formula is derived:

$$\big(\text{Expected flow rate into } n \text{ state}\big) = \lambda \cdot P_{n-1} + \mu \cdot P_{n+1}$$

Similarly:

$$\big(\text{Expected flow rate out of } n \text{ state}\big) = \lambda \cdot P_n + \mu \cdot P_n$$

Model 1: Single-Server Queue Model with Infinite Arrival Source

According to these two formulas, the balance equation is written as follows:

$$\lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n, \quad n = 1, 2, \ldots$$

For $n = 0$, the balance equation is written as follows:

$$\lambda P_0 = \mu P_1, \tag{1}$$

The balance equation can be solved recursively. That is, for $n = 1$:

$$\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1, \tag{2}$$

Obtained by substituting (1) into (2):

$$P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0,$$

can be written. Similarly, for $n = 2$:

$$P_3 = \left(\frac{\lambda}{\mu}\right)^3 P_0,$$

can be obtained. This expression can be generalized as follows:

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

Model 1: Single-Server Queue Model with Infinite Arrival Source

P_0 can be determined from the fact that the sum of all probabilities is 1:

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \left[\left(\frac{\lambda}{\mu}\right)^n P_0 \right] = P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = P_0 \lim_{n\to\infty} \frac{1 - \left(\frac{\lambda}{\mu}\right)^{n+1}}{1 - \frac{\lambda}{\mu}}$$

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

Thus, the probability of the system being empty, P_0, can be calculated as follows:

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

Conversely, the probability of the system being busy is calculated as follows:

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

The probability of having n customers in the system is:

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right).$$

Model 1: Single-Server Queue Model with Infinite Arrival Source

L_s: Expected number of customers in the system

$$L_s = \mathbb{E}(n) = \sum_{n=0}^{\infty} nP_n = \sum_{n=1}^{\infty} nP_n = \sum_{n=1}^{\infty} n\left(\frac{\lambda}{\mu}\right)^n P_0 = P_0 \sum_{n=1}^{\infty} n\left(\frac{\lambda}{\mu}\right)^n$$

Here, if we make a definition for the first m sums:

$$S_m = \frac{\lambda}{\mu} + 2\left(\frac{\lambda}{\mu}\right)^2 + 3\left(\frac{\lambda}{\mu}\right)^3 + \ldots + m\left(\frac{\lambda}{\mu}\right)^m$$

$$\Longrightarrow \quad -\frac{\lambda}{\mu} S_m = -\left(\frac{\lambda}{\mu}\right)^2 - 2\left(\frac{\lambda}{\mu}\right)^3 - 3\left(\frac{\lambda}{\mu}\right)^4 - \ldots - m\left(\frac{\lambda}{\mu}\right)^{m+1}$$

By summing these two equations:

$$S_m - \frac{\lambda}{\mu} S_m = \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \ldots + \left(\frac{\lambda}{\mu}\right)^m - m\left(\frac{\lambda}{\mu}\right)^{m+1}$$

$$\underbrace{\left(1 - \frac{\lambda}{\mu}\right)}_{P_0} S_m = \frac{\lambda}{\mu} \frac{1 - \left(\frac{\lambda}{\mu}\right)^m}{1 - \frac{\lambda}{\mu}} - m\left(\frac{\lambda}{\mu}\right)^{m+1}$$

is obtained.

Model 1: Single-Server Queue Model with Infinite Arrival Source

In the limit,

$$\lim_{m\to\infty} P_0 S_m = \lim_{m\to\infty}\left[\frac{\lambda}{\mu}\frac{1-\left(\frac{\lambda}{\mu}\right)^m}{1-\frac{\lambda}{\mu}} - m\left(\frac{\lambda}{\mu}\right)^{m+1}\right] = \frac{\lambda/\mu}{1-\frac{\lambda}{\mu}} = \frac{\lambda}{\mu-\lambda} = L_s$$

L_q: Expected number of customers in the queue

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

W_s: Average time a customer spends in the system

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu-\lambda}$$

W_q: Average time a customer spends in the queue

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

The total cost per unit time is calculated as follows:

$$\left(\text{Total cost per unit time}\right) = \underbrace{\begin{pmatrix}\text{Cost per}\\ \text{service}\end{pmatrix}}_{c_1}\cdot\mu + \underbrace{\begin{pmatrix}\text{Cost per}\\ \text{waiting}\end{pmatrix}}_{c_2}\cdot L_s$$

$$= c_1\mu + c_2 L_s$$

Model 1: Single-Server Queue Model with Infinite Arrival Source

Example

In a factory, the average malfunction time of a machine is 12 minutes, and the average repair time is 8 minutes.

(a) *At any given moment, what is the number of machines that are not in production?*

(b) *How much time should pass for the broken machines to return to production?*

(c) *What is the probability of the repairman being idle (i.e. out of work)?*

(d) *For the case where the probability of malfunction increases by 20%, answer (a), (b), and (c) again.*

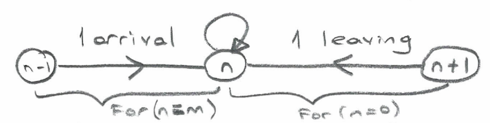
Model 2: Single Source Queue M w/AS & Finite Queue Length

The system can contain at most m customers at any given time. A single server serves a customer, and the queue length cannot exceed $m-1$.

- λ: Arrival rate
- μ: Service rate

Balance Equations:

- For $n=0$: $\lambda P_0 = \mu P_1$
- For $n = 1, 2, ..., m-1$: $\lambda P_{n-1} + \mu P_{n+1} = \lambda P_n + \mu P_n$
- For $n = m$: $\lambda P_{m-1} = \mu P_m$



- For $n=0$: $\lambda P_0 = \mu P_1 \implies P_1 = \frac{\lambda}{\mu} P_0$
- For $n=1$: $\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1 \implies P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0$
- For $n=2$: $\lambda P_1 + \mu P_3 = \lambda P_2 + \mu P_2 \implies P_3 = \left(\frac{\lambda}{\mu}\right)^3 P_0$

Model 2: Single Source Queue M w/AS & Finite Queue Length

- For $n = m$: $\implies P_m = \frac{\lambda}{\mu} P_{m-1} = \left(\frac{\lambda}{\mu}\right)^m P_0$

In this model, due to the assumption of finite queue length, the sum of probabilities for a finite number of states will be 1. Depending on the values of λ and μ, two cases arise:

- In the case of $\lambda = \mu$:

$$\sum_{n=0}^{m} P_n = 1 \implies \sum_{n=0}^{m} \left(\frac{\lambda}{\mu}\right)^n P_0 = (m+1)P_0 = 1 \implies P_0 = \frac{1}{m+1}$$

- In the case of $\lambda \neq \mu$:

$$\sum_{n=0}^{m} P_n = 1 \implies \sum_{n=0}^{m} \left(\frac{\lambda}{\mu}\right)^n P_0 = P_0 \frac{1 - \left(\frac{\lambda}{\mu}\right)^{m+1}}{1 - \left(\frac{\lambda}{\mu}\right)} = 1$$

$$\implies P_0 = \frac{1 - \left(\frac{\lambda}{\mu}\right)}{1 - \left(\frac{\lambda}{\mu}\right)^{m+1}}$$

Model 2: Single Source Queue M w/AS & Finite Queue Length

Accordingly, the probability of the system being empty can be summarized as follows:

$$P_0 = \begin{cases} \frac{1-\left(\frac{\lambda}{\mu}\right)}{1-\left(\frac{\lambda}{\mu}\right)^{m+1}} & , \lambda \neq \mu \\ \frac{1}{m+1} & , \lambda = \mu \end{cases}$$

The probability of having n customers in the system can be calculated as follows:

$$P_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1-\left(\frac{\lambda}{\mu}\right)}{1-\left(\frac{\lambda}{\mu}\right)^{m+1}} & , \lambda \neq \mu \\ \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{m+1} & , \lambda = \mu \end{cases}$$

L_s: Expected number of customers in the system

$$L_s = \mathbb{E}(n) = \sum_{n=0}^{m} nP_n = \sum_{n=1}^{m} nP_n$$

Model 2: Single Source Queue M w/AS & Finite Queue Length

Again, depending on the values of λ and μ, there are two cases for L_s:

- In the case of $\lambda = \mu$:

$$L_s = \sum_{n=1}^{m} nP_n = \sum_{n=1}^{m} n \cdot \frac{1}{m+1} = \frac{1}{m+1} \sum_{n=1}^{m} n = \frac{1}{m+1} \frac{m(m+1)}{2} = \frac{m}{2}$$

- In the case of $\lambda \neq \mu$:

$$L_s = \sum_{n=1}^{m} nP_n = \sum_{n=1}^{m} n \left(\frac{\lambda}{\mu}\right)^n P_0 = P_0 \underbrace{\sum_{n=1}^{m} n \left(\frac{\lambda}{\mu}\right)^n}_{S_m}$$

$$S_m = 1 \cdot \frac{\lambda}{\mu} + 2 \cdot \left(\frac{\lambda}{\mu}\right)^2 + 3 \cdot \left(\frac{\lambda}{\mu}\right)^3 + \ldots + m \cdot \left(\frac{\lambda}{\mu}\right)^m$$

$$-\frac{\lambda}{\mu} S_m = -\left(\frac{\lambda}{\mu}\right)^2 - 2 \cdot \left(\frac{\lambda}{\mu}\right)^3 - 3 \cdot \left(\frac{\lambda}{\mu}\right)^4 - \ldots - m \cdot \left(\frac{\lambda}{\mu}\right)^{m+1}$$

Model 2: Single Source Queue M wIAS & Finite Queue Length

The last two equations are summed side by side:

$$
\begin{aligned}
\left(1-\frac{\lambda}{\mu}\right) S_m &= \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \ldots + \left(\frac{\lambda}{\mu}\right)^m - m\left(\frac{\lambda}{\mu}\right)^{m+1} \\
&= \frac{\lambda}{\mu}\frac{1-\left(\frac{\lambda}{\mu}\right)^m}{1-\frac{\lambda}{\mu}} - m\left(\frac{\lambda}{\mu}\right)^{m+1} \\
&= \frac{\frac{\lambda}{\mu}-\left(\frac{\lambda}{\mu}\right)^{m+1}}{1-\frac{\lambda}{\mu}} - (m+1)\left(\frac{\lambda}{\mu}\right)^{m+1} + \left(\frac{\lambda}{\mu}\right)^{m+1} \\
&= \frac{\frac{\lambda}{\mu}-\left(\frac{\lambda}{\mu}\right)^{m+1}+\left(\frac{\lambda}{\mu}\right)^{m+1}-\left(\frac{\lambda}{\mu}\right)^{m+2}}{1-\frac{\lambda}{\mu}} - (m+1)\left(\frac{\lambda}{\mu}\right)^{m+1} \\
&= \frac{\frac{\lambda}{\mu}\left(1-\left(\frac{\lambda}{\mu}\right)^{m+1}\right)}{1-\frac{\lambda}{\mu}} - (m+1)\left(\frac{\lambda}{\mu}\right)^{m+1}
\end{aligned}
$$

Model 2: Single Source Queue M wIAS & Finite Queue Length

If the expression is rearranged:

$$S_m = \frac{\frac{\lambda}{\mu}\left(1-\left(\frac{\lambda}{\mu}\right)^{m+1}\right)}{\left(1-\frac{\lambda}{\mu}\right)^2} - \frac{(m+1)\left(\frac{\lambda}{\mu}\right)^{m+1}}{1-\frac{\lambda}{\mu}}$$

Substituting this sum into $L_s = P_0 S_m$:

$$\begin{aligned} L_s &= \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{m+1}} \frac{\frac{\lambda}{\mu}\left(1-\left(\frac{\lambda}{\mu}\right)^{m+1}\right)}{\left(1-\frac{\lambda}{\mu}\right)^2} - \frac{(m+1)\left(\frac{\lambda}{\mu}\right)^{m+1}}{1-\frac{\lambda}{\mu}} \\ &= \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} - \frac{(m+1)\left(\frac{\lambda}{\mu}\right)^{m+1}}{1-\left(\frac{\lambda}{\mu}\right)^{m+1}} \end{aligned}$$

Accordingly, the expected number of customers in the system can be summarized as follows:

$$L_s = \begin{cases} \frac{\lambda}{\mu-\lambda} - (m+1)\frac{\left(\frac{\lambda}{\mu}\right)^{m+1}}{1-\left(\frac{\lambda}{\mu}\right)^{m+1}} & ,\lambda \neq \mu \\ \frac{m}{2} & ,\lambda = \mu \end{cases}$$

Model 2: Single Source Queue M wIAS & Finite Queue Length

- P_m: Probability of the system being busy

$$P_m = 1 - P_0$$

- L_q: Expected number of customers in the queue

$$L_q = L_s - P_m = L_s - (1 - P_0)$$

- λ_e: Effective arrival rate

$$\lambda_e = \lambda(1 - P_m)$$

- W_s: Average waiting time in the system

$$W_s = \frac{L_s}{\lambda_e} = \frac{L_s}{\lambda(1 - P_m)}$$

- W_q: Average waiting time in the queue

$$W_q = \frac{L_q}{\lambda_e} = \frac{L_q}{\lambda(1 - P_m)}$$

Model 3: Infinite Arrival Rate Multi-Server Queue Model

In this model, it is assumed that there are s parallel servers, and each parallel server is identical.

- λ: Arrival rate
- μ: Service rate of each server
- s: Number of parallel servers
- n: Number of customers in the system

The effect of using parallel servers is a proportionate increase in the facility service rate:

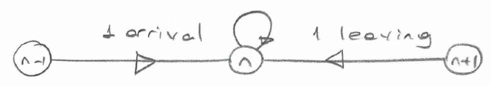
- $n \leq s \implies$ No queue forms
- $n > s \implies$ s customers are in service, and $(n - s)$ customers are waiting in the queue.

In the previous models, it was assumed that $\lambda < \mu$. In this multi-server model, due to s parallel servers, it is assumed that $\lambda < \mu \cdot s$. Here, the product $\mu \cdot s$ can be interpreted as the service capacity.

Model 3: Infinite Arrival Rate Multi-Server Queue Model

Balance Equations: For $0 \leq n < s$;

- For $n = 0, 1, 2, ..., s, \implies \lambda P_{n-1} + (n+1)\mu P_{n+1} = \lambda P_n + n\mu P_n$



- For $n = 0 \implies \lambda P_0 = \mu P_1 \implies P_1 = \frac{\lambda}{\mu} P_0$
- For $n = 1 \implies \lambda P_0 + 2\mu P_2 = \lambda P_1 + \mu P_1 \implies P_2 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 P_0$
- For $n = 2 \implies \lambda P_1 + 3\mu P_3 = \lambda P_2 + 2\mu P_2 \implies P_3 = \frac{1}{3 \cdot 2}\left(\frac{\lambda}{\mu}\right)^3 P_0$
- $\vdots$
- For $n = s \implies P_s = \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s P_0$

Model 3: Infinite Arrival Rate Multi-Server Queue Model

Balance Equations: For $s \leq n$;

- For $n = s \implies \lambda P_{s-1} + s\mu P_{s+1} = \lambda P_s + s\mu P_s \implies P_{s+1} = \frac{\lambda}{s\mu} P_s$
- For $n = s+1 \implies \lambda P_s + s\mu P_{s+2} = \lambda P_{s+1} + s\mu P_{s+1} \implies P_{s+2} = \left(\frac{\lambda}{s\mu}\right)^2 P_s$
- $\vdots$
- For $n = s + k \implies P_{s+k} = \left(\frac{\lambda}{s\mu}\right)^k P_s$

By writing $k = n - s$ in the last expression, the probability of having n customers in the system for $s \leq n$ is obtained:

$$P_n = \left(\frac{\lambda}{s\mu}\right)^{n-s} P_s = \left(\frac{\lambda}{s\mu}\right)^{n-s} \frac{1}{s!} \left(\frac{\lambda}{\mu}\right)^s P_0 = \frac{1}{s!} \frac{1}{s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n P_0$$

For all cases, the probability of having n customers in the system can be summarized as follows:

$$P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n P_0 & , 0 \leq n < s \\ \frac{1}{s!} \frac{1}{s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n P_0 & , s \leq n \end{cases}$$

Model 3: Infinite Arrival Rate Multi-Server Queue Model

Considering the sum of all probabilities:

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{s-1} P_n + \sum_{n=s}^{\infty} P_n = \sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 + \underbrace{\sum_{n=s}^{\infty} \frac{1}{s!}\frac{1}{s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n P_0}_{T}$$

T can be calculated as follows:

$$T = \sum_{n=s}^{\infty} \frac{1}{s!}\frac{1}{s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n P_0 = \sum_{n=s}^{\infty} \frac{1}{s!}\frac{1}{s^{-s}}\left(\frac{\lambda}{s\mu}\right)^n P_0 = \frac{P_0}{s!s^{-s}}\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^n$$

Under the assumption $\lambda < \mu \cdot s$;

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

Substituting this into the sum of probabilities;

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

Model 3: Infinite Arrival Rate Multi-Server Queue Model

Therefore, the probability of the system being empty is calculated as follows:

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

L_q: Expected number of customers in the queue

$$L_q = \mathbb{E}(n-s) = \sum_{n=s}^{\infty} (n-s) P_n = \frac{P_0}{s! s^{-s}} \sum_{n=s}^{\infty} (n-s) \left(\frac{\lambda}{s\mu} \right)^n = \frac{\left(\frac{\lambda}{\mu} \right)^s \frac{\lambda}{s\mu}}{s! \left(1 - \frac{\lambda}{s\mu} \right)^2} P_0$$

L_s: Expected number of customers in the system

$$L_s = L_q + s \frac{\lambda}{s\mu} = L_q + \frac{\lambda}{\mu}$$

Model 3: Infinite Arrival Rate Multi-Server Queue Model

W_s: Average waiting time in the system

$$W_s = \frac{L_s}{\lambda}$$

W_q: Average waiting time in the queue

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

Probability of waiting for service

$$\mathbb{P}(n \geq s) = \sum_{n=s}^{\infty} P_n = \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!\left[1 - \frac{\lambda}{s\mu}\right]} P_0.$$

Model 3: Infinite Arrival Rate Multi-Server Queue Model

Example

There are 3 service desks at a post office. Approximately 192 customers arrive every day. Each business day consists of 8 hours. The average service time for each customer is 5 minutes. Therefore;

a) *What is the probability of having no customers in the post office?*
b) *What is the probability of at least one service desk being busy?*
c) *What is the probability of waiting for service?*
d) *What is the expected number of customers in the queue?*
e) *What is the expected number of customers in the system?*
f) *What is the average waiting time for each customer in the queue?*