Appendix B

Queue Networks

Introduction

Queue network can be regarded as a group of 'k' inter-connected nodes, where each node represents a service facility of some kind with s_i servers at node i. ($s_i \geq 1$)

Series Queues

A *series queue* model or a *tandem queue* model is, in general, one in which (i) customers may arrive from outside the system at any node and may leave the system from any node.

(ii) customers may enter the system at some node, traverse from node to node in the system and leave the system from some node, not necessarily following the same order of nodes and (iii) customers may return to the nodes previously visited, skip some nodes and even choose to remain in the system for ever.

In the following sections, we shall discuss a few kinds of series queues.

Series queues with blocking

This is a sequential queue model consisting of two service points S_1 and S_2, at each of which there is only one server and where no queue is allowed to form at either point.

Note: It is a misnomer by which we called the model as a series queue model!

An entering customer will first go to S_1. After he gets the serivce completed at S_1, he will go to S_2 if it is empty or will wait in S_1 until S_2 becomes empty [viz; the system is blocked for a new customer.] This means that a potential customer will enter the system only when S_1 is empty, irrespective of whether S_2 is empty or not, since the model is a sequential model, viz., all the customers require service at S_1 and then at S_2.

Let us now proceed to find the steady-state probabilities $P(m, n)$ that there is m customer ($m = 0$ or 1) in S_1 and n customer ($n = 0$ or 1) in S_2. Any state of the model will be denoted by (m, n). The possible states of the system are given below with their interpretation:

State	*Interpretation*
(0, 0)	No customer in either service point
(1, 0)	Only one customer in S_1.
(0 1.)	Only one customers in S_2.
(1, 1)	One customer each in S_1 and S_2.
$(b, 1)$	One customer each in S_1 and S_2, but the customer in S_1, having finished his work at S_1, is waiting for S_2 to become free, while the customer in S_2 is being served.

We assume that potential customers arrive in accordance with a Poisson process with parameter λ and the service times at S_1 and S_2 follow exponential distributions with parameters μ_1 and μ_2 respectively. To get the values of $P(m, n)$, we shall first write down the steady-state balance equations using a state transition diagram, as given below, which consists of a small circle for each state and directed lines labeled by the rates at which the process goes from one state to another.

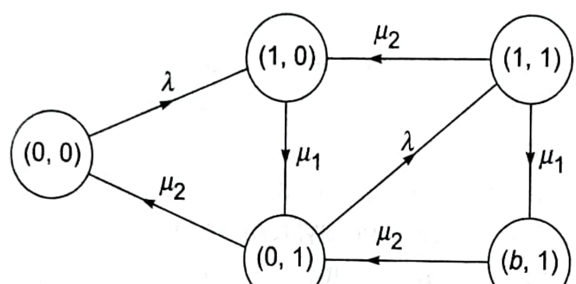

Fig. 1 State Transition Diagram

Note: The arrow from state (0, 0) to state (1, 0) is labeled λ. This means that when the system (consisting of S_1 and S_2) is empty, it goes to state (1, 0) through an arrival which enters the system at a rate λ. Similar explanation holds good for the arrow from (0, 1) to (1, 1). The transition from (1, 0) to (0, 1) takes place, when the customer finsihes his job at S_1 at the rate μ_1. Similar explanation holds good for the arrow from (1, 1) to $(b, 1)$. When in state $(b, 1)$, the process will go to state (0, 1), when the customer at S_2 completes his service, that occurs at the rate μ_2. Similar explanations hold good for the transitions from (1, 1) to (1, 0) and from (0, 1) to (0, 0).

Now the balance equation corresponding to any state (m, n) is obtained by equating the rate at which the process leaves that state and the rate at which the process enters that state. The rate at which the process leaves the state (m, n) $= P(m, n) \times$ sum of the labels of the arrows that leave the state (m, n). The rate at which the process enters the state (m, n) = sum of the labels of the arrows that enter (m, n) multiplied by the relevant probabilities of the states from which they emanate. Accordingly we get the following balance equations for the 5 states of the system:

State	*Balance equation*	
(0, 0)	$\lambda P(0, 0) = \mu_2 P(0, 1)$	(1)
(1, 0)	$\mu_1 P(1, 0) = \lambda P(0, 0) + \mu_2 P(1, 1)$	(2)
(0, 1)	$(\lambda + \mu_2) P(0, 1) = \mu_1 P(1, 0) + \mu_2 P(b, 1)$	(3)
(1, 1)	$(\mu_1 + \mu_2) P(1, 1) = \lambda P(0, 1)$	(4)
(*b*, 1)	$\mu_2 P(b, 1) = \mu_1 P(1, 1)$	(5)

Since the process has to be in any one of the 5 mutually exclusive and exhaustive states, we have

$$P(0, 0) + P(1, 0) + P(0, 1) + P(1, 1) + P(b, 1) = 1 \qquad (6)$$

Solving the above six equations, we can get the five steady-sate probabilities.

Two stage (service Point) Series queues (Two stage tandem queues)

Let us consider a two (service) stage queuing system in which customers arrive from outside at a Poisson rate λ to S_1. After being served at S_1, they then join the queue in front of S_2. After receiving service at S_2, they leave the system. It is assumed that there is infinite waiting (queueing space) at each service point. Each server serves one customer at a time and the service times at S_1 and S_2 follow exponential distributions with parameters μ_1 and μ_2 respectively.

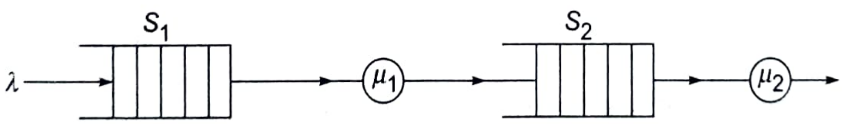

Fig. 2

To find the steady-state joint probability $P(m, n)$ of m customers in S_1 and n customers in S_2, where $m \geq 0$ and $n \geq 0$, we shall write down the balance equations using the state transition diagrams given below. By taking (m, n) as the central state, we have shown only the arrows entering into and leaving from it.

Note: The state transition diagram [Fig. (4)] for the central state (0, 0) is extracted from Fig. (3) by retaining only those states for which $m \geq 0$ and $n \geq 0$ and the connecting arrows. Similarly the state transition diagrams [Fig. (5) and Fig. (6)] for the central states (m, 0) and (n, 0) are obtained.

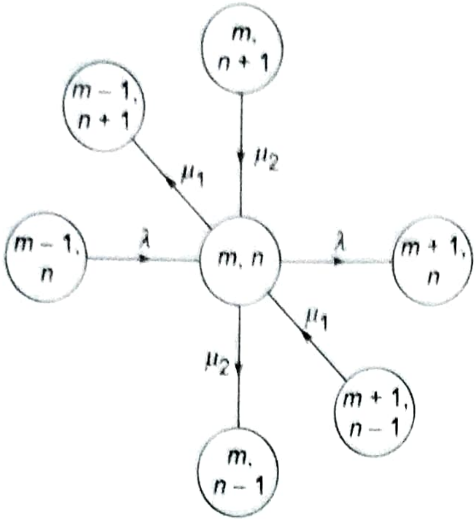

Fig. 3

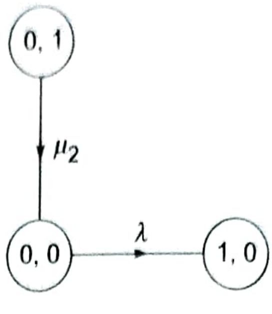

Fig. 4

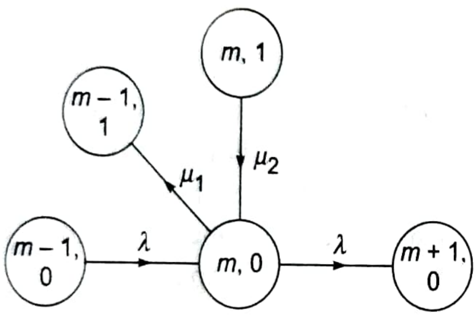

Fig. 5

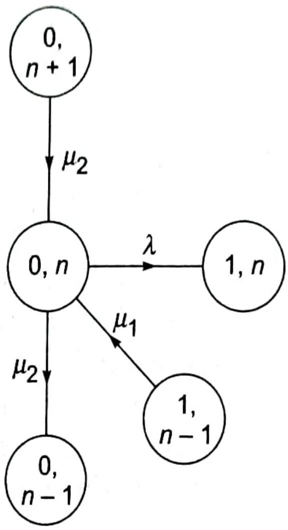

Fig. 6

The balance equations for this model are obtained as given below:

$$\lambda P(0, 0) = \mu_2 P(0, 1) \tag{1}$$

$$(\lambda + \mu_1) P(m, 0) = \lambda P(m-1, 0) + \mu_2 P(m, 1) \; [m > 0] \tag{2}$$

$$(\lambda + \mu_2) P(0, n) = \mu_1 P(1, n-1) + \mu_2 P(0, n+1) \; [n > 0] \tag{3}$$

$$(\lambda + \mu_1 + \mu_2) P(m, n) = \lambda P(m-1, n) + \mu_1 P(m+1, n-1) + \mu_2 P(m, n+1) \tag{4}$$

Also
$$\sum_m \sum_n P(m, n) = 1 \tag{5}$$

We shall not attempt to solve the above balance equations directly, but shall guess the solution by using *Burke's theorem* which is stated below without proof:

For an *M*/*M*/*s* queueing system, the output (departure) process is also Poisson with the same rate λ as the input (arrival) process in the steady-state, where $s \geq 1$.

This means that the arrival process to S_2, which is the same as the departure process from S_1, is also Poisson with parameter λ. Thus the queueing systems at both S_1 and S_2 are $M/M/1$ models (if we assume that there is only one server at S_1 and at S_2).

From the discussion on the characteristics of $M/M/1$ model

$$P\ (m \text{ customers at } S_1 \text{ system}) = \left(\frac{\lambda}{\mu_1}\right)^m \left(1-\frac{\lambda}{\mu_1}\right) \text{ and}$$

$$P\ (n \text{ customers at } S_2 \text{ system}) = \left(\frac{\lambda}{\mu_2}\right)^n \left(1-\frac{\lambda}{\mu_2}\right)$$

Since the number of customers at S_1 and S_2 are independent random variables, the joint probability of m customers at S_1 and n customers at S_2 is given by

$$P\ (m, n) = \left(\frac{\lambda}{\mu_1}\right)^m \cdot \left(1-\frac{\lambda}{\mu_1}\right)^m \cdot \left(\frac{\lambda}{\mu_2}\right)^n \cdot \left(1-\frac{\lambda}{\mu_2}\right)^n, \text{ for } m \geq 0 \text{ and } n \geq 0 \qquad \ldots(6)$$

Note: It is easily verified that (6) satisfies the balance equations (1) to (5). For example, on using (6) in (4),

$$R.\ S. \text{ of } (4) = \left(1-\frac{\lambda}{\mu_1}\right)\left(1-\frac{\lambda}{\mu_2}\right)$$

$$\left[\lambda \cdot \left(\frac{\lambda}{\mu_1}\right)^{m-1} \cdot \left(\frac{\lambda}{\mu_2}\right)^{n} + \mu_1\left(\frac{\lambda}{\mu_1}\right)^{m+1} \cdot \left(\frac{\lambda}{\mu_2}\right)^{n-1} + \mu_2\left(\frac{\lambda}{\mu_1}\right)^{m} \cdot \left(\frac{\lambda}{\mu_2}\right)^{n+1}\right]$$

$$= \left(1-\frac{\lambda}{\mu_1}\right)\left(1-\frac{\lambda}{\mu_2}\right)\left(\frac{\lambda}{\mu_1}\right)^{m-1}\left(\frac{\lambda}{\mu_2}\right)^{n-1}\left[\frac{\lambda^2}{\mu_2}+\frac{\lambda^2}{\mu_1}+\frac{\lambda^3}{\mu_1\mu_2}\right]$$

$$= \left(1-\frac{\lambda}{\mu_1}\right)\left(1-\frac{\lambda}{\mu_2}\right)\left(\frac{\lambda}{\mu_1}\right)^{m-1}\left(\frac{\lambda}{\mu_2}\right)^{n-1}\frac{\lambda^2}{\mu_1\mu_2}(\lambda + \mu_1 + \mu_2)$$

$$= (\lambda + \mu_1 + \mu_2)\left[\left(\frac{\lambda}{\mu_1}\right)^m\left(1-\frac{\lambda}{\mu_1}\right)\right]\left[\left(\frac{\lambda}{\mu_2}\right)^n\left(1-\frac{\lambda}{\mu_2}\right)\right]$$

$$= (\lambda + \mu_1 + \mu_2)\ P\ (m, n)$$
$$= L.\ S. \text{ of } (4)]$$

The average number of customers in the system is given

$$L_S = \sum_m \sum_n (m + n)\ P\ (m, n)$$

$$= \sum_{m=0}^{\infty} m \left(\frac{\lambda}{\mu_1}\right)^m \left(1-\frac{\lambda}{\mu_1}\right) + \sum_{n=0}^{\infty} n \left(\frac{\lambda}{\mu_2}\right)^n \left(1-\frac{\lambda}{\mu_2}\right)$$

$$= \left(1-\frac{\lambda}{\mu_1}\right)\cdot\frac{\lambda}{\mu_1}\cdot\left(1-\frac{\lambda}{\mu_1}\right)^{-2} + \left(1-\frac{\lambda}{\mu_2}\right)\cdot\frac{\lambda}{\mu_2}\cdot\left(1-\frac{\lambda}{\mu_2}\right)^{-2}$$

$$= \frac{\lambda}{\mu_1-\lambda} + \frac{\lambda}{\mu_2-\lambda} = L_{S1} + L_{S2}$$

The average waiting time of a customer in the system is given by

$$E(W_S) = \frac{1}{\lambda} L_S, \text{ by Little's formula.}$$

$$= \frac{1}{\mu_1-\lambda} + \frac{1}{\mu_2-\lambda} = E(W_{S1}) + E(W_{S2})$$

Note (1): As the arrival rate λ in a 2-state tandem queue model increases, the node with the larger value of $\rho_i = \frac{\lambda}{\mu_i}$ will introduce instability. Hence the node with the larger value of ρ_i is called the *bottleneck* of the system.

Note (2): The discussion given above holds good, even when there are many servers at each stage, provided the service time of each server at stage i (i = 1, 2) is exponential with mean $\frac{1}{\mu_i}$.

In this case, it is to be noted that the queueing system at each stage is an *M/M/s* model and hence the relevant results corresponding to this model should be used.

Open Jackson Networks

A network of k service facilities or nodes is called an *open Jackson network*,if it satisfies the following characteristics:

(1) Arrivals, from outside, to the node i follow a Poisson process with mean rate r_i and join the queue at i and wait for his turn for service.

(2) Service times at the channels at node i are independent and each exponentially distributed with parameter μ_i.

(3) Once a customer gets the service completed at node i, he joins the queue at node j with probability P_{ij} (whatever be the number of customers waiting at j for service), where i = 1, 2,..., k and j = 0, 1, 2, ..., k. P_{io} represents the probability that a customer leaves the system from node i after getting the service at i.

Note: Derivations and the methods of solving the balance equations for this model are beyond the scope of the book. However we just give below Jackson's solution for the balance equations of this model

If we denote the total arrival rate of customers to server j [viz., the sum of the arrival rate r_j (Note: It is not λ_j) to j coming from outside and the rates of departure λ_i from the servers i] by λ_j., then

$$\lambda_j = r_j + \sum_{i=1}^{k} \lambda_i P_{ij};\ j = 1, 2, \ldots, k \qquad \ldots(1)$$

P_{ij} is the probability that a departure from server i joins the queue at server j and hence $\lambda_i P_{ij}$ is the rate of arrival to server j from among those coming out from server i.

Equations (1) are called Traffic equations or Flow balance equations.

Jackson has proved that the steady-state solutions of these traffic equations with single server at each node is

$$P(n_1, n_2, \ldots, n_k) = \rho_1^{n_1}(1-\rho_1)\cdot \rho_2^{n_2}(1-\rho_2)\ldots\rho_k^{n_k}(1-\rho_k) \qquad \ldots(2)$$

where $\rho_j = \dfrac{\lambda_j}{\mu_j}$, provided $\rho_j < 1$ for all j.

Since $P(n_1, n_2, \ldots, n_k) = \left[\rho_1^{n_1}(1-\rho_1)\right]\left[\rho_2^{n_2}(1-\rho_2)\right]\ldots\left[\rho_k^{n_k}(1-\rho_k)\right]$

viz., the joint probability is equal to the product of the marginal probabilities, we can interpret that the network acts as if the queue at each node i is an independent $M/M/1$ queue with rates λ_i and μ_i.

Note (1): Eventhough (2) may be misinterpreted such the the arrival process at node i is Poisson in this network model, it need not be so, as the customer may visit a server more than once (known as feedback situation)

Thus whether the arrival processes at various nodes are Poisson or not, step (2) alone holds good and hence the network behaves as if its nodes were independent $M/M/1$ queue models.

Note(2): $E(N_S) = \sum_{i=1}^{k} E\{N_{si}\}$, but $E(W_S) = \dfrac{1}{\sum_{i=1}^{k} r_i} E(N_s)$

Note (3): Jackson's open network concept can be extended when the nodes are multi server nodes. In this case the network behaves as if each node is an independent $M/M/s$ model.

Closed Jackson Networks

A queueing network of k nodes is called a *closed Jackson network*, if new customers never enter into and the existing customers never depart from the system. viz., if $r_i = 0$ and $P_{i0} = 0$ for all i. In other words, it is equivalent to a finite source queueing system of N customers who traverse continuously inside the network where the service time of server i is exponentially distributed with rate μ_i; $i = 1, 2, \ldots, k$.

When a customer completes service at S_i, he then joins the queue at S_j, $j = 1, 2, \ldots, k$ with probability P_{ij} where it is assumed that $\sum_{j=1}^{k} P_{ij} = 1$ for all $i = 1, 2, \ldots k$. We note that the matrix $P = [P_{ij}]$ is similar to one-step transition probability matrix of a Markov chain, that is stochastic and irreducible.

The flow balance equations of this model become

$$\lambda_j = \sum_{i=1}^{k} \lambda_i P_{ij} \ ; \ j = 1, 2, \ldots, k \ [\because r_j = 0] \ldots \quad (1)$$

The martix $[P_{ij}]$ is called the *routing probability martrix* in this context. Jackson has proved that the steady-state solution of equation (1) is

$$P(n_1, n_2, \ldots, n_k) = C_N \ \rho_1^{n_1} \rho_2^{n_2} \ldots \rho_k^{n_k}, \text{ where}$$

$$C_N^{-1} = \sum_{n_1 + n_2 + \ldots + n_k = N} \rho_1^{n_1} \rho_2^{n_2} \ldots \rho_k^{n_k}, \text{ where } \rho_j = \frac{\lambda_j}{\mu_j}$$

Note: *N Customers must be allocated among k nodes, such that $n_1 + n_2 + \ldots + n_k = N$. This allocation can be done in $\binom{N+k-1}{N}$ ways.*

If there are s_i servers at node i, the solution is given by

$$P(n_1, n_2, \ldots, n_k) = C_N \frac{\rho_1^{n_1}}{a_1(n_1)} \cdot \frac{\rho_2^{n_2}}{a_2(n_2)} \ldots, \frac{\rho_k^{n_k}}{a_k(n_k)},$$

where $$C_N^{-1} = \sum_{n_1+n_2+\ldots+n_k = N} \frac{\rho_1^{n_1}}{a_1(n_1)} \frac{\rho_2^{n_2}}{a_2(n_2)} \cdots \frac{\rho_k^{n_k}}{a_k(n_k)},$$

where $$a_i(n_i) = \begin{cases} \lfloor n_i, & \text{if } n_i < s_i \\ s_i^{n_i - s_i} \lfloor s_i, & \text{if } n_i \geq s_i. \end{cases}$$