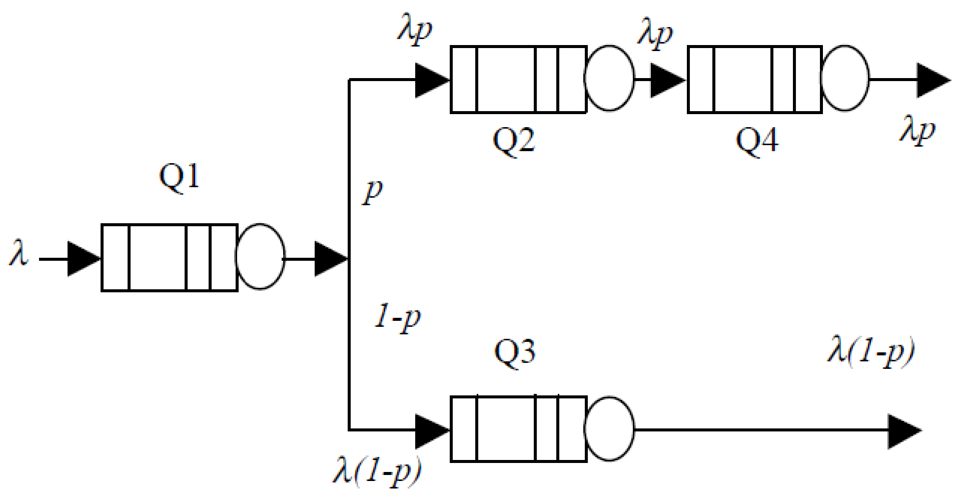
Basic Queueing Theory

Departures, Method of Stages, Batch Arrivals

Lecture 4



Knowledge of the nature of the departure process from a queue would be useful as we can then use it to analyze simple cases of queueing networks as shown.

The key result here is that *the departure process from a M/M/m/∞ queue is also Poisson* with the same rate as the arrival rate entering the queue.

It should also be noted that the result of randomly splitting or combining independent Poisson processes also yields a Poisson process

The result on the departure process of a M/M/m/∞ queue follows from *Burke's Theorem*. This theorem states that -

[A] The departure process from a *M/M/m/∞* queue is Poisson in nature.

[B] For a *M/M/m/∞* queue, at each time *t*, the number of customers in the system is independent of the sequence of departure times prior to *t*.

[C] For a *M/M/m/∞* FCFS queue, given a customer departure at time *t*, the arrival time of this customer is independent of the departure process prior to *t*.

Time Reversibility Property of Irreducible, Aperiodic Markov Chains

Consider a discrete time, irreducible, aperiodic Markov Chain X_1, X_2, ……, X_{n-1}, X_n, X_{n+1}, …….. for which the transition probabilities are given to be $\{p_{ij}\}$.

Now consider the same chain backwards in time, i.e. the chain ……X_{n+1}, X_n, ……, X_3, X_2, X_1. This would also be a Markov Chain since we can show that

$$P\{X_m = j \mid X_{m+1} = i, X_{m+2} = i_2, \ldots\ldots, X_{m+k} = i_k\}$$

$$= \frac{P\{X_m = j, X_{m+1} = i, X_{m+2} = i_2, \ldots\ldots, X_{m+k} = i_k\}}{P\{X_{m+1} = i, X_{m+2} = i_2, \ldots\ldots, X_{m+k} = i_k\}}$$

$$= \frac{P\{X_m = j, X_{m+1} = i\}P\{X_{m+2} = i_2, \ldots\ldots, X_{m+k} = i_k \mid X_m = j, X_{m+1} = i\}}{P\{X_{m+1} = i\}P\{X_{m+2} = i_2, \ldots\ldots, X_{m+k} = i_k \mid X_{m+1} = i\}}$$

$$= \frac{p_j p_{ji}}{p_i} = p_{ij}^*$$

State Transition Probability of the Reverse Chain

- The Markov Chain is considered to be time reversible for the special case where

$$p_{ij}^{*} = p_{ij} \ \forall\, i, j.$$

- The reverse chain will have the following properties –
- The reversed chain is also irreducible and aperiodic like the forward chain
- The reversed chain has the same stationary state distribution as the forward chain
- The chain is time reversible only if the detailed balance equation $p_i p_{ij} = p_j p_{ji}$ holds for $\forall\ i, j \geq 0$

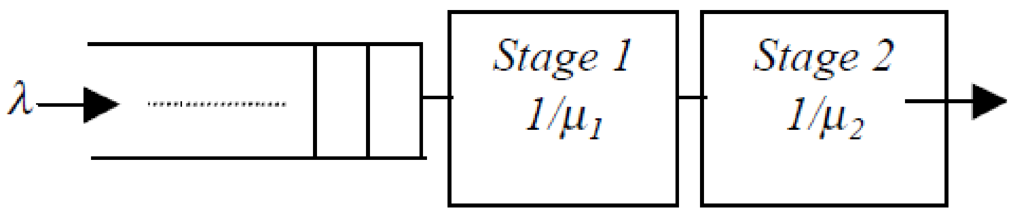
How can we handle queues where the service time distribution is not exponential?

[A] If we can express the actual service time as combinations of exponentially distributed time intervals, then the *Method of Stages* may be used. (Section 2.9)

[B] The M/G/1 queue and its variations may be analyzed. (Chapters 3 and 4)

[C] Approximation methods may be used if the mean and variance of the service time are given. (GI/G/m approximation of Section 6.2)

- Method of Stages

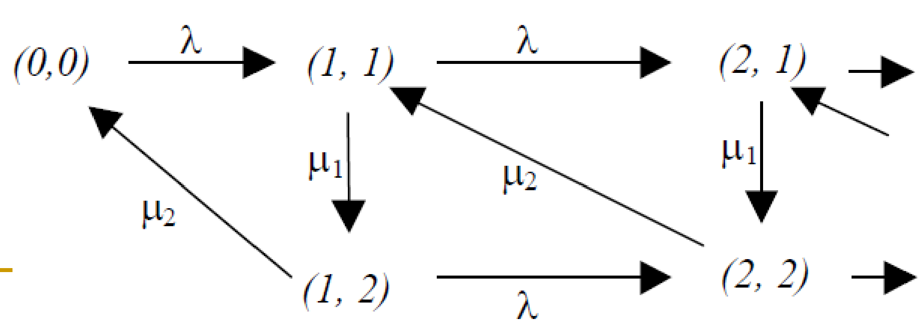


Consider a M/-/1/∞ example where the actual service time is the sum of two random variables, each of which is exponentially distributed.

State of the system represented as *(n, j)* where *n* is the total number of customers in the system where the customer currently being served is at

Stage *j*, $n=0,1,\ldots\ldots,\infty$, $j=1, 2$

State *(0,0)* represents the state when the system is empty



State Transition Diagram of the System

Balance Equations for the System

$$
\left\{
\begin{aligned}
&\lambda p_{00} = \mu_2 p_{12} \\
&(\lambda + \mu_1) p_{11} = \lambda p_{00} + \mu_2 p_{22} \\
&(\lambda + \mu_2) p_{12} = \mu_1 p_{11} \\
&(\lambda + \mu_1) p_{21} = \lambda p_{11} + \mu_2 p_{32} \\
&(\lambda + \mu_2) p_{22} = \lambda p_{12} + \mu_1 p_{21} \\
&etc.......
\end{aligned}
\right. \qquad (2.38)
$$

These Balance Equations may be solved along with the appropriate *Normalization Condition* to obtain the state probabilities of the system.

Once these are known, performance parameters of the queue maybe appropriately evaluated.

The method illustrated for the M/-/1/∞ example may be extended for the following types systems.

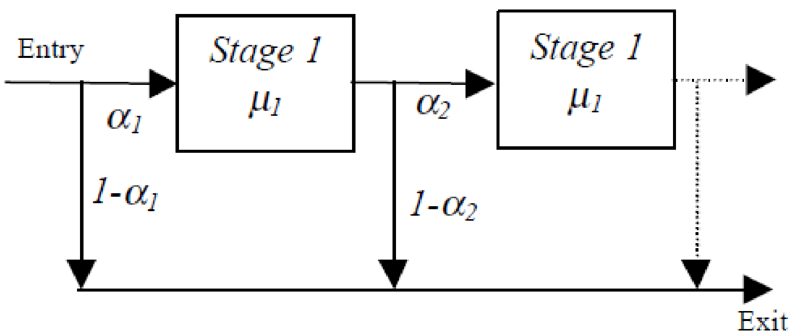
1. *Have k stages of service times* - more rows in the state transition diagram

2. *Finite Number of Waiting Positions in the Queue* - make the arrival rate a function of the number in the system and make it go to zero once all the waiting positions have been filled

3. *Multiple Servers* - approximate this by allowing more than one job to enter service at a time

4. *More General Service Time Distributions* - see next slide

For more general service time distributions, the *Method of Stages* may be used if the *Laplace Transform* of the pdf of the service time may be represented as a rational function of *s*, $L_B(s)=N(s)/D(s)$, with simple roots.



With multiple stages like this, the *L.T.* of the service time pdf will be of the form -

$$L_B(s) = (1-\alpha_1) + \sum_j \alpha_1 \ldots\ldots \alpha_{j-1}(1-\alpha_j)\prod_{i=1}^{j} \frac{\mu_i}{s+\mu_i}$$

This leads to -

$$L_B(s) = \beta_0 + \sum_i \frac{\beta_i}{s+\mu_i}$$

Given a service time pdf as $L_B(s)=N(s)/D(s)$ with simple roots –

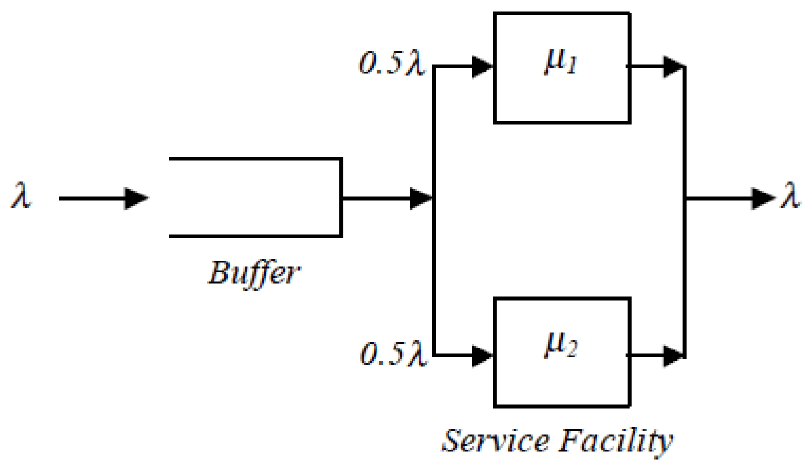
1. Obtain the multiple stage representation in the form shown earlier

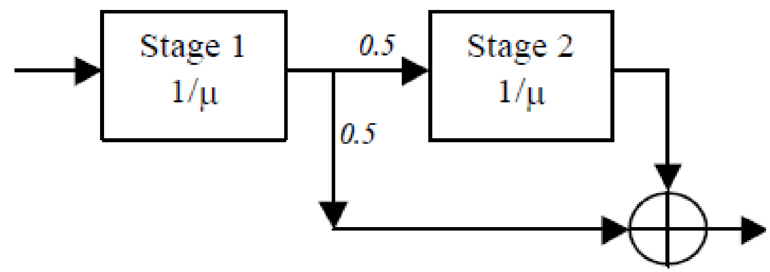
2. Draw the corresponding state transition diagram and identify the flows between the various states

3. Write and solve the flow balance equations along with the normalization condition to obtain the state probabilities

4. Use the state probabilities to obtain the required performance parameters

Some Examples of Service Time Distributions





$$L_B(s) = \frac{0.5s(\mu_1 + \mu_2) + \mu_1\mu_2}{(s+\mu_1)(s+\mu_2)}$$

$$L_B(s) = \frac{0.5\mu}{s+\mu} + \frac{0.5\mu^2}{(s+\mu)^2}$$

Queues with Bulk (or Batch) Arrivals (Section 2.10)

$M^{[X]}$ Poisson Batch Arrival Process

Batches arriving as a Poisson process with exponentially distributed inter-arrival times between batches
Batch size = Number of jobs in a batch (random variable)

λ = Average Batch Arrival Rate

$\beta_r = P\{r \text{ jobs in a batch}\} \quad r = 1, 2, \ldots.$

$$\beta(z) = \sum_{r=1}^{\infty} \beta_r z^r$$

$$\overline{\beta} = \sum_{r=1}^{\infty} r\beta_r$$

The M[X]/M/1 Queue

$$\left.\begin{aligned}
&\lambda p_0 = \mu p_1 && for \quad k=0 \\
&(\lambda+\mu)p_k = \mu p_{k+1} + \sum_{i=0}^{k-1} \lambda \beta_{k-i} p_i && for \quad k \geq 1
\end{aligned}\right\} \text{Balance Equations}$$

Though these may be solved in the standard fashion, we will consider a solution approach for directly obtaining *P(z),* the Generating Function for the number in the system. For this, we would need to multiply the k^{th} equation above by z^k and sum from $k=1$ to $k=\propto$.

$$P(z) = \sum_{n=0}^{\infty} p_n z^n$$

$$(\lambda+\mu)\sum_{k=1}^{\infty} p_k z^k = \frac{\mu}{z}\sum_{k=1}^{\infty} p_{k+1} z^{k+1} + \sum_{k=1}^{\infty}\sum_{i=0}^{k-1} \lambda p_i \beta_{k-i} z^k$$

Simplifying, we get $$(\lambda+\mu)[P(z)-p_0]=\frac{\mu}{z}[P(z)-p_0-p_1 z]+\lambda P(z)\beta(z)$$

$$P(z)=\frac{\mu p_0(1-z)}{\mu(1-z)-\lambda z[1-\beta(z)]}$$

Define $\rho=\dfrac{\lambda\overline{\beta}}{\mu}$ as the offered traffic

Note that, *P(1)=1* is effectively the same as the Normalization Condition. Using this, we get $p_0=1-\rho$

$$P(z)=\frac{\mu(1-\rho)(1-z)}{\mu(1-z)-\lambda z[1-\beta(z)]} \tag{2.42}$$

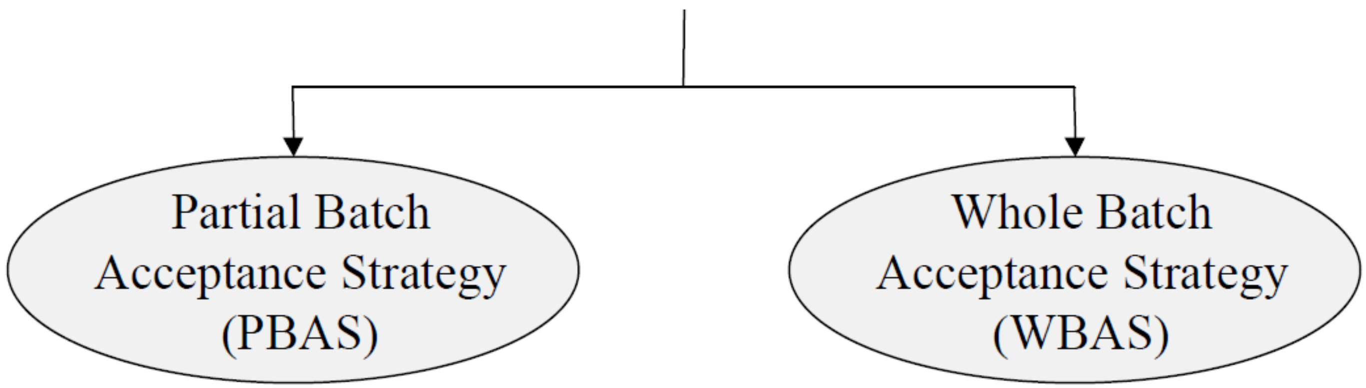
We can invert *P(z)* or expand it as a power series in z^i $i=0,1,\ldots$ to get the state probability distribution. The mean number *N* in the system may be directly calculated from *P(z)* as -

$$N=\left.\frac{dP(z)}{dz}\right|_{z=0}=\frac{\rho(\overline{\beta}+\overline{\beta^2})}{2(1-\rho)} \tag{2.43}$$

The $M^{[X]}$/-/-/K Queue

Batch Arrival Queue with Finite Capacity

For operating queues of this type, one must also specify the *batch acceptance strategy* to be followed if a batch of size *k* or more arrives in a system where the number of waiting positions available is less than *k.*



Randomly choose as many jobs from the batch as may be accommodated in the buffer

Accept the batch only if all its jobs may be accommodated; otherwise, reject all jobs of the batch

M[X}/M/-/- types of queues may be operated and analyzed under either the PBAS or the WBAS strategy

See Section 2.10 where this analysis is done for a M[X}/M/s/s queue. The state distribution for this queue are given by

$$p_j = \frac{\lambda}{j\mu} \sum_{i=0}^{j-1} p_i \phi_{j-i} \qquad j = 1, 2, \ldots, s \qquad (2.46)$$

where $\phi_i = \sum_{k=i}^{\infty} \beta_k \qquad i = 1, 2, \ldots\ldots$

Example

- For a M/M/m/∞ queue, consider a particular customer A, who on arrival finds all m servers busy and n other customers waiting for service. The queue follows a FCFS discipline. Assume that no new customers come to the queue for service after A's arrival. Let $1/\mu$ be the mean service time for customers

  (a) Find the expected length of time customer A spends waiting for service.

  (b) Find the expected length of the time interval T measured from A's arrival to the time when the system becomes completely empty

  Ans:

  (a) The expected length of time customer A spends waiting for service $= \frac{(n+1)}{m\mu}$

  (b) The expected length of time from A's arrival to the time when the system becomes empty $= \frac{(n+1)}{m\mu} + \frac{1}{\mu}\sum_{i=1}^{m}\frac{1}{i}$

(c) Define X be the order of completion of service of customer A where X=k if A is the kth customer to complete service after its arrival to the system. Find P{X=k} for k=1,2,…., (m+n-1)

Ans: (c)

For $k=1,......,(n+1)$ $\quad P\{X=k\} = 0$

For $k=n+2$,

$P\{X=k\}$ = P$\{X_A <$ residual service time of each X_i, $i=1,..,(m-1)$ in queue$\}$

$$= \int_0^\infty \mu e^{-\mu\tau}[P\{X > \tau]^{m-1} d\tau = \int_0^\infty \mu e^{-\mu\tau} e^{-(m-1)\mu\tau} d\tau = \frac{1}{m}$$

For $k=n+3$,

$$P\{X=k\} = \int_0^\infty \mu e^{-\mu\tau}(m-1)(1-e^{-\mu\tau})e^{-(m-2)\mu\tau} d\tau = \frac{1}{m}$$

In general, for $k=n+2+i$, $i=0,......(m-1)$, using $x=e^{-\mu\tau}$

$$P\{X=k\} = \binom{m-1}{i}\int_0^1 (1-x)^i x^{m-1-i} dx$$

$$= \binom{m-1}{i}\int_0^1 \sum_{j=0}^{i}(-1)^j \binom{i}{j} x^{m-1-i+j} dx = \binom{m-1}{i}\sum_{j=0}^{i}\binom{i}{j}\frac{(-1)^j}{m-i+j} = \frac{1}{m}$$

For proving the above, use the result that for $i=0,...., (m-1)$

$$I(m-1,i) = \int_0^1 \binom{m-1}{i}(1-x)^i x^{m-1-i} dx = I(m-1,i-1)$$ and that $I(m-1,0\} = I\{m-1,1\} = 1$

- (d) Find the probability that customer A is able to complete its service before the customer who is immediately ahead of him/her in the queue.

Ans:

Service to the customer served before customer A and the service to customer A will be as shown in Fig. 1.1 when A finishes service before the former. Here τ is the time interval between the start of service for these two customers in the queue. Let X_A be the duration of service for customer A and let X_1 be the service duration of the other customer.

The probability P that we need to find is then $P=P\{\tau+X_A<X_1\}$ as obtained next where $f_\tau(t)=(m-1)\mu e^{-(m-1)\mu t}$ and $f_{XA}(t)=\mu e^{-\mu t}$ for $t\geq 0$. Let $Y=\tau+X_A$ and since $\tau\perp X_A$, we can write that

$$L_Y(s) = L_\tau(s)L_{X_A}(s) = \frac{\mu^2(m-1)}{(s+\mu)[s+(m-1)\mu]}$$

$$= \frac{\mu(m-1)}{(m-2)}\left[\frac{1}{s+\mu} + \frac{1}{s+(m-1)\mu}\right]$$

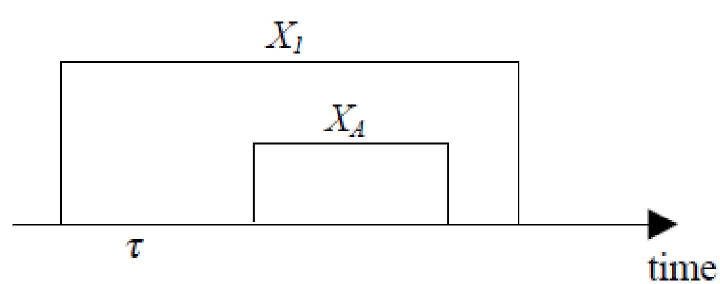


Figure 1.1. Service to A finishes before the service ends for the earlier customer

Therefore $f_Y(y) = \dfrac{\mu(m-1)}{(m-2)}\left[e^{-\mu y} - e^{-(m-1)\mu y}\right]$ for $y \geq 0$

Using this, we can find

$$P\{Y < \tau\} = \int_0^{\tau} f_Y(y)dy = \left(\frac{m-1}{m-2}\right)\left[(1 - e^{-\mu\tau}) - \frac{1}{(m-1)}(1 - e^{-(m-1)\mu\tau}\right]$$

and therefore

$$P = \int_{\tau=0}^{\infty} P\{Y < \tau\}\mu e^{-\mu\tau} d\tau = \left(\frac{m-1}{m-2}\right)\left[1 - \frac{1}{2} - \frac{1}{(m-1)}\left(1 - \frac{1}{m}\right)\right] = \frac{1}{2}\left(1 - \frac{1}{m}\right)$$

(e) Let w be the random amount of time customer A waits for service. Find its cumulative distribution function (c.d.f.)

Ans: From the definition of the Erlang-n distributions as the sum of *(n+1)* i.i.d exponentially distributed random variables, we get

$$P\{w \leq x\} = \int_0^{x} \frac{\mu(\mu\alpha)^n}{n!} e^{-\mu\alpha} d\alpha$$

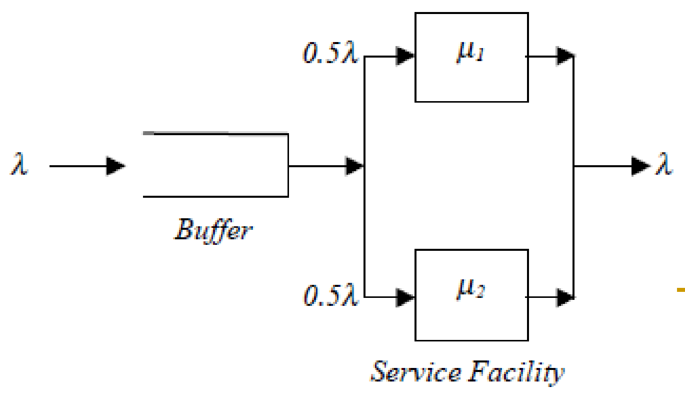
Example

- Consider the M/-/1/3 queue, which is limited to having a maximum of 3 users in the system. Assume that the arrivals come from a Poisson process of average rate l and that the L.T. if the service time distribution is

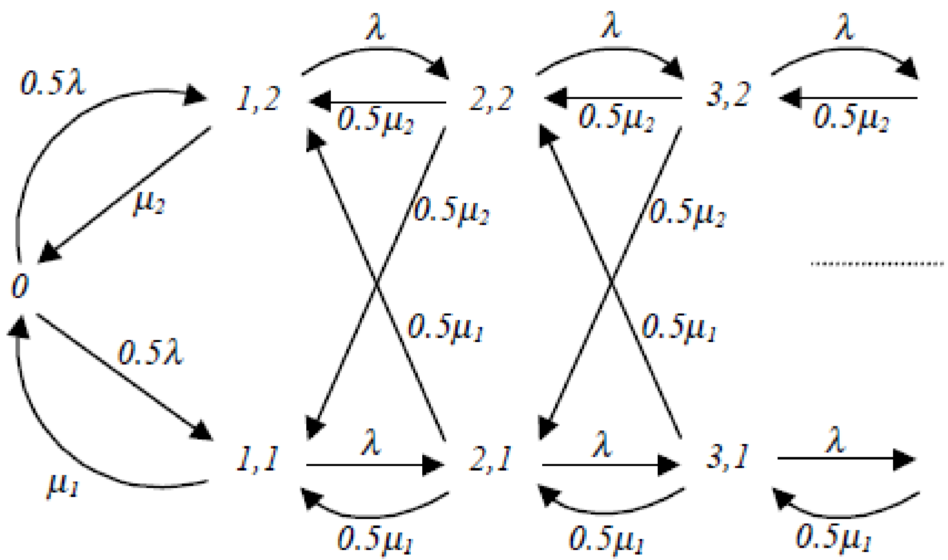
$$L_B(s) = \frac{2\mu^2}{(s+\mu)(s+2\mu)}$$

Draw the state transition diagram of the system and write its balance equations.

Ans:
We can rewrite $L_B(s)$ as $L_B(s) = \frac{1}{2}\frac{\mu_1}{(s+\mu_1)} + \frac{1}{2}\frac{\mu_2}{(s+\mu_2)}$ which implies that the service centre looks like the following.



Customers entering the service centre choose either of the two exponential servers. A new customer does not enter the service facility until the previous customer has departed. Let the state of the system be denoted by {*n, j*} where *n* is the number in the system and *j* is the server currently used by the customer in the service centre. The state transition diagram may then be drawn as follows.



State Transition Diagram

The system can now be completely solved by writing the appropriate balance equations and solving them for the individual state probabilities. In particular, we will find that

$$p_{12} = p_0 \frac{\lambda(\lambda + \mu_1)}{\lambda(\mu_1 + \mu_2) + 2\mu_1\mu_2}$$

$$p_{11} = p_0 \frac{\lambda(\lambda + \mu_2)}{\lambda(\mu_1 + \mu_2) + 2\mu_1\mu_2}$$

and

$$p_1 = p_{11} + p_{12} = p_0 \frac{\lambda(2\lambda + \mu_1 + \mu_2)}{\lambda(\mu_1 + \mu_2) + 2\mu_1\mu_2}$$

We can similarly find that

$$p_2 = p_0\left(\frac{2\lambda^2\left[(\lambda + \mu_1)^2 + (\lambda + \mu_2)^2\right]}{[\lambda(\mu_1 + \mu_2) + 2\mu_1\mu_2]^2}\right)$$