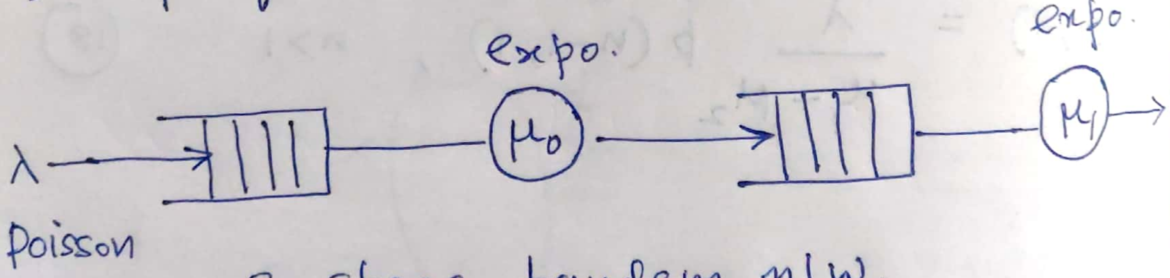
Networks of Queues

Two types of networks:

(i) Open Queuing N/W: is characterized by one or more sources of job arrivals and correspondingly one or more sinks that absorb jobs departing from the n/w.

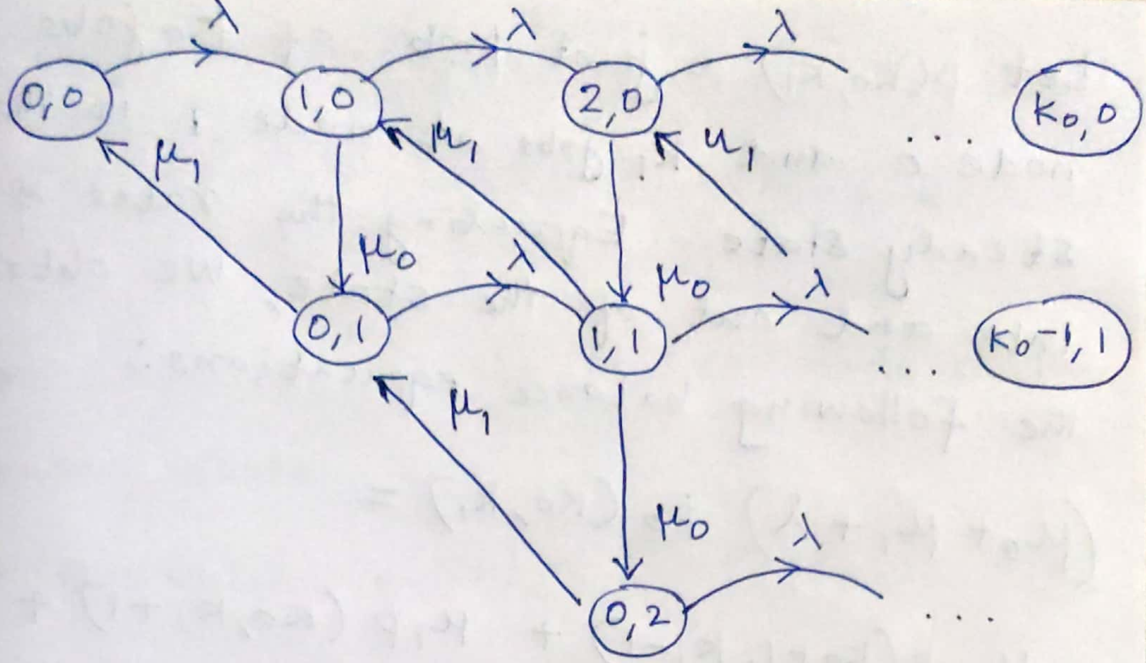
(ii) Closed Queuing n/w: jobs neither enter nor depart from the n/w.

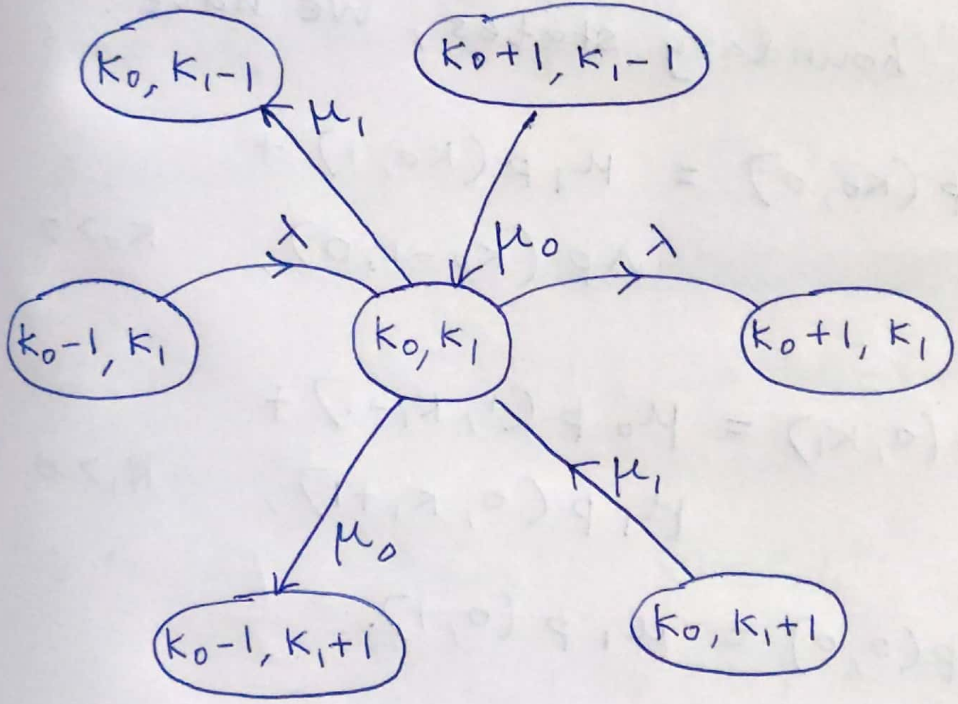
The behavior of jobs within the n/w is characterized by the prob. of transitions b/w service centres & the distr. of job service times at each center. For each center, the no. of servers, the scheduling descipline, & the size of the Queue must be specified.



2-stage tandem n/w.

This system can be modeled as a Stochastic process whose states are specified by pairs (k_0, k_1), $k_0 \geq 0$, $k_1 \geq 0$ Where k_i $(i=0,1)$ is the no. of jobs at server i in the steady state.


0,0
1,0
2,0
$k_0,0$
0,1
1,1
$k_0-1,1$
0,2
λ
μ_1
μ_0


k_0, k_1-1
k_0+1, k_1-1
k_0-1, k_1
k_0, k_1
k_0+1, k_1
k_0-1, k_1+1
k_0, k_1+1
λ
μ_1
μ_0

Let $p(k_0, k_1)$ = joint prob. of k_0 jobs at node 0 and k_1 jobs at node 1 in the steady state. Equating the rates of flow into and out of the state, we obtain the following balance equations:

$$(\mu_0 + \mu_1 + \lambda)\, p(k_0, k_1) =$$

$$\mu_0\, p(k_0+1, k_1-1) + \mu_1\, p(k_0, k_1+1) + \lambda\, p(k_0-1, k_1), \quad k_0, k_1 > 0.$$

For the boundary states, we have:

$$(\mu_0 + \lambda)\, p(k_0, 0) = \mu_1\, p(k_0, 1) + \lambda\, p(k_0-1, 0), \quad k_0 > 0$$

$$(\mu_1 + \lambda)\, p(0, k_1) = \mu_0\, p(1, k_1-1) + \mu_1\, p(0, k_1+1), \quad k_1 > 0$$

$$\lambda\, p(0,0) = \mu_1\, p(0,1).$$

The normalization is provided by:

$$\sum_{k_0 \geq 0} \sum_{k_1 \geq 0} p(k_0, k_1) = 1.$$

It is shown by direct substitution that the following eqn. is the soln to the above balance equations.

$$p(k_0, k_1) = (1-\rho_0)\rho_0^{k_0} \cdot (1-\rho_1)\rho_1^{k_1} \quad -(1)$$

where $\rho_0 = \frac{\lambda}{\mu_0}$ and $\rho_1 = \frac{\lambda}{\mu_1}$, both <1.

pmf of no. of jobs N_0 at node 0 in steady state

$$p(N_0 = k_0) = p_0(k_0) = (1-\rho_0)\rho_0^{k_0} \quad -(2)$$

o/p of an M/M/1 Queue is also poisson with rate λ.

pmf of no. of jobs N_1 at node 1 is

$$\cancel{p(k_0,k_1)}\ p(N_1 = k_1) = p(k_1) = (1-\rho_1)\rho_1^{k_1} \quad -(3)$$

$$\cancel{p_0(k_0)\, p_1(k_1)}$$

The joint prob. of k_0 jobs at node 0 and k_1 jobs @ node 1 is:

$$p(k_0, k_1) = (1-\rho_0)\rho_0^{k_0}(1-\rho_1)\rho_1^{k_1}$$

$$= p_0(k_0) \cdot p_1(k_1) \quad -(4).$$

Hence, $p(k_0, k_1)$ is the prod. of marginal probabilities $p_0(k_0)\ p_1(k_1)$; hence r.v. N_0 and N_1 are independent in the steady state. The 2 queues are independent M/M/1 Queues.

Ex: A repair facility shared by a large no. of m/c's has 2 sequential stations with 1/hr & 2/hr. respectively. The cummulative failure rate of all machines is 0.5/hr. Determine the average repair time.

Soln: Given $\lambda = 0.5$, $\mu_0 = 1$, $\mu_1 = 2$.

$\rho_0 = \frac{0.5}{1} = 0.5$; $\rho_1 = \frac{0.5}{2} = 0.25$.

Avg. length of queue @ station i $(i=0,1)$ is

$$E[N_i] = \frac{\rho_i}{1-\rho_i}$$

$$E[N_0] = \frac{0.5}{1-0.5} = 1$$

$$E[N_1] = \frac{0.25}{1-0.25} = \frac{0.25}{0.75} = \frac{1}{3}.$$

Using Little's formula; the repair delay at the two stations is

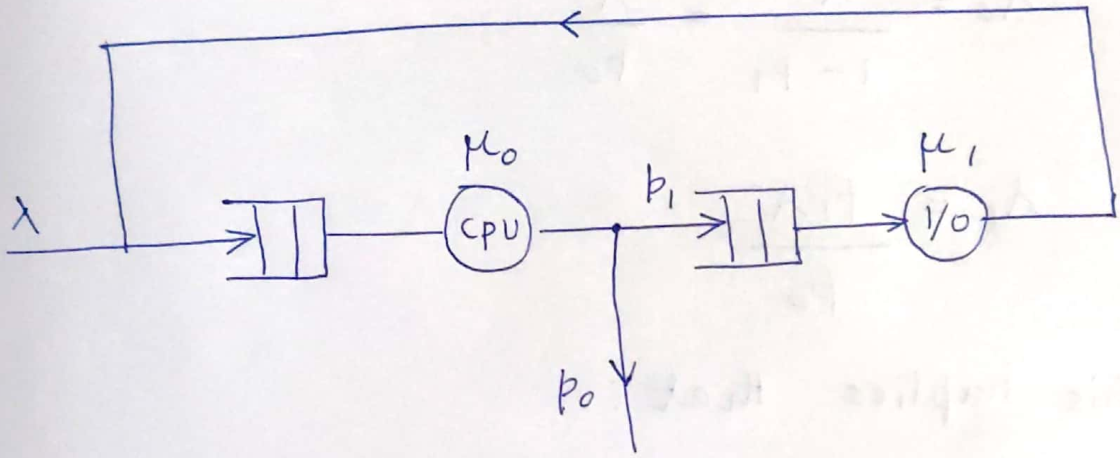
$$E[R_0] = \frac{E[N_0]}{\lambda} = \frac{1}{0.5} = 2 \text{ hrs.}$$

$$E[R_1] = \frac{E[N_1]}{\lambda} = \frac{2}{3} \text{ hrs.}$$

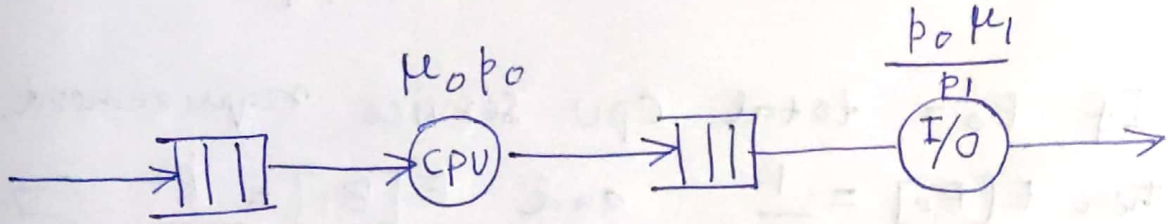
Hence, the avg. repair time is :

$$E[R] = E[R_0] + E[R_1] = 2 + \frac{2}{3} = \frac{8}{3} \text{ hrs.}$$

Open Queuing N/W :



(a) with feedback.



(b) without feedback.

$$p(k_0, k_1) = (1-\rho_0)\rho_0^{k_0} \cdot (1-\rho_1)\rho_1^{k_1} \quad — (5)$$

where $\rho_0 = \frac{\lambda_0}{\mu_0}$ & $\rho_1 = \frac{\lambda_1}{\mu_1}$.

Total arrival rate @ CPU node is

$\lambda_0 = \lambda + \lambda_1$ (outside + I/O)

Avg. arrival rate @ device 1 is:

$\lambda_1 = \lambda_0 p_1$

Thus,

$$\lambda_0 = \frac{\lambda}{1 - p_1} = \frac{\lambda}{p_0}$$

$$\lambda_1 = \frac{p_1 \lambda}{p_0}$$

This implies that:

$$\rho_0 = \frac{\lambda}{p_0 \mu_0} \quad \text{and} \quad \rho_1 = \frac{p_1 \lambda}{p_0 \mu_1}.$$

If B_0 = total CPU service requirement,

then $E[B_0] = \frac{1}{(p_0 \mu_0)}$ and $E[B_1] = \frac{p_1}{(p_0 \mu_1)}$ (I

If $\rho_0 > \rho_1$, then $E[B_0] > E[B_1]$ or

CPU is bottleneck or CPU-bound,

Else I/O bound.

Avg. response time

$$E[R] = \left[\frac{\rho_0}{1-\rho_0} + \frac{\rho_1}{1-\rho_1} \right] \frac{1}{\lambda}$$

or

$$= \frac{1}{p_0 \mu_0 - \lambda} + \frac{1}{\frac{p_0 \mu_1}{p_1} - \lambda}$$

$$= \frac{E[B_0]}{1 - \lambda E[B_0]} + \frac{E[B_1]}{1 - \lambda E[B_1]} \quad - (6)$$