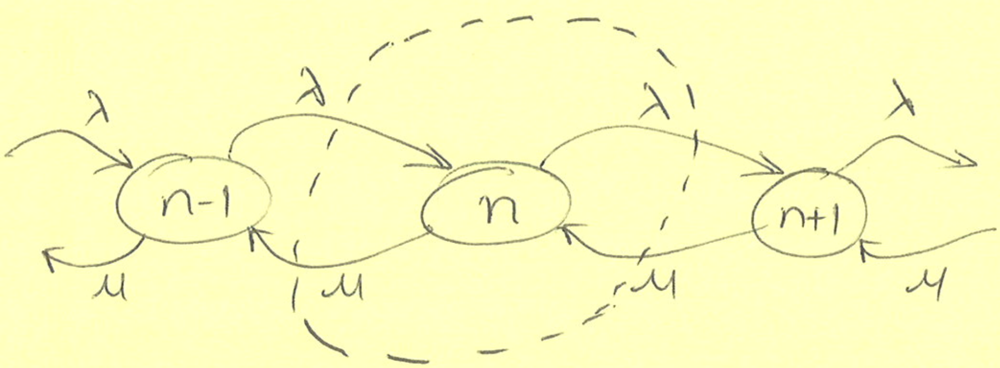
Global balance.

Why? Because, consider this:



Electric current
||
Similar to Kirchoff's law

Rate of entering state n is $\boxed{\lambda p_{n-1} + \mu p_{n+1}}$

prob. flux into state

Rate of exiting state n is $\lambda p_n + \mu p_n$ $\boxed{= (\lambda+\mu) p_n}$

prob. flux out of state.

In other words, ~~mean~~ rate at which a state is entered = ~~mean~~ rate at which it is exited.

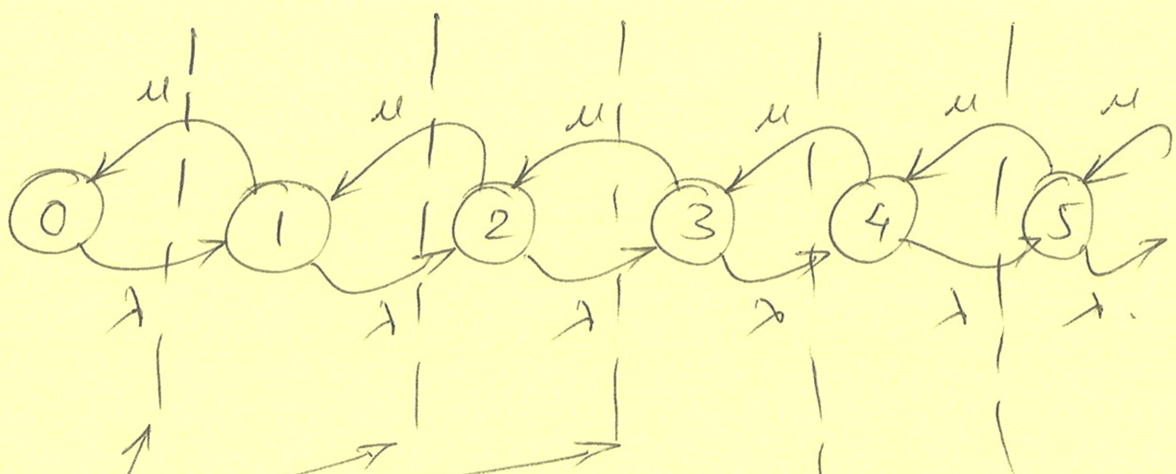
(

→ Global balance gives us N linear simultaneous equations, for a system with N states.

i.e. 1 equation for every state.

☆ → note $p_0 + p_1 + p_2 + \dots + p_N = 1$. (why?)

Local balance.



local balance boundaries.

Solving for local balance:

$$\underline{\lambda p_0 = \mu p_1} \quad \text{i.e.} \quad p_1 = \frac{\lambda}{\mu} p_0$$

$$\underline{\lambda p_1 = \mu p_2} \quad \text{i.e.} \quad p_2 = \frac{\lambda}{\mu} p_1$$

$$\therefore \; p_2 = \frac{\lambda}{\mu} \cdot \left(\frac{\lambda}{\mu} p_0\right)$$

$$\therefore \; p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0 .$$

similarly $\lambda p_2 = \mu p_3$ gives us.

$$p_3 = \left(\frac{\lambda}{\mu}\right)^3 p_0$$

By mathematical induction, we get.

$$\boxed{p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 .} \quad \text{—— (19).}$$

also note that $\sum_{n=0}^{\infty} p_n = 1.$ {i.e. all s.s. probabilities must sum to 1.}

Using (19), we get:

$$\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n p_0 = 1$$

$$\therefore \; \boxed{p_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n}} \quad \text{—— (20).}$$

For a stable system, $0 \leq \frac{\lambda}{\mu} < 1$.

λ/μ is the load $\left(= \frac{\text{arrival rate}, \lambda}{\text{departure rate}, \mu} \right)$

let $\lambda/\mu = \rho$.

then $0 \leq \rho < 1$. for stability.

Then Eq. (20) can be rewritten as:

$$p_0 = \frac{1}{\sum_{n=0}^{\infty} \rho^n} = \frac{1}{\frac{1}{1-\rho}} \quad \text{if } 0 \leq \rho < 1$$

[Since $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$, if $0 \leq x < 1$]

$$\therefore \boxed{p_0 = 1 - \rho} \qquad (21)$$

Substitute this in (19) to get

$$\boxed{p_n = \rho^n \cdot (1 - \rho)} \qquad (22)$$

System utilization, (U)

The Utilization for M/M/1 system is defined as the fraction of time for which the server has at least 1 customer.

~~i.e. $U = 1 - \sum_{n=1}^{\infty} P$~~

In other words, it is the probability that the system has at least 1 customer.

i.e. $U = \sum_{n=1}^{\infty} P_n = P_1 + P_2 + P_3 + \dots + P_{\infty}$

$$\therefore U = 1 - P_0$$

$$\therefore U = 1 - (1 - \rho)$$

$$\therefore \boxed{U = \rho} \qquad \text{———} \qquad (23).$$

Average number of customers in system :

$$\bar{n} = E[N] = \sum_{n=0}^{\infty} n p_n = \sum_{n=1}^{\infty} n p_n$$

(i.e. a summation of n, weighted by the corresponding probability p_n, for all possible values of n.)

from Eq.(22) & above:

$$\therefore \bar{n} = \sum_{n=\emptyset}^{\infty} n \cdot \rho^n (1-\rho)$$

$$\therefore \bar{n} = (1-\rho) \sum_{n=\emptyset}^{\infty} n \rho^n$$

→ or directly $= \frac{\rho}{(1-\rho)^2}$

see pg. 338 Robertazzi

$$= (1-\rho) \cdot \rho \sum_{n=\emptyset}^{\infty} n \cdot \rho^{n-1}$$

note

$$= (1-\rho) \cdot \rho \cdot \sum_{n=1}^{\infty} \frac{d}{d\rho} \rho^n$$

Taking the derivative outside the sum
(this is legal if both the derivative & the sum exist)

$$\therefore \bar{n} = \rho \cdot (1-\rho) \cdot \frac{d}{d\rho} \sum_{n=1}^{\infty} \rho^n$$

$$= \rho \cdot (1-\rho) \cdot \frac{d}{d\rho} \left(\sum_{n=0}^{\infty} \rho^n - \rho^0 \right)$$

→ note again

$$= \rho \cdot (1-\rho) \cdot \frac{d}{d\rho} \left(\frac{1}{1-\rho} - 1 \right)$$

$\left\{ \text{since } \sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \text{ if } 0 \leq x < 1 \right.$

$$= \rho (1-\rho) \cdot \left[\frac{0 - (-1)}{(1-\rho)^2} - 0 \right]$$

$$\boxed{\therefore \bar{n} = \frac{\rho}{1-\rho}} \quad \ldots (24)$$

Miscellaneous measures.

Variance of the number of customers in the system:

$$\mathrm{Var}[N] = \frac{\rho}{(1-\rho)^2}$$ Derivation not provided.

End

For any system (not just M/M/1 system)

Little's law (or Little's Result).

$$\boxed{\bar{n} = \lambda \bar{T}} \quad \text{--- (25)}$$

where $\bar{n}$ = mean # of customers in system
λ = arrival rate of customers.
$\bar{T}$ = mean response time.

Proof: not part of this course.
But there are at least 4 different ways to prove it.

1. Using calculus ($\int$ and Σ)
2. Using areas (geometry) and a little bit of calculus (limits)
3. Using heuristics
4. Using Laplace transforms.

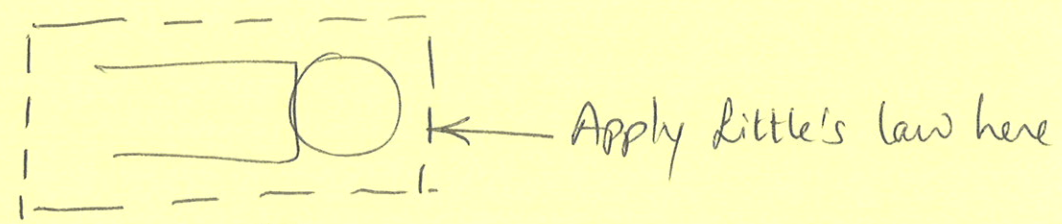
—x—

A little more about Little's law:

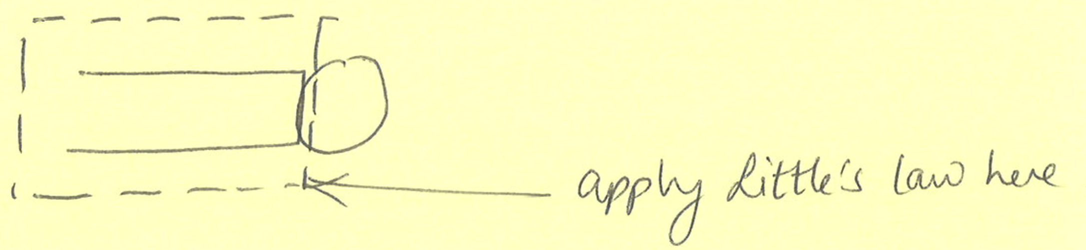
$\bar{n} = \lambda \bar{T}$ ..... from (25).

The above equation can be applied to the whole system or it can be can be applied to just the queue (without the server).

i.e.



or



∴ We can redesignate equation (25) as:

$\bar{n}_{system} = \lambda \bar{T}_{system}$ where $\bar{T}_{system} = \bar{T}_{waiting} + \bar{T}_{service}$

& $\bar{n}_{system} = \cancel{\bar{n}_{queue}}$ cab be obtained as shown in previous derivation.

∴ we can get $\bar{T}_{system}$.

or

$\bar{n}_{queue} = \lambda \bar{T}_{waiting}$ where $\bar{n}_{queue} = \bar{n}_{system} - 1$.

where $\bar{n}_{system}$ is obtained as shown in previous derivations.

<u>note that</u> λ is independent of how you choose to apply Little's law

The M/M/1/N queuing system (Finite Buffer case)

Local balance eqns

$\lambda p_0 = \mu p_1 \quad \therefore p_1 = \frac{\lambda}{\mu} p_0$

$\lambda p_1 = \mu p_2 \quad \therefore p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0$

$\vdots$

$\lambda p_{N-1} = \mu p_N \quad \therefore p_N = \left(\frac{\lambda}{\mu}\right)^N p_0$ $\left[\text{in general, } p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 \text{ where } 0 \le n \le N\right]$

& $p_0 = 1 - \sum_{n=1}^{N} p_n = 1 - \sum_{n=1}^{N} p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n$

$$\therefore p_0 = \frac{1}{1 + \sum_{n=1}^{N} \rho^n} \quad \text{where } \rho = \frac{\lambda}{\mu}.$$

$$\therefore p_0 = \frac{1}{\sum_{n=0}^{N} \rho^n}$$

$$\therefore p_0 = \frac{1}{\frac{1-\rho^{N+1}}{1-\rho}} \quad \left\{ \text{since } \sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x} \right\}$$

$$\therefore \boxed{p_0 = \frac{1-\rho}{1-\rho^{N+1}}} \quad \text{------} (26)$$

Substitute (26) in $p_n = (\rho)^n \cdot p_0$ to get:

$$\boxed{p_n = \frac{\rho^n \cdot (1-\rho)}{1-\rho^{N+1}}} \quad [0 \leq n \leq N] \qquad \text{— (27)}.$$

When the system is full, note that there are a maximum of N customers in the system (i.e. 1 in server & N-1 in the queue). Any new arrivals at this point will be rejected (i.e. new calls will be ~~dropped~~ blocked / new packets will be discarded.).

∴ The state probability P_N is:

p_N = probability that there are N customers in system
= probability that system is full
= probability that new incoming calls/pkts will be dropped.

∴ P_N is also called the "blocking probability" or the "probability of call rejection" or the "packet-drop probability". ($\equiv P_B$)

Thus, if λ packets arrive into the system per second, on the average, then the ~~probability that a packet~~ no. of packets that will be dropped due to full buffer is $\underline{\lambda p_n}$. per unit time.

In percentage terms, $(\lambda p_n \times 100)\%$ of incoming packets will be dropped!

Performance figures for M/M/1/N:

1. Utilization, $U = \sum_{n=1}^{\infty} p_n$

$$= 1 - p_0$$

$$= 1 - \frac{1-\rho}{1-\rho^{N+1}} = \frac{\cancel{1} - \rho^{N+1} - \cancel{1} + \rho}{1-\rho^{N+1}}$$

$$\therefore \boxed{U = \frac{\rho(1-\rho^N)}{1-\rho^{N+1}}} \quad \text{——— (28).}$$

2. Avg. # of customers in system:

$$\bar{n} = E[N] = \sum_{n=0}^{N} n\, p_n$$

$$= \sum_{n=0}^{N} n \cdot \frac{\rho^n (1-\rho)}{1-\rho^{N+1}}$$

$$= \frac{(1-\rho)}{1-\rho^{N+1}} \cdot \sum_{n=0}^{N} n \cdot \rho^n$$

~~solve further~~

$$= \frac{\rho(1-\rho)}{1-\rho^{N+1}} \cdot \sum_{\underline{\underline{n=1}}}^{N} n \cdot \rho^{n-1} \longleftarrow \text{note !!!}$$

$$= \frac{\rho(1-\rho)}{1-\rho^{N+1}} \cdot \sum_{n=1}^{N} \frac{d}{d\rho} \rho^n$$

$$= \frac{\rho(1-\rho)}{1-\rho^{N+1}} \cdot \frac{d}{d\rho} \sum_{n=0}^{N} \cdot \rho^N - \rho^0$$

$$\therefore \bar{n} = \frac{\rho(1-\rho)}{1-\rho^{N+1}} \cdot \cancel{\tfrac{\rho}{n}} \cdot \frac{d}{d\rho}\left(\frac{1-\rho^{N+1}}{1-\rho} - 1\right)$$

$$\left(\text{since } \sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x}\right)$$

(solve this yourself to get ....)

check

$$\boxed{\bar{n} = \frac{\rho^N(N\rho - N - 1) + 1}{(1-\rho)^2}} \times \frac{\rho(1-\rho)}{1-\rho^{N+1}} \quad \text{--- (29)}$$

$\Rightarrow$ Let me know if there is an error in eqn. (29)!

$\therefore$ we know $\bar{n}$ (from eq. 29). We also know λ $\Rightarrow$ usually given!

$\therefore$ from little's law, we can find the ~~waiting time~~ response time $\bar{T}$, in the system.

$$\therefore \bar{n} = \frac{\rho^{N+1}(N\rho - N - 1) + 1}{(1-\rho)(1-\rho^{N+1})}$$

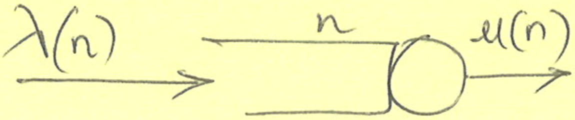
Note: pages 28 & 29 do not exist

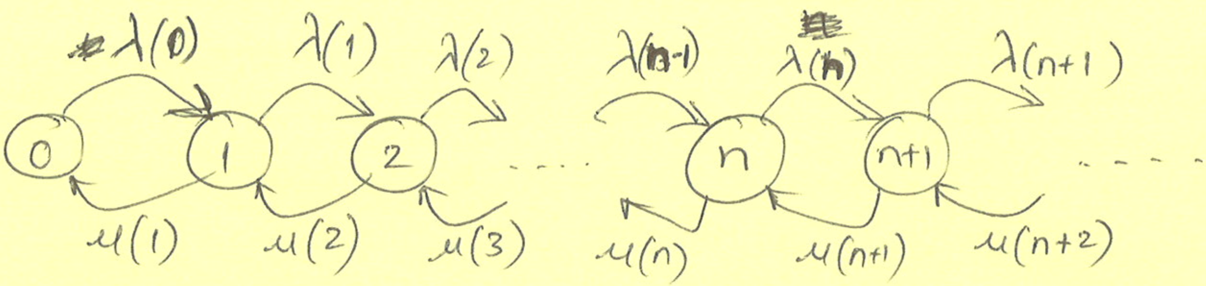
State Dependent M/M/1 system:

Note that arrival rate & service rates can change without affecting the arrival & service time distributions. i.e. the means of the distribution can change without affecting the distribution itself.

When the arrival & service rates depend on the states themselves, then the system is called a state dependent system.

i.e. consider:





where $\lambda(0) \neq \lambda(1) \neq \lambda(2) \neq \cdots \neq \lambda(n) \neq \cdots$
and $\mu(1) \neq \mu(2) \neq \cdots \qquad \cdots$ $\Big\}$ (A)

Note that the strict inequality above is not necessary to classify a ~~markovian~~ system as state dependent. Even one inequality in (A) is enough.

* Home exercise: *

Use local balance equations for the abovementioned state dependent M/M/1 system to show that

$$\boxed{p_n = \left(\prod_{i=1}^{n} \frac{\lambda(i-1)}{\mu(i)} \right) p_0} \quad \text{------ (30)}$$

and

$$\boxed{p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda(i-1)}{\mu(i)}}} \quad \text{------ (31)}$$

~~Response time~~

Perf. measures of interest to designers.

- U Utilization
- $\bar{J}$ Response time, waiting time, service time
- $\bar{Y}$ mean throughput.
- $\bar{n}$ mean # of customers

For any system:

$$U = 1 - p_0$$

$\bar{J} \rightarrow$ from Little's law

$\bar{n} \rightarrow$ from our derivations, using p_n terms.

$\bar{Y} \rightarrow$? (we usually need p_n before we can compute $\bar{Y}$).

for $M/M/1/\infty$,

$\bar{y} = \lambda$ (as measured on the input side). iff $\rho < 1$

also $\bar{y} = \underbrace{\mu \cdot \sum_{n=1}^{\infty} p_n}_{\text{dep. rate} \times P(\text{at least 1 cust in sys.})}$ (if measured on the output side).

for $M/M/1/N$

$\bar{y} = \overbrace{\lambda \cdot (1 - P_B)}^{\text{arr. rate} \times P(\text{packet will be accepted})}$ {if measured on the i/p side}

and $\bar{Y} = \mu \cdot \sum_{n=1}^{N} p_n$ {if measured on the o/p side}.

for M/M/1/∞ , state-dependent :

$$\bar{Y} = \sum_{n=0}^{\infty} \lambda(n) \cancel{p_n} \, p_n \qquad \{\text{if calculated on i/p side}\}$$

and $$\bar{Y} = \sum_{n=1}^{\infty} \mu(n) \, p_n \qquad \{\text{if calculated on o/p side}\}.$$

for state-dependent M/M/1/N :

$$\bar{Y} = \sum_{n=0}^{N-1} \lambda(n) \, p_n . \qquad \{\text{if calculated on i/p side}\}$$

and $$\bar{Y} = \sum_{n=1}^{N} \mu(n) \, p_n \qquad \{\text{if calculated on o/p side}\}$$

The M/G/1 queuing system:

M/G/1 = M/G/1/∞
where G = "General"
i.e. service time
probability distribution is
arbitrary.

For M/G/1 systems:
$E[N]$ i.e. $\bar{n}$ is very hard to compute using the methods ~~men~~ we tried before.
But we have a very powerful result called the "Pollaczek-Khinchin" mean value formula.
According to the PK formula:

$$\boxed{E[N] = \frac{2\rho - \rho^2 + \lambda^2 \sigma_s^2}{2(1-\rho)}} \qquad (32)$$

where ρ = normalized load (i.e. $\frac{\lambda}{\mu}$)
λ = arrival rate
σ_s^2 = variance of service time distribution ~~distributed~~
~~e.g.~~

Suppose we want to see if we can get the $\bar{n}$ for an M/M/1 system. Then, we let the G in M/G/1 stand for a Markovian system. i.e. the service times are no longer general; they are exponential (for M/M/1).

Then for an M/M/1 system.

$$E[N] = \frac{2\rho - \rho^2 + \lambda^2 \cdot \left(\frac{1}{\mu^2}\right)}{2(1-\rho)}$$

since $\sigma_s^2 = \frac{1}{\mu^2}$ for exponential distribution

$$\therefore \boxed{E[N] = \frac{\rho}{1-\rho}} \longrightarrow$$

check this with eqn. (24)

Eureka!

Q: Why is the PK result so powerful?

Ans: Because: (in M/G/1 system):

1. We only know that the arrivals are Poisson.
2. The service time distribution is "General". i.e we have no clue as to what it is!
3. Even so, we need only the 2nd moment (i.e. variance) of that service time distribution to calculate E[N]. NO HIGHER ORDER MOMENTS ARE NEEDED

∴ E[N] computation is relatively easy.

4. Once we get E[N], we can get $\bar{T}$ from Little's law.

Examples from text:

2.16) a) The arrival rate in a queuing sys. is 10 cust/sec. mean time for a customer to move through the system is 2 seconds. What is the mean # of customers in the queuing system?

Sol^n: Use Little's law: $\bar{n} = \lambda \bar{T} = 10 \times 2 = 20$ customers.

b) $\lambda = 10$ per sec in an M/M/1 queueing system. service rate $\mu = 20$ cust. per sec. ∴ Utilization?

sol^n: $U = \rho = 1/2$ (ie. $\frac{10}{20}$).

c) Calculate the Utilization for the data in (b), assuming that the system has finite buffer (ie. M/M/1/N)

Soln:

2.17) For the state independent M/M/1 queuing system. find the ratio P_{100}/P_0 when $\lambda = 1.0$ and $\mu = 2.0$. Comment on the potential numerical problems.

$$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 = 0.5^n P_0$$

$$\therefore \frac{P_{100}}{P_0} = \frac{0.5^n \cancel{P_0}}{\cancel{P_0}} = (0.5)^n = (0.5)^{100}$$

$$= 7.88 \times 10^{-31}$$

∴ even for only 100 customers (pkts) the state probabilities span large orders of magnitude. This implies that the numerical solution of realistically-sized models may be prone to underflow – overflow problems.

2.19) In an M/M/1 system, the $\rho = 0.3$. There is approx. a 97% probability that 'm' or less customers are present in the queuing system at any time. Find m.

Sol^n.

~~i.e.~~ $$\sum_{n=0}^{m} p_n \doteq 0.97$$

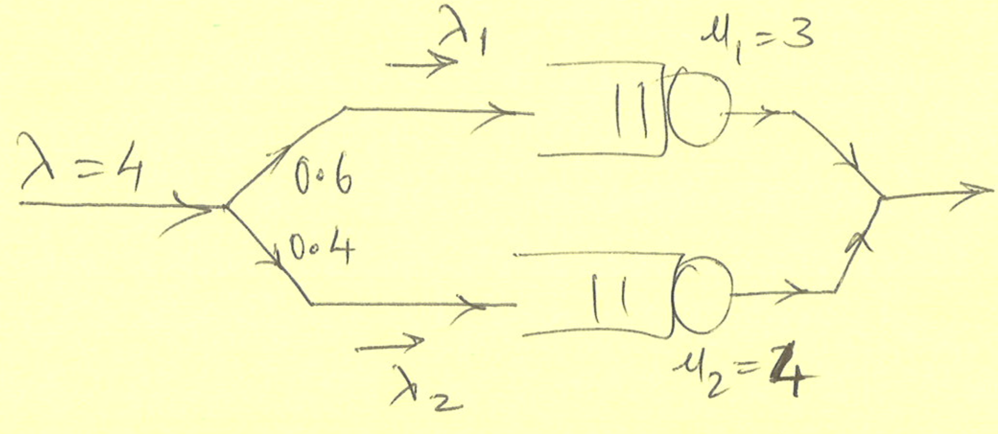
For M/M/1 sys, $p_n = \rho^n(1-\rho) = 0.3^n(1-0.3)$

$$\therefore \sum_{n=0}^{m} 0.3^n \times 0.7 \doteq 0.97$$

$$\therefore \quad m = 2$$

~~2.20~~

Problem :



In the above diagram of ~~a~~ standard M/M/1 queues, the fractions (0.4 & 0.6) indicate the iid probabilities of routing arriving jobs to the two branches. Determine the expected number of customers for the whole system. What is the expected response time for the whole system?

Soln:

(1) $\lambda_1 = 0.6 \times 4 = 2.4$
$\lambda_2 = 0.4 \times 4 = 1.6$

$\therefore \rho_1 = \lambda_1/\mu_1 = 2.4/3 = 0.8$
$\rho_2 = \lambda_2/\mu_2 = 1.6/4 = \cancel{0.8}\ 0.4$

$$E[N_1] = \frac{\rho_1}{1-\rho_1} = \frac{0.8}{1-0.8} = \underline{\underline{4}}$$

$$E[N_2] = \frac{\rho_2}{1-\rho_2} = \frac{0.4}{1-0.4} = \frac{.4}{.6} = 2/3$$

$\therefore E[\text{Total } N \text{ in entire system}] = E[N_1] + E[N_2] = 4 + \frac{2}{3}$

$\therefore E[N_{\cancel{total}\ system}] = \underline{\underline{4.666}}$

(2) $E[T_{system}] = ?$

$= \cancel{0.6 \times E[T_1] + E[T_2]}$

$= 0.6 \times E[T_1] + 0.4 \times E[T_2]$.

$$= 0.6 \times \frac{E[N_1]}{\lambda_1} + 0.4 \times \frac{E[N_2]}{\lambda_2}$$

$$= 0.6 \times \frac{4}{2.4} + 0.4 \times \frac{0.667}{1.6}$$

$=$ ______