M/M/1 and M/M/m Queueing Systems

M. Veeraraghavan; March 20, 2004

1. Preliminaries

1.1 Kendall's notation: G/G/n/k queue

G: General - can be any distribution.

First letter: Arrival process; M: memoryless - exponential interarrival times - Poisson arrival process

Second letter: Service times distribution - M: exponential, D - Deterministic

Third letter: Number of servers, n

Fourth letter: Number in system (including number in queue and number being served)

Important: **In all cases,** service times s_n are mutually independent of each other, and are also independent of interarrival times. Successive interarrival times are iid with the distribution specified by the first letter. Successive service times are also iid with the distribution specified by the second letter. This means the arrival process and departure process are both renewal processes - iid random sequence.

1.2 Relation between M/M/1 queue and MC

Why can the process $N(t)$, the number of customers in the system at time t in an M/M/1 queue, be modeled as a Markov chain?

Answer: [4, pg. 126] Given that the memoryless property PLUS the independence assumption of interarrival and service times, the number of customers in the system at a future time $t + h$ only depends upon the number in the system now (at time t) and the arrivals or departures that occur within the interval h. Past history of how the system got to its state at time t is irrelevant. Additional time needed to complete service of customer being served observes the memoryless property.

$$N(t+h) = N(t) + X(h) - Y(h), \tag{1}$$

where $X(h)$ is the number of arrivals in the time interval $(t, t+h)$, and $Y(h)$ is the number of departures in the time interval $(t, t+h)$. $X(h)$ is dependent only on h because the arrival process is Poisson. $Y(h)$ is 0 if the service time of the customer being served, $s_1 > h$. $Y(h)$ is 1, if $s_1 \le h$ and $s_2 + s_1 > h$, and so on. As the service times $s_1, s_2, \ldots, s_n$ are independent, neither $X(h)$ nor $Y(h)$ depend on what happened prior to t. Thus, $N(t+h)$ only depends upon $N(t)$ and not the past history. Hence it is a CTMC.

2. Analysis of the M/M/1 queue using CTMC results: [3], page 365-371.

First consider a special case of an irreducible time-homogeneous MC, i.e., a birth-death process. A homogeneous CTMC is a birth-death process if there exists constants λ_i, $i = 0, 1, \ldots$, and μ_i, $i = 0, 1, \ldots$ such that the transition rates are given by:

$$q_{i,\,i+1} = \lambda_i,\ q_{i,\,i-1} = \mu_i,\ q_i = \lambda_i + \mu_i \text{ and } q_{ij} = 0 \text{ for } |i-j| > 1\ . \tag{2}$$

The reason $q_i = \lambda_i + \mu_i$ is as follows: See (11) in Markov Chain lecture, which states that

$$p_{jj}(t, t+h) = 1 - q_j(t) \bullet h + o(h) \text{, which for a homogeneous MC can be rewritten as} \tag{3}$$

$$p_{jj}(h) = 1 - q_j \bullet h + o(h) \tag{4}$$

Remembering that the sum of all transition probabilities out of a state is 1, $p_{jj}(t, t+h)$ should be equal to $1 - \sum_{\forall i} p_{ji}(t, t+h)$. This is equal to $1 - (\lambda_i + \mu_i)h + o(h)$. Therefore $q_j = \lambda_i + \mu_i$.

Use the global balance equations derived for steady-state solution of an irreducible, homogeneous CTMC (eqn 32 of MC.pdf):

$$\sum_{k \ne j} p_k q_{kj} - q_j p_j = 0 \tag{5}$$

combined with the equation $\sum_j p_j = 1$, where p_j is the steady-state limiting probability of the system being in state j and q_{kj} and q_j are the transition rates. Applying this to our birth-death MC, we get

$$-(\lambda_j + \mu_j)p_j + p_{j+1}\mu_{j+1} + p_{j-1}\lambda_{j-1} = 0 \tag{6}$$

$$-\lambda_0 p_0 + \mu_1 p_1 = 0 \tag{7}$$

Eqn. (7) is the same as the detailed balance equation we showed for a birth-death DTMC in MC.pdf. See Fig. 1 for a birth-death CTMC.

Note the difference between the state diagram of a CTMC and the state diagram of a DTMC. In the latter the arcs are labeled with conditional probabilities; in the former they are labeled with transition rates - so the latter is sometime called transition-rate diagram - this from [3], page 367.

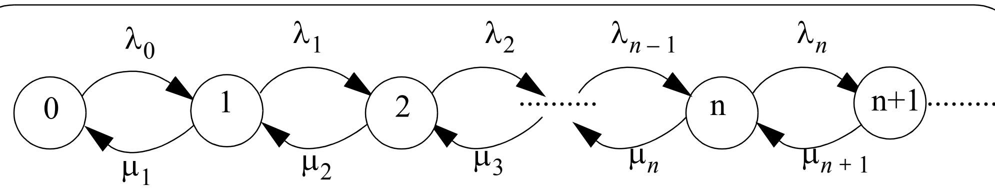


Figure 1:CTMC for a birth-death process; interesting that we don't show the self-loops

Rearranging (6), we get:

$$\lambda_j p_j - \mu_{j+1} p_{j+1} = \lambda_{j-1} p_{j-1} - \mu_j p_j \tag{8}$$

Similarly,

$$\lambda_{j-1} p_{j-1} - \mu_j p_j = \lambda_{j-2} p_{j-2} - \mu_{j-1} p_{j-1} \tag{9}$$

and so on until:

$$\lambda_1 p_1 - \mu_2 p_2 = \lambda_0 p_0 - \mu_1 p_1 \ . \tag{10}$$

Therefore:

$$\lambda_j p_j - \mu_{j+1} p_{j+1} = \lambda_{j-1} p_{j-1} - \mu_j p_j = \lambda_{j-2} p_{j-2} - \mu_{j-1} p_{j-1} = \ldots = \lambda_0 p_0 - \mu_1 p_1 \tag{11}$$

From (7), since $\lambda_0 p_0 - \mu_1 p_1 = 0$,

$$\lambda_{j-1} p_{j-1} - \mu_j p_j = 0 \quad \text{and hence} \quad p_j = \left(\frac{\lambda_{j-1}}{\mu_j}\right) p_{j-1} \quad \text{for } j \geq 1 \tag{12}$$

Therefore:

$$p_j = \frac{\lambda_{j-1}\lambda_{j-2}\ldots\lambda_0}{\mu_j \mu_{j-1}\ldots\mu_0} p_0 = p_0 \prod_{i=0}^{j-1} \frac{\lambda_i}{\mu_{i+1}} \quad \text{for } j \geq 1 \tag{13}$$

Since $\sum_j p_j = 1$:

$$p_0 = \frac{1}{1 + \sum_{j \geq 1} \prod_{i=0}^{j-1} \frac{\lambda_i}{\mu_{i+1}}}. \tag{14}$$

The number of customers in an M/M/1 queue is a homogeneous, irreducible birth-death CTMC in which $\lambda_i = \lambda$ for $\forall i$ and $\mu_i = \mu$ for $\forall i$. Applying (13) and (14) to this case yields:

$$p_j = \left(\frac{\lambda}{\mu}\right)^j p_0 \text{ and } p_0 = \frac{1}{1 + \sum_{j \geq 1} \left(\frac{\lambda}{\mu}\right)^j} = \frac{1}{\sum_{j \geq 0} \left(\frac{\lambda}{\mu}\right)^j} = 1 - \frac{\lambda}{\mu}, \text{ if } \frac{\lambda}{\mu} < 1. \tag{15}$$

We define traffic intensity, $\rho = \lambda/\mu$. $\rho < 1$ for a stable system. Server utilization $U = 1 - p_0 = \rho$.

The mean number of customers in the system in the steady-state can be computed:

$$E[N] = \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n\rho^n(1-\rho) = \rho(1-\rho)\sum_{n=0}^{\infty} n\rho^{n-1} \tag{16}$$

$$E[N] = \rho(1-\rho)\sum_{n=0}^{\infty}\left(\frac{\partial}{\partial\rho}\rho^n\right) = \rho(1-\rho)\frac{\partial}{\partial\rho}\left(\sum_{n=0}^{\infty}\rho^n\right) = \rho(1-\rho)\frac{\partial}{\partial\rho}\left(\frac{1}{(1-\rho)}\right) \tag{17}$$

$$E[N] = \rho(1-\rho)\frac{1}{(1-\rho)^2} = \frac{\rho}{(1-\rho)} \tag{18}$$

$$\text{Derive } Var[N] = \frac{\rho}{(1-\rho)^2} \tag{19}$$

Mean response time (using Little's Law):

$$E[T] = \frac{E[N]}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu-\lambda} \tag{20}$$

Mean waiting time in queue:

$$E[W] = E[T] - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\rho}{\mu-\lambda} \tag{21}$$

Mean number of customers in queue (again using Little's Law):

$$E[N_Q] = \lambda E[W] = \frac{\rho^2}{(1-\rho)}; \text{ also } E[N_Q] = E[N] - \rho = \frac{\rho}{(1-\rho)} - \rho = \frac{\rho^2}{(1-\rho)} \tag{22}$$

Subtract ρ and not 1 because ρ is the probability that the server is busy.

Above results hold for disciplines other than FCFS. It holds for any scheduling discipline as long as [3], page 370):

1. The server is not idle when there are jobs waiting for service (work conserving)
2. The scheduler is not allowed to use any deterministic a priori information about job service times (e.g., Shortest Remaining Processing Time First - SPRT will reduce E[R]).
3. The service time distribution is not affected by the scheduling discipline.

3. Derivation of M/M/1 queue results using DTMC

Both [4] and [5] analyze the M/M/1 queue using a DTMC. Focus attention on the time instants: $0, \delta, 2\delta, 3\delta, \ldots$, where δ is a small positive number. Let N_k be the number of customers in the system at time $k\delta$. The process $\{N_k | k = 0, 1, \ldots\}$ is a DTMC with the same steady-state occupancy distribution as those of the CTMC $N(t)$. The DTMC for an M/M/1 queue is shown in Fig.

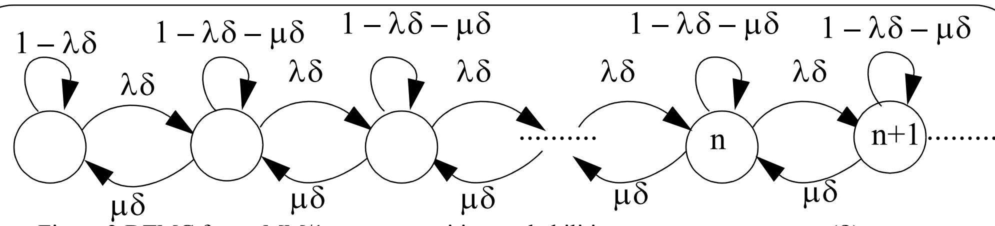


Figure 2:DTMC for an MM/1 queue; transition probabilities are correct up to an $o(\delta)$ term

2. It is a birth-death DTMC. The transition probabilities $P_{ij} = P\{N_{k+1} = j | N_k = i\}$ is independent of k for a time-homogeneous DTMC.

$$P_{00} = 1 - \lambda\delta + o(\delta) = e^{-\lambda\delta} \text{ (Poisson arrival process)} \tag{23}$$

$$P_{ii} = 1 - \lambda\delta - \mu\delta + o(\delta) = e^{-\lambda\delta}e^{-\mu\delta} \quad \text{for } i \ge 1, \tag{24}$$

which is the prob. of 0 arrivals and 0 departures in interval $(k\delta, (k+1)\delta)$.

$$P_{i,\, i+1} = \lambda\delta + o(\delta) = e^{-\mu\delta}\lambda\delta e^{-\lambda\delta} \text{ for } i \ge 0, \tag{25}$$

which is the prob. of 1 arrival and 0 departures in interval $(k\delta, (k+1)\delta)$

$$P_{i,\, i-1} = \mu\delta + o(\delta) = e^{-\lambda\delta}\mu\delta e^{-\mu\delta} \text{ for } i \ge 1, \tag{26}$$

which is the prob. of 0 arrivals and 1 departure in interval $(k\delta, (k+1)\delta)$

$$P_{ij} = o(\delta) \text{ for } |i-j| > 1 \tag{27}$$

Neglect all terms in the second order of δ.

With this DTMC, we can find the stationary distribution (steady-state) using the derivation in MC.pdf for time-homogeneous irreducible DTMCs:

$$\nu_n = \lim_{k \to \infty} P\{N_k = n\} = \lim_{t \to \infty} P\{N(t) = n\} \tag{28}$$

Using the detailed balance equations (see the lecture MC.pdf) - since it is a birth-death process:

$$\nu_n p_{n(n+1)} = \nu_{n+1} p_{(n+1)n} \qquad n = 0, 1, \ldots \tag{29}$$

Applying it to the DTMC of the M/M/1 queue shown in Fig. 2:

$$p_n \lambda\delta + o(\delta) = p_{n+1}\mu\delta + o(\delta) \tag{30}$$

As $\delta \to 0$,

$$p_{n+1} = p_n\left(\frac{\lambda}{\mu}\right) \text{ for } n = 0, 1, 2, \ldots \tag{31}$$

Setting $\rho = \lambda/\mu$, we get the same results as with the CTMC derivation.

4. Distribution of response time

This DOES depend upon the scheduling discipline (unlike the average response time). We derive the response time distribution assuming the FCFS (First Come First Serve) scheduling discipline. Let the arriving job find n jobs in the system. Response time $R = S + S'_1 + S_2 + \ldots + S_n$, where S is the service time of the arriving job, S'_1 is the remaining time of the job in service and S_i, $i = 2, \ldots, n$ are the service times of the $(n-1)$ jobs in the queue. These are n independent random variables. Given the memoryless property of the exponential distribution, we can write the Laplace transform of the conditional distribution of R given $N = n$ as

$$L_{R|N}(s|n) = \left(\frac{\mu}{s+\mu}\right)^{n+1} \tag{32}$$

Laplace transform of a nonnegative continuous random variable X is defined as [3], page 196:

$$L_X(s) = L(s) = M_X(-s) = \int_0^{\infty} e^{-sx} f(x)dx \tag{33}$$

If X is a nonnegative integer-valued discrete random variable, then we define its z-transform as:

$$G_X(z) = G(z) = E[z^X] = M_X(\ln z) = \sum_{i=0}^{\infty} p_X(i) z^i, \tag{34}$$

where moment generating function $M_X(\theta)$ is given by (if X is a r.v., $X\theta$ is also a r.v.):

$$M_X(\theta) = E[e^{X\theta}] \tag{35}$$

The convolution theorem: Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables and let $Y = \sum_{i=1}^{n} X_i$. Then if $M_{X_i}(\theta)$ exists for all i, then:

$$M_Y(\theta) = M_{X_1}(\theta) M_{X_2}(\theta) \ldots M_{X_n}(\theta), \tag{36}$$

i.e., the moment generating function of a sum of independent random variables is the product of the moment generating functions.

Proof:

$$M_Y(\theta) = E[e^{(X_1 + X_2 + \ldots + X_n)\theta}] = E\left[\prod_{i=1}^{n} e^{X_i\theta}\right] = \prod_{i=1}^{n} E[e^{X_i\theta}] \text{ by independence} \tag{37}$$

$$\text{Hence } M_Y(\theta) = \prod_{i=1}^{n} M_{X_i}(\theta) \tag{38}$$

Sum of mutually exclusive events $P(A \cup B) = P(A) + P(B)$ and product of independent random variables $P(X_1 X_2) = P(X_1) P(X_2)$. So for a sum of independent random variables, we use the theorem that the product of moment generating functions of the r.vs. is the moment generating function of the sum.

Going back to (32), and applying the theorem of total Laplace transform (like the theorem of total probability), and using (15) for p_n, we get:

$$L_R(s) = \sum_{n=0}^{\infty} \left(\frac{\mu}{s+\mu}\right)^{n+1} (1-\rho)\rho^n \tag{39}$$

$$L_R(s) = \frac{\mu(1-\rho)}{s+\mu}\left(\frac{1}{1-\frac{\mu\rho}{s+\mu}}\right) = \frac{\mu(1-\rho)}{s+\mu-\mu\rho} = \frac{\mu(1-\rho)}{s+\mu(1-\rho)} \tag{40}$$

This means R is exponentially distributed with parameter $\mu(1-\rho)$.

Laplace transform for an exponentially distributed random variable X is:

$$L_X(s) = \int_0^\infty e^{-sx}\lambda e^{-\lambda x}dx = \frac{\lambda}{s+\lambda} \tag{41}$$

5. M/M/m queue

The underlying Markov chain is again a birth-death process with

$$\lambda_k = \lambda \text{ for } k = 0, 1, 2, \dots \text{ and} \tag{42}$$

$$\mu_k = \begin{cases} k\mu & 0 < k < m \\ m\mu & k \geq m \end{cases} \tag{43}$$

Plugging these into (13) yields:

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda}{(i+1)\mu} = p_0\left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \text{ for } k < m\text{, and} \tag{44}$$

$$p_k = p_0 \prod_{i=0}^{m-1} \frac{\lambda}{(i+1)\mu} \prod_{j=m}^{k-1} \frac{\lambda}{m\mu} = p_0\left(\frac{\lambda}{\mu}\right)^k \frac{1}{m!m^{k-m}} \text{ for } k \geq m. \tag{45}$$

Here is the interesting thing regarding traffic intensity. We called ρ **traffic intensity** in the M/M/1 queue derivation as it is called in [3], page 368. In [6], page 824, ρ is called **traffic load**. In an M/M/1 queue, the utilization $(1-p_0)$ (i.e., probability that the server is not idle) turns out to be equal to ρ. For an M/M/m queue, the condition for stability is $\frac{\lambda}{m\mu} < 1$. Traffic load should not be dependent on m. But here both [3] and [4] set $\rho = \lambda/(m\mu)$. NOTE THIS REDEFINITION OF ρ (confusing to redefine ρ but it helps the analytics).

Using (45) and $\sum_{k=0}^{\infty} p_k = 1$, we derive p_0 as:

$$p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \sum_{k=m}^{\infty} \frac{(m\rho)^k}{m!m^{k-m}}\right]^{-1} \tag{46}$$

$$p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \sum_{k=m}^{\infty} \rho^{k-m}\right]^{-1} = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \sum_{i=0}^{\infty} \rho^{i}\right]^{-1} \tag{47}$$

$$p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \frac{1}{(1-\rho)}\right]^{-1} \tag{48}$$

$$E[N] = \sum_{k=0}^{\infty} kp_k = m\rho + \rho\frac{(m\rho)^m}{m!}\frac{p_0}{(1-\rho)^2} \text{ (Exercise: derive this)} \tag{49}$$

If M is an r.v. denoting the number of busy servers, then:

$$E[M] = \sum_{k=0}^{m-1} kp_k + m\sum_{k=m}^{\infty} p_k = \sum_{k=0}^{m-1} kp_k + m\sum_{k=m}^{\infty}\left(p_0(m\rho)^k\frac{1}{m!m^{k-m}}\right) \tag{50}$$

$$E[M] = \sum_{k=0}^{m-1} kp_k + \frac{mp_0}{m!}(m\rho)^m\left(\frac{1}{1-\rho}\right) = \sum_{k=0}^{m-1} kp_k + \frac{mp_m}{1-\rho} \tag{51}$$

$$\text{where } p_m = p_0(m\rho)^m\frac{1}{m!} \tag{52}$$

Exercise: Show that $E[M] = m\rho = \lambda/\mu$. **Therefore the utilization of any individual server is $(E[M])/m = \rho$, while the average number of busy servers is equal to traffic intensity λ/μ. I'd define utilization of the system to be $E[M]$ and hence equal to the traffic intensity (same as in M/M/1 queue). See [7], page 492: ρ is called per-server utilization and ρm is called total traffic intensity.** Unit of λ/μ is defined in [7], page 492 as measured in the *Erlangs*. λ/μ is also referred to as *total load* on the system. Same as in [6].

Probability of an arriving job finding that all m servers are busy is called **Erlang-C formula** and is given by:

$$P[queueing] = \sum_{k=m}^{\infty} p_k = \frac{p_m}{1-\rho} \text{ (see (51) and (52))} \tag{53}$$

$$P_Q = P[queueing] = \frac{(m\rho)^m}{m!}\frac{p_0}{1-\rho} \tag{54}$$

The above is a call queueing system in a telephone network with an infinite buffer. It's also called Erlang's delayed-call formula.

Average number in queue (not in service) is:

$$E[N_Q] = \sum_{k=m+1}^{\infty} (k-m)p_k = \sum_{k=m+1}^{\infty} (k-m)p_0\left((m\rho)^k \frac{1}{m!m^{k-m}}\right) \tag{55}$$

$$E[N_Q] = p_0\frac{(m\rho)^m}{m!}\sum_{k=m+1}^{\infty}(k-m)\rho^{k-m} = p_0\frac{(m\rho)^m}{m!}\sum_{n=0}^{\infty} n\rho^n \tag{56}$$

Using (54) and $(1-\rho)\left(\sum_{n=0}^{\infty} n\rho^n\right) = \frac{\rho}{1-\rho}$, we can express the mean number waiting in queue

as

$$E[N_Q] = P_Q(1-\rho)\frac{\rho}{(1-\rho)^2} = P_Q\frac{\rho}{(1-\rho)} \tag{57}$$

Using Little's Law,

$$E[W] = P_Q\frac{\rho}{\lambda(1-\rho)} \text{ and} \tag{58}$$

Using $\rho = \lambda/(m\mu)$, we get

$$E[T] = E[W] + \frac{1}{\mu} = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda} \tag{59}$$

Using Little's Law,

$$E[N] = \left(\frac{1}{\mu} + \frac{P_Q}{m\mu-\lambda}\right)\lambda = m\rho + \frac{\rho P_Q}{1-\rho} \text{, which is the same as (49).} \tag{60}$$

Compare M/M/1 queue with a fast server operating at $m\mu$ with an M/M/m system:

Average packet delay (mean response time) in the M/M/m system is:

$$E[T] = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda} \tag{61}$$

Average packet delay in an M/M/1 queue with service rate $m\mu$ is:

$$E[\hat{T}] = \frac{1}{m\mu - \lambda} = \frac{1}{m\mu} + \frac{\hat{P}_Q}{m\mu - \lambda} \tag{62}$$

In an M/M/1 queue

$$\hat{P}_Q = \frac{(m\rho)^m}{m!}\frac{p_0}{1-\rho} = \rho \text{ because } p_0 = 1 - \rho\,. \tag{63}$$

$$E[T] = \frac{1}{m\mu - \lambda} = \frac{1}{m\mu} + \frac{\rho}{m\mu - \lambda} = \frac{(m\mu - \lambda) + \rho m\mu}{m\mu(m\mu - \lambda)} = \frac{1}{(m\mu - \lambda)}. \tag{64}$$

Under light loads $\rho \ll 1$, $P_Q = \hat{P}_Q \cong 0$ then

$$\frac{E[T]}{E[\hat{T}]} \cong m \tag{65}$$

Under heavy loads, $P_Q = \hat{P}_Q \cong 1$ and $1/\mu \ll 1/(m\mu - \lambda)$, then

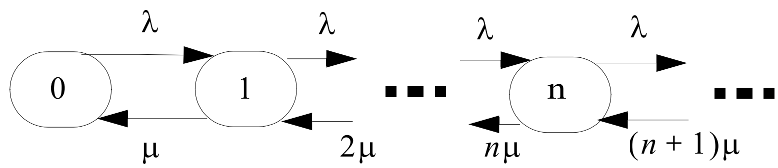
$$\frac{E[T]}{E[\hat{T}]} \cong 1 \tag{66}$$

The above is a comparison of statistical multiplexing (where whole link capacity can be used) and TDM/FDM schemes where once a circuit capacity is assigned, it is not increased during the life-time of the call.

6. M/M/∞

Suppose $P(X(t) = n)$ is the probability that n telephone lines are busy at time t. Assume that infinitely many lines are available and that the call arrival rate is λ while average call duration is $1/\mu$. Find $P(X(t) = n)$ as $t \to \infty$. Let $E[N(t)]$ denote the average number of busy lines at time t. Find the value of $E[N(t)]$ when the system reaches steady state. Also find the average response time for this queueing system in the steady state.

This is a M/M/∞ queueing system.



State transition diagram

As $t \to \infty$, we have

$$P(X = n) = \frac{\lambda}{n\mu} P(X = n - 1) = \left(\frac{\lambda}{\mu}\right)^n \frac{P(X = 0)}{n!} \tag{67}$$

Thus,

$$P(X = 0) = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}} = e^{-\lambda/\mu} = e^{-\rho} \tag{68}$$

$$E[N] = \sum_{n=1}^{\infty} nP(X = n) = \lambda/\mu \tag{69}$$

Since there are always enough telephone lines for any call, the response time follows exponential distribution with the average value of $1/\mu$.

References

[1] K. S. Trivedi, "Probability, Statistics with Reliability, Queueing and Computer Science Applications," Second Edition, Wiley, 2002, ISBN 0-471-33341-7.
[2] R. Yates and D. Goodman, "Probability and Stochastic Processes," Wiley, ISBN 0-471-17837-3.
[3] K. S. Trivedi, "Probability, Statistics with Reliability, Queueing and Computer Science Applications," First Edition, Prentice Hall, 1982, ISBN 0-13-711564-4r.
[4] D. Bertsekas and R. Gallager, "Data Networks," Prentice Hall, 1986, ISBN 0-13-196825-4.
[5] Prof. Boorstyn's notes, Polytechnic University, NY.
[6] A. Leon Garcia and I. Widjaja, "Communication Networks," McGraw Hill, 2000, First Edition.
[7] Mischa Schwartz, "Telecommunications Networks, Protocols, Modeling and Analysis," Addison Wesley, 1987.