Analysis of the M/G/1/-/N Queue

We consider here the basic M/G/1 queue again, but with the limitation, that the number of sources generating jobs for the queue is finite. Specifically, we assume that there are N sources generating jobs that require service. These jobs are buffered in the system's queue if their service cannot start immediately. Moreover, a source that has already generated a job for which service is pending (i.e. the job is either buffered in the system or is presently in service) cannot generate a new job until the previous (pending) job is served. A source may therefore be considered to be in one of two states - *thinking* when it can generate a new job and *blocked* when its previous job has not yet completed service.

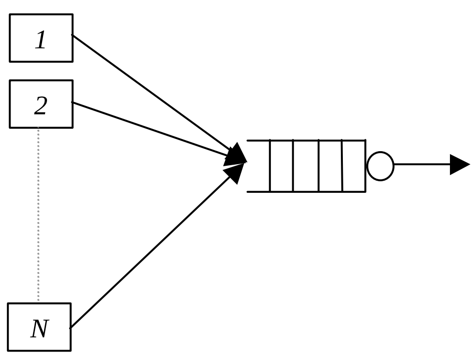

Figure 1. A M/G/1/-N Queue

The overall system of N sources and the M/G/1 queue may be considered to be in state $n(t)$ at time t if there are $n(t)$ jobs (waiting and in-service) in the queue at that time. A thinking user is assumed to generate jobs following a Poisson process at rate l while a blocked user cannot generate new jobs until its previous job is served. A system of this type may be represented as shown in Fig. 1. Note that the overall arrival rate to a system in state n will also be Poisson with rate $(N-n)l$. The service times of the individual jobs are assumed to be generally distributed with pdf $b(t)$ and cdf $B(t)$ as before. We also assume that $L_B(t)$ is the Laplace Transform of the pdf $b(t)$ and that the mean service time is $E\{X\}=m^{-1}$.

To analyse this system, we again imbed the Markov Chain at the departure instants of the jobs leaving the system (i.e. leaving the queue after service). (It should be noted that to maintain consistency with our earlier definitions, we consider the system to be the queue with its server.) Let n_i be the number in the system left behind when the i^{th} departure leaves the system after getting its required service. We can then see that the system state at the $(i+1)^{th}$ departure will be related to the system state at the i^{th} departure as

$$
\begin{aligned}
n_{i+1} &= a_{i+1} && n_i = 0 \\
&= n_i - 1 + a_{i+1} && n_i \geq 1
\end{aligned}
\qquad (1)
$$

where a_{i+1} is the number of jobs arriving to the system during the $(i+1)^{th}$ service time. Note that this system will be unconditionally stable. As l increases, in the worst case, all the N sources will be blocked and their corresponding jobs either will be at the server or will be waiting in the queue. In that case, the system state will saturate to N and cannot blow up to infinity to cause instability. It may also be noted that the state at the job departure instants may only range between 0 and $(N-1)$ - this is because the job finishing service and departing is not considered in the representation of the system state.

Considering the system at equilibrium, let $p_{d,k}$ be the probability of state k for this chain (at the departure instants) and let $p_{d,jk}$ be its one-step transition probability for going from state j to state k, $0£ j,k£(N-1)$. We can then write

$$
\begin{aligned}
p_{d,0k} &= \binom{N-1}{k} \int_{t=0}^{\infty} e^{-(N-1-k)lt} (1-e^{-lt})^{k}\, b(t)dt && j = 0 \\
p_{d,jk} &= \binom{N-j}{k-j+1} \int_{t=0}^{\infty} e^{-(N-1-k)lt} (1-e^{-lt})^{k-j+1} b(t)dt && j = 1,......,k+1 \\
p_{d,jk} &= 0 && otherwise
\end{aligned}
\qquad (2)
$$

The corresponding state probabilities $p_{d,k}$ $k=0,1,.....(N-1)$ may then be found by solving the following set of equations

$$\begin{aligned} p_{d,k} &= \sum_{j=0}^{N-1} p_{d,j}\, p_{d,jk} \qquad k = 0,1,........,(N-1) \\ \sum_{k=0}^{N-1} p_{d,k} &= 1 \qquad \textit{normalization condition} \end{aligned} \tag{3}$$

This is most conveniently done by using the generating function $P_d(z)$, defined as

$$P_d(z) = \sum_{k=0}^{N-1} p_{d,k}\, z^{N-k-1} \tag{4}$$

The somewhat unusual form of the z-transform of the sequence $p_{d,0}$,, $p_{d,2}$,....$p_{d,\,N-1}$ should be noted. Unlike the usual definition of the z-transform, the definition given in (4) reverses the sequence for computational convenience. Substituting (3) in (4), we get

$$P_d(z) = \sum_{j=0}^{N-1} p_{d,j} \sum_{k=0}^{N-1} p_{d,jk}\, z^{N-k-1} \tag{5}$$

Using (2) in the inner summation gives

$$\begin{aligned} \sum_{k=0}^{N-1} p_{d,0k}\, z^{N-k-1} &= \int_{t=0}^{\infty} \left(1+(z-1)e^{-lt}\right)^{N-1} b(t)dt \qquad j=0 \\ \sum_{k=0}^{N-1} p_{d,jk}\, z^{N-k-1} &= \int_{t=0}^{\infty} \left(1+(z-1)e^{-lt}\right)^{N-j} b(t)dt \qquad j=1,....,(N-1) \end{aligned} \tag{6}$$

Substituting these in (5) yields

$$\begin{aligned} P_d(z) = {} & p_{d,0} \int_{t=0}^{\infty} \left(1+(z-1)e^{-lt}\right)^{N-1} b(t)dt \\ & + \sum_{j=1}^{N-1} p_{d,j} \int_{t=0}^{\infty} \left(1+(z-1)e^{-lt}\right)^{N-j} b(t)dt \end{aligned} \tag{7}$$

Simplifying (6) gives

$$\begin{aligned} P_d(z) &= \int_{t=0}^{\infty} \left[\sum_{j=0}^{N-1} p_{d,j} \left(1+(z-1)e^{-lt}\right)^{N-j} \right] b(t)dt \\ &\quad + p_{d,0} \int_{t=0}^{\infty} \left[\left(1+(z-1)e^{-lt}\right)^{N-1} - \left(1+(z-1)e^{-lt}\right)^{N} \right] b(t)dt \\ &= \int_0^{\infty} \left(1+(z-1)e^{-lt}\right) P_d\left(1+(z-1)e^{-lt}\right) b(t)dt \\ &\quad + p_{d,0}(1-z) \int_0^{\infty} \left(1+(z-1)e^{-lt}\right)^{N-1} e^{-lt} b(t)dt \end{aligned} \tag{8}$$

Note that (8) gives a functional equation for the generating function $P_d(z)$ as defined in (4). This would have to be solved to get the actual generating function from which the state probabilities at the departure instants may be calculated. To solve this, we consider $P_d(z)$ as an expansion defined differently as

$$P_d(z) = \sum_{n=0}^{N-1} a_n (z-1)^n \tag{9}$$

Expanding this and matching its terms with those of (4), we can show that

$$a_n = \sum_{k=n}^{N-1} \binom{k}{n} p_{d,N-k-1} \qquad n=0,1,\ldots\ldots,(N-1) \tag{10}$$

and

$$p_{d,k} = \sum_{n=N-k-1}^{N-1} a_n \binom{n}{N-k-1} (-1)^{n-N+k+1} \qquad k=0,1,\ldots\ldots,(N-1) \tag{11}$$

In order to use the above, we define $y=z-1$ and substitute this in (9) to get

$$P_d(y) = \sum_{n=0}^{N-1} a_n y^n \tag{12}$$

A similar substitution in (8) gives

$$P_d(y) = \sum_{n=0}^{N-1} a_n \int_0^{\infty} (1 + ye^{-lt})\left(ye^{-lt}\right)^n b(t)dt \\ - p_{d,0}\, y \int_0^{\infty} e^{-lt}\left(1 + ye^{-lt}\right)^{N-1} b(t)dt \tag{13}$$

Simplifying (13) using the Laplace Transform $L_B(s)$ of the pdf $b(t)$ and using (12), we can write

$$\sum_{n=0}^{N-1} a_n y^n = \sum_{n=0}^{N-1} a_n \left[y^n L_B(nl) + y^{n+1} L_B((n+1)l) \right] \\ - p_{d,0}\, y \sum_{n=0}^{N-1} \binom{N-1}{n} y^n L_B((n+1)l) \tag{14}$$

Comparing the coefficients of y^n on both sides of (14) for $n=1$, ..., $N-1$ will allow us to find recurrence relations for the various a_n for $n=1$,....., $N-1$ as given subsequently in (16). However, we cannot find a_0 with this approach as can be verified by considering the coefficient of y^0 in (14). To find a_0, we need to use the normalisation condition on (12) as

$$a_0 = P_d(y)\big|_{y=0} = P_d(z)\big|_{z=1} = 1 \tag{16}$$

As stated earlier, by comparing the coefficients of y^n on both sides of (14) for $n=1$, ..., $N-1$, we get

$$\begin{aligned} a_n &= \left[a_n + a_{n-1} - p_{d,0} \binom{N-1}{n-1} \right] L_B(nl) & n &= 1, \ldots\ldots, (N-1) \\ 0 &= a_{N-1} L_B(nl) - p_{d,0} L_B(nl) & n &= N \end{aligned} \tag{17}$$

Note that (16) and (17) give $N+1$ equations which need to be solved for the $N+1$ unknowns $a_0, a_1, \ldots\ldots, a_{N-1}$ and $p_{d,0}$. Defining a sequence b_i $i=0,1,\ldots, N-1$ as

$$\begin{aligned} b_0 &= 1 & i &= 0 \\ b_i &= \prod_{j=1}^{i} \frac{L_B(jl)}{1 - L_B(jl)} & i &= 1, \ldots\ldots, (N-1) \end{aligned} \tag{18}$$

it may be shown that (17) may be written in the form

$$\frac{a_{n-1}}{b_{n-1}} - \frac{a_n}{b_n} = \binom{N-1}{n-1}\frac{p_{d,0}}{b_{n-1}} \qquad n = 1,......,(N-1)$$
$$a_{N-1} = p_{d,0} \tag{19}$$

To find a_k, we sum the first equation of (19) from $n=(k+1)$ to $n=(N-1)$ to get

$$\frac{a_k}{b_k} - \frac{a_{N-1}}{b_{N-1}} = \sum_{n=k+1}^{N-1}\binom{N-1}{n-1}\frac{p_{d,0}}{b_{n-1}} \qquad k=1,........,(N-1)$$

Using the result $a_{N-1}=p_{d,0}$ from (19), we can simplify this to

$$a_k = b_k\, p_{d,0} \sum_{n=k}^{N}\binom{N-1}{n}(b_n)^{-1} \qquad k=1,........,(N-1) \tag{20}$$

Similarly summing the first equation of (19) from $n=1$ to $n=(N-1)$, we get

$$\frac{a_0}{b_0} - \frac{a_{N-1}}{b_{N-1}} = \sum_{n=1}^{N-1}\binom{N-1}{n-1}\frac{p_{d,0}}{b_{n-1}}$$

Simplifying this with $a_{N-1}=p_{d,0}$, $a_0=1$ and $b_0=1$, gives

$$p_{d,0} = \frac{1}{\sum_{n=0}^{N-1}\binom{N-1}{n}(b_n)^{-1}} \tag{21}$$

and

$$a_k = b_k \frac{\sum_{n=k}^{N}\binom{N-1}{n}(b_n)^{-1}}{\sum_{n=0}^{N-1}\binom{N-1}{n}(b_n)^{-1}} \qquad k=0,1,........,(N-1) \tag{22}$$

Note that since b_i $i=0,1,...,$ N-1 are known from (18), the sequence a_k $k=0,1,...,$ N-1 may be calculated using (22). These can, in turn, be used in (11), to get the state equilibrium probabilities $p_{d,k}$ $k=0,1,...,$ N-1 at the job departure instants.

In order to find the system performance parameters for this queue, consider the jobs generated by one of the N sources in the system. Note that once a job is generated, it will spend a mean time W in the system (waiting and in service). After this job is serviced, the source can generate another job again. The mean time between completion of the job from the i^{th} source and the generation of the next job by this source would be l^{-1} - this comes from the memoryless exponential interarrival time distribution of the Poisson source. The throughput rate of an individual source will then be $1/(W+l^{-1})$. The system throughput rate g will be the sum of the throughputs of all the N (identical) sources and will be given by

$$g = \frac{N}{W + \frac{1}{l}} \tag{23}$$

The overall system throughput rate g multiplied by the mean service time m^{-1} may also be interpreted as the fraction of time (probability) that the server at the queue is not idle. Let p_0 be the equilibrium probability that the server is found idle when the system is examined at an arbitrary time instant. (Note that this is not the same as $p_{d,0}$ which is the probability of finding the server idle at the job departure instants.) We can then get

$$1 - p_0 = \frac{g}{m} \tag{24}$$

Eliminating g from (23) and (24), we have

$$W = \frac{N}{(1 - p_0)m} - \frac{1}{l} \tag{25}$$

As in Section 3.4, we can consider the time axis to be divided in *Idle Periods* (IP, i.e. when the server is idle) and *Busy Periods* (BP, i.e. when the server is busy). The probability p_0 may then be expressed as

$$p_0 = \frac{E\{IP\}}{E\{BP\} + E\{IP\}} \tag{26}$$

When the server is idle, the arrival process to the queue is Poisson with rate Nl and, therefore, $E\{IP\}=1/(Nl)$. Moreover, since the busy period terminates with a job departure instant which leaves the system empty, the mean length of the busy period will be $1/(mp_{d,0})$. (This latter may be seen by observing that if $p_{d,0}$ is the probability that the system is empty after a job finishes service then $1/p_{d,0}$ will be the average number of jobs that are served in the busy period.) Using these in (26), we can calculate p_0 as

$$p_0 = \frac{\frac{1}{N l}}{\frac{1}{N l} + \frac{1}{m p_{d,0}}} = \frac{p_{d,0}}{p_{d,0} + \frac{N l}{m}} \tag{27}$$

Substituting this result in (24) and (25), we can get the throughput g and the mean time spent in system (waiting and in service) to be

$$g = \frac{N l}{p_{d,0} + \frac{N l}{m}} \tag{28}$$

$$W = \frac{N}{m} - \frac{1 - p_{d,0}}{l} \tag{29}$$

Actual values for these may be found by substituting the value of $p_{d,0}$ from (21) in (28) and (29) to get the throughput and mean delay of this M/G/1/-/N queue.

It is important to note that the equilibrium probabilities $p_{d,k}$ $k=0,1,...,(N-1)$ found in the above analysis only give the state probabilities of the system at the job departure instants, i.e. the number left behind in the system as seen by a job leaving after it gets its required service. These probabilities will not be the same as the state probabilities p_k $k=0,1,....,N$ observed if the system is examined at an arbitrary time instant. (We have only found one of these probabilities, i.e. p_0, in (27) above.) The equilibrium state probabilities obtained by examining the system at an arbitrary instant of time may be found using a level crossing argument (originally due to Takacs) is given next. Other approaches for this may be found in [Takagi2].

Consider the M/G/1/-/N system and let D_k be the rate of state transitions from state $k+1$ to state k because of a service completion, i.e. the rate of the down transitions. Similarly, let U_k be the rate of state transitions from state k to state $k+1$ because of an arrival, i.e. the rate of the up transitions. (Note that the transitions can only be value +1 or -1 as multiple arrivals and service completions are not possible.)

Since $E\{IP\}=(N l)^{-1}$ and $E\{BP\}=(m p_{d,0})^{-1}$, the mean length of a cycle (i.e. an idle period followed by a busy period) will be $[(N l)^{-1}+(m p_{d,0})^{-1}]$. If we now consider a long time interval of length T, then the number of cycles in that interval would be $T/[(Nl)^{-1}+(mp_{d,0})^{-1}]$. However, this must also be equal to the mean number of transitions from state *1* to state *0*, which would be given by D_0T. (Note that this follows from the fact that we would start a new cycle every time the system becomes idle.) Therefore, equating the two, we will have

$$D_0 = \frac{1}{\dfrac{1}{mp_{d,0}} + \dfrac{1}{Nl}} = \left[\frac{Nl}{p_{d,0} + \dfrac{Nl}{m}}\right] p_{d,0} = g\, p_{d,0} \qquad (30)$$

where g is the system's equilibrium throughput rate obtained earlier.

We now need the results from a theorem on the *visit ratio* for Markov Chains which states D_k will be proportional to $p_{d,k}$ (i.e. the transition rate from state $k+1$ to k will be proportional to the probability of state k at the departure instant). This implies that

$$\frac{D_k}{p_{d,k}} = \frac{D_j}{p_{d,j}} \qquad j,k=0,1,......,(N-1) \qquad (31)$$

Since (31) will also hold for $j=0$, we can use the result of (30) to claim that

$$D_k = \left[\frac{Nl}{p_{d,0} + \dfrac{Nl}{m}}\right] p_{d,k} = g\, p_{d,k} \qquad k=0,1,......,(N-1) \qquad (32)$$

Now consider the number of transitions from state k to $k+1$ for $k= 0,1,......,(N-1)$. Measured over a long time interval T, this will be $(N-k)lp_kT$, since the average arrival rate in state k will be $(N-k)l$. This gives

$$U_k = (N-k)\, l p_k \qquad k=0,1,......,(N-1) \qquad (33)$$

where it should be noted that p_k is the probability of finding the system in state k at an arbitrary time instant.

Note that since the system is being considered at equilibrium, the up transition rate from state k (i.e. U_k) must equal the down transition rate to state k (i.e. D_k). Therefore, equating (32) and (33) for the same value of k will give

$$p_k = \frac{g\, p_{d,k}}{(N-k)\, l} \qquad k=0,1,......,(N\text{-}1) \qquad (34)$$

as the required state probabilities at an arbitrary time instant for *k=0,1,.. ....,(N-1)*. To find the probability of the remaining state p_N (i.e. the probability that all sources are blocked at the queue), we can use the normalisation result. This will give

$$p_N = 1 - \sum_{k=0}^{N-1} p_k \qquad (35)$$

Since we had earlier found the state probabilities $p_{d,k}$ *k=0,1,......,(N-1)* at the departure instants and the throughput g , (34) and (35) may now be used to find the state probabilities of the system at an arbitrary instant of time.