4.6.1 The Geo/G/1 Queue

We consider here the case where only one arrival, at the most, can occur in a slot and the service duration of a job is an integral (random) multiple of the slot duration. The queue has only one server and has an infinite number of waiting positions. We consider the analysis of this queue separately for the late arrival and early arrival models.

Late Arrival Model

Let n_i be the number of jobs in the queue immediately after the service completion of the i^{th} job. Let a_i be the number of jobs arriving during the service time of the i^{th} job. This leads to the following for the late arrival model

$$\begin{aligned} n_{i+1} &= a_{i+1} & n_i &= 0 \\ &= n_i + a_{i+1} - 1 & n_i &\geq 1 \end{aligned} \qquad (4.91)$$

The random variables a_i $i=1,2,......$ are independent and identically distributed random variables with the generating function *A(z)* and mean r. The state of the queue n_i will then form a homogenous Markov chain. Under conditions of equilibrium, the steady-state distribution p_k of this Markov chain will be

$$p_k = \lim_{i\to\infty} P\{n_i = k\} \qquad k = 0,1,2,....... \qquad (4.92)$$

with the generating function *P(z)* found using

$$P(z) = \sum_{k=0}^{\infty} p_k z^k \qquad (4.93)$$

Note that Eq. (4.91) is in the same form as Eq. (3.11), which was the equivalent state expression derived for the M/G/1 queue. It can therefore be solved in the same fashion as Section 3.2 to find *P(z)* as

$$P(z) = \frac{(1-r)(1-z)A(z)}{A(z)-z} \qquad (4.94)$$

with

$$p_0 = 1 - r \tag{4.95}$$

As in Section 3.2, *A(z)* is the generating function for the number of jobs arriving during a service interval. Using Eq. (4.87), it can be shown that for the Geo/G/1 queue we will get

$$A(z) = B(1 - l + lz) \tag{4.96}$$

This leads to

$$P(z) = \frac{(1 - r)(1 - z)B(1 - l + lz)}{B(1 - l + lz) - z} \tag{4.97}$$

as the generating function for the number in the system at the instant of service completion. In this case also, it can be shown as in Section 3.2, that the generating function of the queue states will be given by *P(z)* even at the time instant immediately after each slot boundary. (This will actually also be true at any arbitrary time instant on the continuous time axis.). Therefore, we can use *P(z)*, as given in Eq. (4.97), to find the equilibrium state distribution of the queue (i.e. the probability distribution of the number in the system) immediately after each slot boundary.

The discrete-time equivalent of the PASTA property may also be proved. This property is sometimes referred to as BASTA (i.e. Bernoulli arrivals see time averages) or GASTA (geometric arrivals see time averages). Using this property, we can then also claim that the state distribution as given by *P(z)* of Eq. (4.97) will also hold for the number in the system as seen by an arriving customer.

Using *P(z)* from Eq. (4.97), the mean number *N* in the system may be obtained to be

$$N = \frac{l^2 b^{(2)} - lr}{2(1 - r)} + r \tag{4.98}$$

The discrete-time version of Little's result also holds and is stated as

$$N = lW \tag{4.99}$$

where *N* is the mean number in the system as given by Eq. (4.98), l is the mean number of arrivals in one slot and *W* is the mean time spent in system (in units of slots) by an arriving customer. This may then be used to obtain the mean time spent in system. Once W has been obtained, the mean time W_q spent waiting for service (in number of slots) will be given as

$$W_q = W - b = \frac{l b^{(2)} - r}{2(1 - r)} \qquad (4.100)$$

Little's result ($N_q = l W_q$) may then be applied once again to find the mean number waiting in queue in the system, prior to service.

Other results, similar to the ones derived in Chapter 3 for the M/G/1 queue may also be similarly derived for this Geo/G/1 queue. For example, consider a FCFS Geo/G/1 queue. In this case, the number left behind in the system by a departing customer will be the same as the number arriving to the system, while that customer was in service. Let $G_W(z)$ be the generating function for the number of slots for which an arriving job stays in the system. We can then show that

$$P(z) = G_W(1 - l + lz) \qquad (4.101)$$

This will allow us to find the required generating function $G_W(z)$ as

$$G_W(z) = \frac{(1 - r)(1 - z)B(z)}{(1 - z) - l(1 - B(z))} \qquad (4.102)$$

Since the waiting time in queue for a job will be independent of its service time (each one of them measured in units of slots), we can write

$$G_W(z) = G_{Wq}(z)B(z) \qquad (4.103)$$

where $G_{Wq}(z)$ is the generating function of the number of slots spent waiting in queue by a job before its service actually starts. Therefore

$$G_{Wq}(z) = \frac{(1 - r)(1 - z)}{(1 - z) - l(1 - B(z))} \qquad (4.104)$$

Early Arrival Model

In this case, let the state n_i be the system state after the completion of the i^{th} service and before the possible arrival point. Let the number of arrivals a_i in a service duration be defined as for the late arrival case. At equilibrium, the generating functions for these are $P(z)$ and $A(z)$, respectively. In addition we define $\tilde{a}_i$ as the “number of jobs arriving in the service time of the i^{th} job minus one slot” with equilibrium values of its probability distribution and generating function given by

$$P\{\tilde{a}=k\}=\sum_{j=k+1}^{\infty}\binom{j-1}{k} l^{k}(1-l)^{j-1-k}b(j) \qquad k=0,1,2,.......$$

$$\begin{aligned}\tilde{A}(z) &= \sum_{k=0}^{\infty} z^{k} P\{\tilde{a}=k\} \\ &= \sum_{k=0}^{\infty} z^{k} \sum_{j=k+1}^{\infty}\binom{j-1}{k} l^{k}(1-l)^{j-1-k}b(j) \\ &= \sum_{j=1}^{\infty} b(j)\sum_{k=0}^{j-1}\binom{j-1}{k}(lz)^{k}(1-l)^{j-1-k} \\ &= \sum_{j=1}^{\infty} b(j)(1-l+lz)^{j-1} \\ &= \frac{B(1-l+lz)}{1-l+lz}\end{aligned} \qquad (4.105)$$

The state transition equation (corresponding to Eq. (4.91) for the late arrival case) may then be written as

$$\begin{aligned} n_{i+1} &= \tilde{a}_{i+1} & n_i &= 0 \\ &= n_i + a_{i+1} - 1 & n_i &\geq 1 \end{aligned} \qquad (4.106)$$

This leads to the following generating function for the system states at equilibrium

$$P(z)=\frac{(1-r)[A(z)-z\tilde{A}(z)]}{(1-l)[A(z)-z]} \qquad (4.107)$$

Simplifying Eq. (4.107) by substitution using Eqs. (4.96) and (4.105), we get the final expression for the generating function *P(z)* as

$$P(z)=\frac{(1-r)(1-z)B(1-l+lz)}{(1-l+lz)[B(1-l+lz)-z]} \qquad (4.108)$$

Comparing this with the result of Eq. (4.97) for the late arrival model, we see that the expressions differ merely by a scaling factor of $(1-l+lz)$. The queue size in the early arrival model is lower than in the late arrival model. This happens because in the early arrival model, we are observing the queue size before the possible arrival point in a slot. Note that since the slot in which the job arrives is counted in the early arrival model, when we calculate the time spent by it in the queue, the expression for $G_W(z)$ given in Eq. (4.101) for the late arrival model will get modified to be

$$P(z) = \frac{1}{(1-l+lz)} G_W(1-l+lz) \tag{4.109}$$

Using this, the generating function $G_W(z)$ for the number of slots spent in the system by a job (waiting and in service) may be obtained as

$$G_W(z) = \frac{(1-r)(1-z)B(z)}{(1-z)-l(1-B(z))} \tag{4.110}$$

Note that this is the same as the result of Eq. (4.102) for the late arrival model. This is expected since the number of slots that are spent in the system by a job will be the same in both cases. Other results, such as those obtained for the late arrival model, may also be similarly obtained.