4.3 M/G/1 Queue with Exceptional First Service

We consider here a special kind of M/G/1 queue where the first customer to get serviced when a busy period starts gets a different kind of service (i.e. with a different distribution of the service time) than the other customers served during the busy period. This kind of situation may arise in many practical situations as the server starting work after a period of idling may work slower (or faster) than the way it otherwise would. In a computer network, this may also arise because the first packet in a sequence may require special processing for route establishment and buffer set-up and will therefore require a different kind of service than the subsequent packets.

Once again, it is possible to analyse this using either the residual lifetime approach or the method of imbedded Markov chain. In the following, the latter method will be described based on the choice of the customer departure instants $\{t_i\}$ as the imbedded time points and the system state represented by the number in the system $\{n_i\}$ as seen by a departing customer. Note that both PASTA and Kleinrock's Result will still be applicable to this queue allowing us to generalize the distributions obtained at the departure instants to both the arrival instants and the ergodic system averages.

Let $b(t)$ (with L.T. $L_B(s)$) be the probability density function of the normal service time and let $b^*(t)$ (with L.T. $L_{B^*}(s)$) be the probability density function of the service time of the first customer being served in a busy period, i.e. that of the exceptional first service time. In the following we will use superscript * to denote quantities relevant to the exceptional first service. For this system, we can then show that

$$n_{i+1} = n_i - U(n_i) + a^*_{i+1} + (a_{i+1} - a^*_{i+1})U(n_i) \tag{4.16}$$

where a^*_{i+1} is the number of arrivals in the first (exceptional) service time of the busy period and a_{i+1} is the number of arrivals in the normal service times. The generating functions for these will be denoted by $A^*(z)$ and $A(z)$, respectively and may be found using $A^*(z)=L_{B^*}(l-lz)$ and $A(z)=L_B(l-lz)$. Taking expectations of the RHS and LHS of Eq. (4.16) at equilibrium (i.e. by dropping the subscript i), we get that

$$1 - p_0 = \frac{\overline{a^*}}{1-\overline{a}+\overline{a^*}} \quad \text{with} \quad \overline{a} = l\overline{X} \quad \text{and} \quad \overline{a^*} = l\,\overline{X^*} \tag{4.17}$$

Therefore,

$$p_0 = \frac{1 - l\overline{X}}{1 - l\overline{X} + l\,\overline{X^*}} \tag{4.18}$$

Using Eq. (4.16), we can also find that

$$P(z) = \frac{p_0[A(z) - zA^*(z)]}{A(z) - z} \tag{4.19}$$

where $A^*(z)=L_{B^*}(l-lz)$ and $A(z)=L_B(l-lz)$.

It should be noted that the above results will hold even if the queue is not FCFS in nature. If a FCFS service discipline is assumed, we can find the distribution of the total time T spent in system by an arriving customer by identifying as before that $P(z)=L_T(l-lz)$. This would be the L.T. of the probability density function of T and will be given by

$$L_T(s) = \frac{p_0[lL_B(s) - (l - s)L_{B^*}(s)]}{lL_B(s) + s - l} \tag{4.20}$$

This needs to be inverted to find the actual distribution (probability density function) of T, in case that is required. The mean of this will be the mean time W spent by a customer in the system. This and other moments may be found directly from $L_T(s)$ using the moment generating properties of Laplace Transform. Using this approach, the following W may be obtained.

$$W = \frac{\overline{X^*}}{1 - l\overline{X} + l\,\overline{X^*}} + \frac{l\,\overline{X^2}}{2(1 - l\overline{X})} + \frac{l(\overline{X^{*2}} - \overline{X^2})}{2(1 - l\overline{X} + l\,\overline{X^*})} \tag{4.21}$$

The overall mean service time will be $[(1-p_0)E\{X\}+p_0E\{X^*\}]$ taking into account the fact that the first customer in the busy period (probability p_0) will encounter a mean service time of $E\{X^*\}$, whereas the other customers (probability $1-p_0$) will have a mean service time of $E\{X\}$. Using this, the mean queueing delay W_q may be found to be

$$W_q = \frac{l\,\overline{X^2}}{2(1 - l\overline{X})} + \frac{l(\overline{X^{*2}} - \overline{X^2})}{2(1 - l\overline{X} + l\,\overline{X^*})} \tag{4.22}$$

Knowing W and W_q, the mean N number in the system and the mean number N_q waiting in the queue may be found using Little's Result. It should be noted that even though the probability density function (actually the L.T. of the probability density function) result of Eq. (4.20) holds only for the FCFS queue, the mean results of Eqs. (4.21) and (4.22) and the results for N and N_q will hold for any service discipline, i.e. FCFS, LCFS or SIRO.

The above results may also be obtained by using a residual life based analytical approach. This is being left as an exercise for the user (see Problem 1). Another typical way of describing a queue like the one considered in this section is to state that the first service in a busy period requires an additional D seconds of service (D random with its own moments and distribution) over and above the normal service time X. The analysis of such a system may be done in the same manner as above.