1 The $M/M/1$ system with priorities

In this section we consider an $M/M/1$ system processing different types of jobs. To keep it simple we suppose that there are two types only, type 1 and 2 say, but the analysis can easily be extended the situation with more types of jobs. Type 1 and type 2 jobs arrive according to independent Poisson processes with rate λ_1, and λ_2 respectively. The processing times of all jobs are exponentially distributed with the same mean $1/\mu$. We assume that

$$\rho_1 + \rho_2 < 1,$$

where $\rho_i = \lambda_i/\mu$, i.e. the occupation rate due to type i jobs. Type 1 jobs are treated with priority over type 2 jobs. In the following subsections we will consider two priority rules, preemptive-resume priority and non-preemptive priority.

1.1 Preemptive-resume priority

In the preemptive resume priority rule, type 1 jobs have absolute priority over type 2 jobs. Absolute priority means that when a type 2 job is being processed and a type 1 job arrives, the type 2 processing is interrupted and the machine proceeds with the type 1 job. Once there are no more type 1 jobs in the system, the machine resumes the processing of the type 2 job at the point where it was interrupted.

Let the random variable L_i denote the number of type i jobs in the system and S_i the throughput time of a type i job. Below we will determine $E(L_i)$ and $E(S_i)$ for $i = 1, 2$.

For type 1 jobs the type 2 jobs do not exist. Hence we immediately have

$$E(S_1) = \frac{1/\mu}{1-\rho_1}, \qquad E(L_1) = \frac{\rho_1}{1-\rho_1}\,. \tag{1}$$

Since the (residual) processing times of all jobs are exponentially distributed with the same mean, the total number of jobs in the system does not depend on the order in which the jobs are served. So this number is the same as in the system where all jobs are served in order of arrival. Hence,

$$E(L_1) + E(L_2) = \frac{\rho_1 + \rho_2}{1-\rho_1-\rho_2}, \tag{2}$$

and thus, inserting (1),

$$E(L_2) = \frac{\rho_1 + \rho_2}{1-\rho_1-\rho_2} - \frac{\rho_1}{1-\rho_1} = \frac{\rho_2}{(1-\rho_1)(1-\rho_1-\rho_2)}\,,$$

and applying Little's law,

$$E(S_2) = \frac{E(L_2)}{\lambda_2} = \frac{1/\mu}{(1-\rho_1)(1-\rho_1-\rho_2)}\,.$$

Example 1.1 For $\lambda_1 = 0.2$, $\lambda_2 = 0.6$ and $\mu = 1$, we find in case all jobs are treated in order of arrival,

$$E(S) = \frac{1}{1 - 0.8} = 5,$$

and in case type 1 jobs have absolute priority over type 2 jobs,

$$E(S_1) = \frac{1}{1 - 0.2} = 1.25, \qquad E(S_2) = \frac{1}{(1 - 0.2)(1 - 0.8)} = 6.25.$$

1.2 Non-preemptive priority

We now consider the situation that type 1 jobs have nearly absolute priority over type 2 jobs. The difference with the previous rule is that type 1 jobs are not allowed to interrupt the processing of a type 2 jobs. This priority rule is therefore called *non-preemptive*.

For the mean throughput time of type 1 jobs we find

$$E(S_1) = E(L_1)\frac{1}{\mu} + \frac{1}{\mu} + \rho_2\frac{1}{\mu}.$$

The last term reflects that when an arriving type 1 job finds a type 2 job being processed, he has to wait until the processing of this type 2 job has been completed. According to PASTA the probability that he finds a type 2 job being processed is equal to the fraction of time the machine spends on type 2 jobs, which is ρ_2. Together with Little's law,

$$E(L_1) = \lambda_1 E(S_1),$$

we obtain

$$E(S_1) = \frac{(1 + \rho_2)/\mu}{1 - \rho_1}, \qquad E(L_1) = \frac{(1 + \rho_2)\rho_1}{1 - \rho_1}\,.$$

For type 2 jobs it follows from (2) that

$$E(L_2) = \frac{(1 - \rho_1(1 - \rho_1 - \rho_2))\rho_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}\,,$$

and applying Little's law,

$$E(S_2) = \frac{(1 - \rho_1(1 - \rho_1 - \rho_2))/\mu}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}\,.$$

Example 1.2 For $\lambda_1 = 0.2$, $\lambda_2 = 0.6$ and $\mu = 1$, we get

$$E(S_1) = \frac{1 + 0.6}{1 - 0.2} = 2, \qquad E(S_2) = \frac{1 - 0.2(1 - 0.8)}{(1 - 0.2)(1 - 0.8)} = 6.$$