10 Variations of the $M/G/1$ model

In this chapter we treat some variations of the $M/G/1$ model and we demonstrate that the mean value technique is a powerful technique to evaluate mean performance characteristics in these models.

10.1 Machine with setup times

Consider a single machine where jobs are being processed in order of arrival and suppose that it is expensive to keep the machine in operation while there are no jobs. Therefore the machine is turned off as soon as the system is empty. When a new job arrives the machine is turned on again, but it takes some setup time till the machine is ready for processing. So turning off the machine leads to longer production leadtimes. But how much longer? This will be investigated for some simple models in the following sections.

10.1.1 Exponential processing and setup times

Suppose that the jobs arrive according to a Poisson stream with rate λ and that the processing times are exponentially distributed with mean $1/\mu$. For stability we have to require that $\rho = \lambda/\mu < 1$. The setup time of the machine is also exponentially distributed with mean $1/\theta$. We now wish to determine the mean production lead time $E(S)$ and the mean number of jobs in the system. These means can be determined by using the mean value technique (see also section 5.1).

To derive an equation for the mean production lead time, i.e. *the arrival relation*, we evaluate what is seen by an arriving job. We know that the mean number of jobs in the system found by an arriving job is equal to $E(L)$ and each of them (also the one being processed) has an exponential (residual) processing time with mean $1/\mu$. With probability $1-\rho$ the machine is not in operation on arrival, in which case the job also has to wait for the (residual) setup phase with mean $1/\theta$. Further the job has to wait for its own processing time. Hence

$$E(S) = (1-\rho)\frac{1}{\theta} + E(L)\frac{1}{\mu} + \frac{1}{\mu}$$

and together with Little's law,

$$E(L) = \lambda E(S)$$

we immediately find

$$E(S) = \frac{1/\mu}{1-\rho} + \frac{1}{\theta}\,.$$

So the mean production lead time is equal to the one in the system where the machine is always on, plus an extra delay caused by turning off the machine when there is no work. In fact, it can be shown that the extra delay is exponentially distributed with mean $1/\theta$.

10.1.2 General processing and setup times

We now consider the model with generally distributed processing times and generally distributed setup times. The arrivals are still Poisson with rate λ. The first and second moment of the processing time are denoted by $E(B)$ and $E(B^2)$ respectively, $E(T)$ and $E(T^2)$ are the first and second moment of the setup time. For stability we require that $\rho = \lambda E(B) < 1$. Below we demonstrate that also in this more general setting the mean value technique can be used to find the mean production lead time. We first determine the mean waiting time. Then the mean production lead time is found afterwards by adding the mean processing time.

The mean waiting time $E(W)$ of a job satisfies

$$\begin{aligned} E(W) &= E(L^q)E(B) + \rho E(R_B) \\ &+ P(\text{Machine is off on arrival})E(T) \\ &+ P(\text{Machine is in setup phase on arrival})E(R_T), \end{aligned} \tag{1}$$

where $E(R_B)$ and $E(R_T)$ denote the mean residual processing and residual setup time, so (see section 9.3)

$$E(R_B) = \frac{E(B^2)}{2E(B)}, \qquad E(R_T) = \frac{E(T^2)}{2E(T)}.$$

To find the probability that on arrival the machine is off (i.e. not working *and* not in the setup phase), note that by PASTA, this probability is equal to the fraction of time that the machine is off. Since a period in which the machine is not processing jobs consists of an interarrival time followed by a setup time, we have

$$P(\text{Machine is off on arrival}) = (1-\rho)\frac{1/\lambda}{1/\lambda + E(T)}.$$

Similarly we find

$$P(\text{Machine is in setup phase on arrival}) = (1-\rho)\frac{E(T)}{1/\lambda + E(T)}.$$

Substituting these relations into (1) and using Little's law stating that

$$E(L^q) = \lambda E(W)$$

we finally obtain that

$$E(W) = \frac{\rho E(R_B)}{1-\rho} + \frac{1/\lambda}{1/\lambda + E(T)}E(T) + \frac{E(T)}{1/\lambda + E(T)}E(R_T).$$

Note that the first term at the right-hand side is equal to the mean waiting time in the $M/G/1$ without setup times, the other terms express the extra delay due to the setup times. Finally, the mean production lead time $E(S)$ follows by simply adding the mean processing time $E(B)$ to $E(W)$.

10.1.3 Threshold setup policy

A natural extension to the setup policy in the previous section is the one in which the machine is switched on when the number of jobs in the system reaches some threshold value, N say. This situation can be analyzed along the same lines as in the previous section. The mean waiting time now satisfies

$$\begin{aligned} E(W) &= E(L^q)E(B) + \rho E(R_B) \\ &+ \sum_{i=1}^{N} P(\text{Arriving job is number } i)\left(\frac{N-i}{\lambda} + E(T)\right) \\ &+ P(\text{Machine is in setup phase on arrival})E(R_T), \end{aligned} \tag{2}$$

The probability that an arriving job is the i-th one in a cycle can be determined as follows. A typical production cycle is displayed in figure 1.

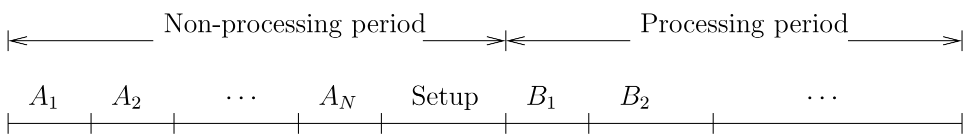

Figure 1: Production cycle in case the machine is switched on when there are N jobs

The probability that a job arrives in a non-processing period is equal to $1-\rho$. Such a period now consists of N interarrival times followed by a setup time. Hence, the probability that a job is the i-th one, given that the job arrives in a non-processing period, is equal to $1/\lambda$ divided by the mean length of a non-processing period. So

$$P(\text{Arriving job is number } i) = (1-\rho)\frac{1/\lambda}{N/\lambda + E(T)}, \qquad i = 1, \ldots, N,$$

and similarly,

$$P(\text{Machine is in setup phase on arrival}) = (1-\rho)\frac{E(T)}{N/\lambda + E(T)}.$$

Substituting these relations into (2) we obtain, together with Little's law, that

$$E(W) = \frac{\rho E(R_B)}{1-\rho} + \frac{N/\lambda}{N/\lambda + E(T)}\left(\frac{N-1}{2\lambda} + E(T)\right) + \frac{E(T)}{N/\lambda + E(T)}E(R_T).$$

10.2 Unreliable machine

In this section we consider an unreliable machine processing jobs. The machine can break down at any time, even when it is not processing jobs. What is the impact of these breakdowns on the production leadtimes? To obtain some insight in the effects of breakdowns we formulate and study some simple models in the following subsections (see also section 5.2).

10.2.1 Exponential processing and down times

Suppose that jobs arrive according to a Poisson stream with rate λ. The processing times are exponentially distributed with mean $1/\mu$. The machine is successively up and down. The up and down times of the machine are also exponentially distributed with means $1/\eta$ and $1/\theta$, respectively.

We begin with formulating the condition under which the machine can handle the amount of work offered per unit time. Let ρ_U and ρ_D denote the fraction of time the machine is up and down, respectively. So

$$\rho_U = \frac{1/\eta}{1/\eta + 1/\theta} = \frac{1}{1+\eta/\theta}, \qquad \rho_D = 1 - \rho_U = \frac{1}{1+\theta/\eta}.$$

Then we have to require that

$$\frac{\lambda}{\mu} < \rho_U. \tag{3}$$

We now proceed to derive an equation for the mean production lead time. An arriving job finds on average $E(L)$ jobs in the system and each of them has an exponential processing time with mean $1/\mu$. So in case of a perfect machine his mean sojourn time is equal to $(E(L)+1)/\mu$. However, it is not perfect. Breakdowns occur according to a Poisson process with rate η. So the mean number of breakdowns experienced by our job is equal to $\eta(E(L)+1)/\mu$, and the mean duration of each breakdown is $1/\theta$. Finally, with probability ρ_D the machine is already down on arrival, in which case our job has an extra mean delay of $1/\theta$. Summarizing we have

$$\begin{aligned} E(S) &= (E(L)+1)\frac{1}{\mu} + \eta(E(L)+1)\frac{1}{\mu}\cdot\frac{1}{\theta} + \rho_D\frac{1}{\theta} \\ &= (1+\frac{\eta}{\theta})(E(L)+1)\frac{1}{\mu} + \frac{\rho_D}{\theta} \\ &= (E(L)+1)\frac{1}{\mu\rho_U} + \frac{\rho_D}{\theta}. \end{aligned}$$

Then, with Little's law stating that

$$E(L) = \lambda E(S),$$

we immediately obtain

$$E(S) = \frac{1/(\mu\rho_U) + \rho_D/\theta}{1 - \lambda/(\mu\rho_U)}.$$

10.2.2 General processing and down times

We now consider the model with general processing times and general down times. The arrivals are still Poisson with rate λ. The first and second moment of the processing time are denoted by $E(B)$ and $E(B^2)$ respectively. The time between two breakdowns is exponentially distributed with mean $1/\eta$ and $E(D)$ and $E(D^2)$ are the first and second moment of the down time. For stability we require that (cf. (3))

$$\lambda E(B) < \frac{1}{1+\eta E(D)}.$$

Below we first determine the mean waiting time by using the mean value technique. The mean production leadtime is determined afterwards.

We start with introducing the generalized processing time, which is defined as the processing time plus the down times occurring in that processing time. Denote the generalized processing time by G. Then we have

$$G = B + \sum_{i=1}^{N(B)} D_i,$$

where $N(B)$ is the number of break-downs during the processing time B and D_i is the i-th down time. For the mean of G we get by conditioning on B and $N(B)$ that

$$\begin{aligned} E(G) &= \int_{x=0}^{\infty} \sum_{n=0}^{\infty} E(G|B=x, N(B)=n) e^{-\eta x} \frac{(\eta x)^n}{n!} f_B(x)dx \\ &= \int_{x=0}^{\infty} \sum_{n=0}^{\infty} (x + nE(D)) e^{-\eta x} \frac{(\eta x)^n}{n!} f_B(x)dx \\ &= \int_{x=0}^{\infty} \sum_{n=0}^{\infty} (x + x\eta E(D)) e^{-\eta x} \frac{(\eta x)^n}{n!} f_B(x)dx \\ &= E(B) + E(B)\eta E(D), \end{aligned}$$

and similarly,

$$E(G^2) = E(B^2)(1+\eta E(D))^2 + E(B)\eta E(D^2).$$

A typical production cycle is shown in figure 2. The non-processing period consists of cycles which start with an up time. When the machine is up two things can happen: a job arrives, in which case the machine starts to work, or the machine goes down and has to be repaired. Hence, an up time is exponentially distributed with mean $1/(\lambda+\eta)$ and it is followed by a processing period (i.e. a job arrives) with probability $\lambda/(\lambda+\eta)$ or otherwise, it is followed by a down time. During a processing period the machine works on jobs with generalized processing times (until all jobs are cleared). For the mean waiting time it holds that

$$\begin{aligned} E(W) &= E(L^q)E(G) + \rho_G E(R_G) \\ &\quad + P[\text{Arrival in a down time in a non-processing period}]E(R_D), \end{aligned} \tag{4}$$

where ρ_G is the fraction of time the machine works on generalized jobs and $E(R_G)$ and $E(R_D)$ denote the mean residual generalized processing time and the mean residual down time, so

$$\rho_G = \lambda E(G), \qquad E(R_G) = \frac{E(G^2)}{2E(G)}, \qquad E(R_D) = \frac{E(D^2)}{2E(D)}.$$

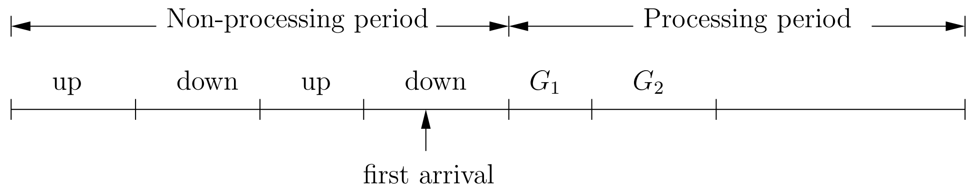

Figure 2: Production cycle of an unreliable machine

The probability that a job arriving in a non-processing period finds that the machine is down, is given by

$$\frac{E(D)\eta/(\lambda+\eta)}{1/(\lambda+\eta)+E(D)\eta/(\lambda+\eta)} = \frac{E(D)\eta}{1+E(D)\eta},$$

Hence, since a job arrives with probability $1-\rho_G$ in a non-processing period, we have

$$P(\text{Arrival in a down time in a non-processing period}) = (1-\rho_G)\frac{E(D)\eta}{1+E(D)\eta}.$$

Substitution of this relation into (4) and using Little's law yields

$$E(W) = \frac{\rho_G E(R_G)}{1-\rho_G} + \frac{E(D)\eta}{1+E(D)\eta}E(R_D), \tag{5}$$

and the mean production lead time finally follows from

$$E(S) = E(W) + E(G).$$

Remark 10.1 The first term at the right-hand side of (5) is precisely the the mean waiting time in case of operational dependent failures (cf. remark 5.3);so the second term is the extra delay due to the fact that the machine may also be down on arrival of the first job in a busy period.

10.3 $M/G/1$ queue with an exceptional first customer in a busy period

The life of a server in an $M/G/1$ queue is an alternating sequence of periods during which no work is done, the so-called idle periods, and periods during which the server helps customers, the so-called busy periods. In some applications the service time of the first customer in a busy period is different from the service times of the other customers served in the busy period. For example, consider the setup problem in section 10.1 again. This problem may, alternatively, be formulated as an $M/G/1$ queue in which the first job in a busy period has an exceptional service time, namely the setup time plus his actual processing time. The problem in section 10.2 can also be described in this way. Here we can take the generalized processing times as the service times. However, on arrival of the first job in a busy period the machine can be down. Then the (generalized) servicing of that job cannot immediately start, but one has to wait till the machine is repaired. This repair time can be included in the service time of the first job. In this case, however, the determination of the distribution, or the moments of the service time of the first job is more difficult than in the setup problem. Below we show how, in the general setting of an $M/G/1$ with an exceptional first customer, the mean sojourn time can be found by using the mean-value approach.

Let B_f and R_f denote the service time and residual service time, respectively, of the first customers in a busy period. For the mean waiting time we have

$$E(W) = E(L^q)E(B) + \rho_f E(R_f) + \rho E(R), \tag{6}$$

where ρ_f is the fraction of time the server works on first customers, and ρ is the fraction of time the server works on other (ordinary) customers. Together with Little's law we then obtain from (6) that

$$E(W) = \frac{\rho_f E(R_f) + \rho E(R)}{1 - \lambda E(B)},$$

and for the mean sojourn time we get

$$E(S) = E(W) + (1 - \rho_f - \rho)E(B_f) + (\rho_f + \rho)E(B).$$

It remains to determine ρ_f and ρ. The probability that an arriving customer is the first one in a busy cycle is given by $1 - \rho_f - \rho$. Hence, the number of first customers arriving per unit of time is equal to $\lambda(1 - \rho_f - \rho)$. Thus ρ_f and ρ satisfy

$$\begin{aligned} \rho_f &= \lambda(1 - \rho_f - \rho)E(B_f), \\ \rho &= \lambda(\rho_f + \rho)E(B), \end{aligned}$$

from which it follows that

$$\rho_f = \frac{\lambda E(B_f)(1 - \lambda E(B))}{1 + \lambda E(B_f) - \lambda E(B)}, \qquad \rho = \frac{\lambda E(B_f)\lambda E(B)}{1 + \lambda E(B_f) - \lambda E(B)}.$$

10.4 $M/G/1$ queue with group arrivals

In this section we consider the $M/G/1$ queue where customers do not arrive one by one, but in groups. These groups arrive according to a Poisson process with rate λ. The group size is denoted by the random variable G with probability distribution

$$g_k = P(G = k), \qquad k = 0, 1, 2, \ldots$$

Note that we also admit zero-size groups to arrive. Our interest lies in the mean waiting time of a customer, for which we can write down the following equation.

$$E(W) = E(L^q)E(B) + \rho E(R) + \sum_{k=1}^{\infty} r_k (k-1) E(B)\,, \tag{7}$$

where ρ is the server utilization, so

$$\rho = \lambda E(G)E(B),$$

and r_k is the probability that our customer is the kth customer served in his group. The first two terms at the right-hand side of (7) correspond to the mean waiting time of the whole group. The last one indicates the mean waiting time due to the servicing of members in his own group.

To find r_k we first determine the probability h_n that our customer is a member of a group of size n. Since it is more likely that our customer belongs to a large group than to a small one, it follows that h_n is proportional to the group size n as well as the frequency of such groups. Thus we can write

$$h_n = Cng_n,$$

where C is a constant to normalize this distribution. So

$$C^{-1} = \sum_{n=1}^{\infty} ng_n = E(G).$$

Hence

$$h_n = \frac{ng_n}{E(G)}, \qquad n = 1, 2, \ldots$$

Given that our customer is a member of a group of size n, he will be with probability $1/n$ the kth customer in his group going into service (of course, $n \geq k$). So we obtain

$$r_k = \sum_{n=k}^{\infty} h_n \cdot \frac{1}{n} = \frac{1}{E(G)} \sum_{n=k}^{\infty} g_n,$$

and for the last term in (7) it immediately follows that

$$\begin{aligned}
\sum_{k=1}^{\infty} r_k(k-1)E(B) &= \frac{1}{E(G)} \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} g_n(k-1)E(B) \\
&= \frac{1}{E(G)} \sum_{n=1}^{\infty} \sum_{k=1}^{n} g_n(k-1)E(B) \\
&= \frac{1}{E(G)} \sum_{n=1}^{\infty} \frac{1}{2} n(n-1) g_n E(B) \\
&= \frac{E(G^2) - E(G)}{2E(G)} E(B). \qquad (8)
\end{aligned}$$

From (7) and (8) and Little's law stating that

$$E(L^q) = \lambda E(G)E(W),$$

we finally obtain

$$E(W) = \frac{\rho E(R)}{1-\rho} + \frac{(E(G^2) - E(G))E(B)}{2E(G)(1-\rho)}.$$

The first term at the right-hand side is equal to the mean waiting time in the system where customers arrive one by one according to a Poisson process with rate $\lambda E(G)$. Clearly, the second term indicates the extra mean delay due to clustering of arrivals.

Example 10.2 *(Uniform group sizes)*
In case the group size is uniformly distributed over $1, 2, \ldots, n$, so

$$g_k = \frac{1}{n}, \qquad k = 1, \ldots, n,$$

we find

$$E(G) = \sum_{k=1}^{n} \frac{k}{n} = \frac{n+1}{2}, \qquad E(G^2 - G) = \sum_{k=1}^{n} \frac{k^2 - k}{n} = \frac{(n-1)(n+1)}{3}.$$

Hence

$$E(W) = \frac{\rho E(R)}{1-\rho} + \frac{(n-1)E(B)}{3(1-\rho)}.$$