

We can interpret this to mean that 44% of the packets enter the network with no delay at the leaky bucket. Similarly, the data buffer is full whenever the $\{X(t), t \geq 0\}$ process is in state $M + D = 24$. Hence the long-run fraction of the time the data buffer is full is given by

$$p_{24} = \frac{1}{25} = .04.$$

This means that 4% of the packets will be dropped due to buffer overflow. ■

Example 4.27. (Manufacturing). Consider the manufacturing setup of Example 4.21. Compute the long-run fraction of the time the machine is off.

Let $X(t)$ and $Y(t)$ be as in Example 4.21. We see that $\{(X(t), Y(t)), t \geq 0\}$ is an irreducible CTMC with state space $\{1, 2, \dots, 6\}$ and rate matrix as given in (4.41). Hence it has a unique limiting distribution that is also its occupancy distribution. The machine is off in states $5 = (4, 0)$ and $6 = (3, 0)$. Thus the long-run fraction of the time that the machine is off is given by $p_5 + p_6$. The balance equations and the normalizing equation can be solved numerically to obtain

$$\begin{aligned} p_1 &= .1585, \quad p_2 = .1902, \quad p_3 = .2282; \\ p_4 &= .1245, \quad p_5 = .1493, \quad p_6 = .1493. \end{aligned}$$

Hence the required answer is $.1493 + .1493 = .2986$. Thus the machine is turned off about 30% of the time! ■

4.7 Cost Models

Following the development in Chapter 2, we now study cost models associated with CTMCs. Recall Example 1.4(a), where we were interested in the expected total cost of operating a machine for T units of time and also in the long-run cost per unit time of operating the machine. In this section, we shall develop methods of computing such quantities.

We assume the following cost model. Let $\{X(t), t \geq 0\}$ be a CTMC with state space $\{1, 2, \dots, N\}$ and rate matrix R . Whenever the CTMC is in state i , it incurs costs continuously at rate $c(i)$, $1 \leq i \leq N$.

4.7.1 Expected Total Cost

In this subsection, we study the ETC, the *expected total cost* up to a finite time T , called the horizon. Note that the cost rate at time t is $c(X(t))$. Hence the total cost up to time T is given by

$$\int_0^T c(X(t)) dt.$$

The expected total cost up to T , starting from state i , is given by

$$g(i, T) = \mathbb{E} \left(\int_0^T c(X(t)) dt \middle| X(0) = i \right), \quad 1 \leq i \leq N. \quad (4.53)$$

The next theorem gives a method of computing these costs. First, it is convenient to introduce the following vector notation:

$$c = \begin{bmatrix} c(1) \\ c(2) \\ \vdots \\ c(N-1) \\ c(N) \end{bmatrix},$$

and

$$g(T) = \begin{bmatrix} g(1, T) \\ g(2, T) \\ \vdots \\ g(N-1, T) \\ g(N, T) \end{bmatrix}.$$

Theorem 4.9. (Expected Total Cost). *Let $M(T)=[m_{i,j}(T)]$ be the occupancy matrix as given in Theorem 4.5. Then*

$$g(T) = M(T)c. \quad (4.54)$$

Proof. We have

$$\begin{aligned} g(i, T) &= \mathbb{E} \left(\int_0^T c(X(t)) dt \middle| X(0) = i \right) \\ &= \int_0^T \mathbb{E}(c(X(t)) | X(0) = i) dt \\ &= \int_0^T \sum_{j=1}^N c(j) \mathbb{P}(X(t) = j | X(0) = i) dt \\ &= \int_0^T \sum_{j=1}^N c(j) p_{i,j}(t) dt \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N c(j) \int_0^T p_{i,j}(t) dt \\
&= \sum_{j=1}^N c(j) m_{i,j}(T),
\end{aligned}$$

where the last equation follows from Theorem 4.4. In matrix form, this yields Equation (4.54). ■

We illustrate this with several examples.

Example 4.28. (Two-State Machine). Consider the two-state machine as described in Example 4.3. Suppose the machine produces revenue at rate A per day when it is up and the repair costs B per day of downtime. What is the expected net income from the machine up to time T (days) if the machine started in the up state?

Let $\{X(t), t \geq 0\}$ be a CTMC on $\{0, 1\}$ as described in Example 4.3. (Assume that the parameters λ and μ are per day.) Since we are interested in revenues, we set $c(0)$, the revenue rate in state 0, to be $-B$ per day of downtime and $c(1) = A$ per day of uptime. From Example 4.18, we get

$$p_{1,1}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t}$$

and

$$p_{1,0}(t) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t}.$$

Using Theorem 4.4, we get

$$\begin{aligned}
m_{1,1}(T) &= \int_0^T p_{1,1}(t) dt \\
&= \int_0^T \left(\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} \right) dt \\
&= \frac{\lambda T}{\lambda + \mu} + \frac{\mu}{(\lambda + \mu)^2} (1 - e^{-(\lambda + \mu)T}).
\end{aligned}$$

Similarly,

$$m_{1,0}(T) = \frac{\mu T}{\lambda + \mu} - \frac{\mu}{(\lambda + \mu)^2} (1 - e^{-(\lambda + \mu)T}).$$

The net income up to T can then be obtained from Theorem 4.9 as

$$\begin{aligned}
g(1, T) &= m_{1,0}(T)c(0) + m_{1,1}(T)c(1) \\
&= \frac{(\lambda A - \mu B)T}{\lambda + \mu} + \frac{\mu(B + A)}{(\lambda + \mu)^2} (1 - e^{-(\lambda + \mu)T}).
\end{aligned}$$

As a numerical example, consider the data from Example 4.19. Assume $A = \$2400$ per day and $B = \$480$ per day. The net revenue in the month of January ($T = 31$) is given by

$$g(1, 31) = -m_{1,0}(31)B + m_{1,1}(31)A.$$

In Example 4.19, we have computed $m_{1,0}(31) = 2.7355$ days and $m_{1,1}(31) = 28.2645$ days. Substituting in the equation above, we get

$$g(1, 31) = -2.7355 \cdot 480 + 28.2645 \cdot 2400 = \$66,521.60. \blacksquare$$

Example 4.29. (General Machine Shop). Consider the general machine shop of Example 4.16. Each machine produces revenue of \$50 per hour of uptime, and the downtime costs \$15 per hour per machine. The repair time costs an additional \$10 per hour per machine. Compute the net expected revenue during the first day, assuming all the machines are up in the beginning.

Let $X(t)$ be the number of up machines at time t . Then $\{X(t), t \geq 0\}$ is a CTMC on $\{0, 1, 2, 3, 4\}$ with the rate matrix given in (4.29). The revenue rates in various states are as follows:

$$\begin{aligned} c(0) &= 0 * 50 - 4 * 15 - 2 * 10 = -80, \\ c(1) &= 1 * 50 - 3 * 15 - 2 * 10 = -15, \\ c(2) &= 2 * 50 - 2 * 15 - 2 * 10 = 50, \\ c(3) &= 3 * 50 - 1 * 15 - 1 * 10 = 125, \\ c(4) &= 4 * 50 - 0 * 15 - 0 * 10 = 200. \end{aligned}$$

The net expected revenue vector is given by

$$\begin{aligned} g(24) &= M(24) * c \\ &= \begin{bmatrix} 1.0143 & 1.0318 & 1.1496 & 3.6934 & 17.1108 \\ 0.0143 & 1.0319 & 1.1538 & 3.7931 & 18.0069 \\ 0.0004 & 0.0320 & 1.1580 & 3.8940 & 18.9155 \\ 0.0001 & 0.0044 & 0.1623 & 3.9964 & 19.8369 \\ 0.0000 & 0.0023 & 0.0876 & 2.2041 & 21.7060 \end{bmatrix} \begin{bmatrix} -80 \\ -15 \\ 50 \\ 125 \\ 200 \end{bmatrix} \\ &= \begin{bmatrix} 3844.69 \\ 4116.57 \\ 4327.23 \\ 4474.96 \\ 4621.05 \end{bmatrix}. \end{aligned}$$

The required answer is given by $g(4, 24) = \$4621.05. \blacksquare$

4.7.2 Long-Run Cost Rates

In Example 4.28, suppose the downtime cost B is given. We want to know how much the revenue rate during uptime should be so that it is economically profitable to operate the machine. If we go by total cost, the answer will depend on the planning horizon T and also the initial state of the machine. An alternative is to compute the long-run net revenue per unit time for this machine and insist that this be positive for profitability. This answer will not depend on T , and, as we shall see, not even on the initial state of the machine. Hence, computing such long-run cost rates or revenue rates is very useful. We shall show how to compute these quantities in this subsection.

First, define the long-run cost rate as

$$g(i) = \lim_{T \rightarrow \infty} \frac{g(i, T)}{T}. \quad (4.55)$$

The next theorem shows an easy way to compute $g(i)$.

Theorem 4.10. (Long-Run Cost Rate). *Suppose $\{X(t), t \geq 0\}$ is an irreducible CTMC with limiting distribution $p = [p_1, p_2, \dots, p_N]$. Then*

$$g = g(i) = \sum_{j=1}^N p_j c(j), \quad 1 \leq i \leq N.$$

Proof. We have

$$\begin{aligned} g(i) &= \lim_{T \rightarrow \infty} \frac{g(i, T)}{T} \\ &= \lim_{T \rightarrow \infty} \frac{\sum_{j=1}^N m_{i,j}(T) c(j)}{T} \\ &\quad \text{(from Theorem 4.9)} \\ &= \sum_{j=1}^N c(j) \lim_{T \rightarrow \infty} \frac{m_{i,j}(T)}{T} \\ &= \sum_{j=1}^N c(j) p_j, \end{aligned}$$

where the last equality follows from Theorem 4.8. This proves the theorem. ■

Note that the theorem is intuitively obvious: the CTMC incurs a cost at rate $c(j)$ per unit time it spends in state j . It spends p_j fraction of the time in state j , regardless of the starting state. Hence it must incur costs at rate $\sum_{j=1}^N p_j c(j)$ in the long run. We illustrate this result with several examples.

Example 4.30. (Two-State Machine). Consider the two-state machine of Example 4.28. Suppose the downtime cost of the machine is B per unit time. What is the minimum revenue rate A during the uptime needed to break even in the long run?

From Example 4.22, the machine is up for $p_1 = \lambda/(\lambda + \mu)$ fraction of the time in the long run and down for $p_0 = \mu/(\lambda + \mu)$ fraction of the time in the long run. Using $c(0) = -B$ and $c(1) = A$ in Theorem 4.10, we obtain as the long-run net revenue per unit time

$$g = \frac{\lambda A - \mu B}{\lambda + \mu}.$$

For breakeven, we must have

$$g \geq 0;$$

i.e.,

$$\frac{A}{\mu} \geq \frac{B}{\lambda}.$$

Note that this condition makes sense: A/μ is the expected revenue during one up-time, and B/λ is the expected cost of one downtime. If the former is greater than the latter, the machine is profitable to operate; otherwise it is not! ■

Example 4.31. (Telephone Switch). Consider the CTMC model of the telephone switch described in Example 4.9. The switch has the capacity to handle at most six calls simultaneously. The call arrival rate is four calls per minute, and the arrival process is Poisson. The average duration of a call is 2 minutes, and the call durations are iid exponential random variables. Compute the expected revenue per minute if each caller is charged 10 cents per minute of the call duration. What is the rate at which revenue is lost due to the switch being full?

Let $X(t)$ be the number of calls carried by the switch at time t . Then $\{X(t), t \geq 0\}$ is a birth and death process with birth rates

$$\lambda_i = 4, \quad 0 \leq i \leq 5,$$

and death rates

$$\mu_i = i/2, \quad 1 \leq i \leq 6.$$

The revenue rate in state i is $c(i) = 10i$ cents per minute. Hence the long-run rate of revenue is

$$g = \sum_{j=0}^6 p_j c(j) = 10 \sum_{j=0}^6 j p_j.$$

Note that the last sum is just the expected number of calls in the switch in steady state. Using the results of Example 4.23, we get

$$\rho_i = \frac{8^i}{i!}, \quad 0 \leq i \leq 6,$$

and

$$p_i = \frac{\rho_i}{\sum_{j=0}^6 \rho_j}, \quad 0 \leq i \leq 6.$$

Hence we get

$$g = 10 \frac{\sum_{i=0}^6 i \rho_i}{\sum_{j=0}^6 \rho_j} = 48.8198.$$

Hence the switch earns revenue at a rate of 48.82 cents per minute.

Note that calls arrive at the rate of four per minute, and the average revenue from each call is 20 cents (duration is 2 minutes). Hence, if the switch had ample capacity so that no call was rejected, the revenue would be $4 * 20 = 80$ cents per minute. But the actual revenue is only 48.82 cents per minute. Hence revenue is lost at the rate $80 - 48.82 = 31.18$ cents per minute. ■

4.8 First-Passage Times

In this section, we study the *first-passage times* in CTMCs, following the development in the DTMCs in Section 2.7. Let $\{X(t), t \geq 0\}$ be a CTMC with state space $\{1, \dots, N\}$ and rate matrix R . The first-passage time into state N is defined to be

$$T = \min\{t \geq 0 : X(t) = N\}. \quad (4.56)$$

As in Section 2.7, we study the expected value $E(T)$. Let

$$m_i = E(T | X(0) = i). \quad (4.57)$$

Note that $m_N = 0$. The next theorem gives a method of computing m_i , $1 \leq i \leq N - 1$.

Theorem 4.11. (First-Passage Times). $\{m_i, 1 \leq i \leq N - 1\}$ satisfy

$$r_i m_i = 1 + \sum_{j=1}^{N-1} r_{i,j} m_j, \quad 1 \leq i \leq N - 1. \quad (4.58)$$

Proof. We shall prove the theorem by conditioning on the time of the first transition and the state of the CTMC after the first transition. Let $X(0) = i$, $1 \leq i \leq N - 1$, and let $X(Y) = j$, where Y is the sojourn time in the initial state. Then, $T = Y$ if $X(Y) = N$, and $T = Y +$ the first passage time into state N from state j if $j \neq N$. Using this observation, we get

$$E(T | X(0) = i, X(Y) = j) = E(Y | X(0) = i) + m_j, \quad 1 \leq i, j \leq N - 1,$$

and

$$E(T | X(0) = i, X(Y) = N) = E(Y | X(0) = i), \quad 1 \leq i \leq N - 1.$$

Now, Y is an $\text{Exp}(r_i)$ random variable if $X(0) = i$. Hence

$$\mathbb{E}(Y|X(0) = i) = \frac{1}{r_i}.$$

From the properties of CTMCs, we also have

$$\mathbb{P}(X(Y) = j|X(0) = i) = \frac{r_{i,j}}{r_i},$$

where $r_{i,i} = 0$ and $r_i = \sum_{j=1}^N r_{i,j}$. Using these facts, we get, for $1 \leq i \leq N - 1$,

$$\begin{aligned} m_i &= \mathbb{E}(T|X(0) = i) \\ &= \sum_{j=1}^N \mathbb{E}(T|X(0) = i, X(Y) = j) \mathbb{P}(X(Y) = j|X(0) = i) \\ &= \sum_{j=1}^{N-1} \left(\frac{1}{r_i} + m_j \right) \frac{r_{i,j}}{r_i} + \frac{1}{r_i} \frac{r_{i,N}}{r_i} \\ &= \frac{1}{r_i} + \sum_{j=1}^{N-1} \frac{r_{i,j}}{r_i} m_j. \end{aligned}$$

Multiplying the last equation by r_i yields (4.58). ■

We illustrate the theorem above with several examples.

Example 4.32. Let $\{X(t), t \geq 0\}$ be the CTMC in Example 4.14. Compute the expected time the CTMC takes to reach state 4 starting from state 1.

Equations (4.58) are

$$\begin{aligned} 5m_1 &= 1 + 2m_2 + 3m_3, \\ 6m_2 &= 1 + 4m_1 + 2m_3, \\ 4m_3 &= 1 + 2m_2. \end{aligned}$$

Solving simultaneously, we get

$$m_1 = 1.2727, m_2 = 1.3182, m_3 = .9091.$$

Thus it takes on average 1.2727 units of time to go from state 1 to state 4. ■

Example 4.33. (Single Server Queue). Consider the single server queue of Example 4.20. Compute the expected time until the queue becomes empty, given that it has one customer at time 0.

Let $X(t)$ be the number of customers in the queue at time t . With the parameters given in Example 4.20, $\{X(t), t \geq 0\}$ is a CTMC with state space $\{0, 1, 2, 3, 4, 5\}$

and rate matrix given by (4.40). Let T be the first-passage time to state 0, and let $m_i = E(T|X(0) = i)$. Equations (4.58) for this case are

$$\begin{aligned} m_0 &= 0, \\ 25m_i &= 1 + 10m_{i+1} + 15m_{i-1}, \quad 1 \leq i \leq 4, \\ 15m_5 &= 1 + 15m_4. \end{aligned}$$

Solving numerically, we get

$$m_1 = .1737, m_2 = .3342, m_3 = .4749, m_4 = .5860, m_5 = .6527.$$

Thus the expected time until the queue becomes empty (starting with one person) is .1737 hours, or 10.42 minutes. ■

As in Section 2.7, we now study the expected time to reach a set of states, A . Analogously to (2.62), define

$$T = \min\{t \geq 0 : X(t) \in A\}. \quad (4.59)$$

Theorem 4.11 can be easily extended to the case of the first-passage time defined above. Let $m_i(A)$ be the expected time to reach the set A starting from state $i \notin A$. Note that $m_i(A) = 0$ for $i \in A$. Following the same argument as in the proof of Theorem 4.11, we can show that

$$r_i m_i(A) = 1 + \sum_{j \notin A} r_{i,j} m_j(A), \quad i \notin A. \quad (4.60)$$

We illustrate this with an example below.

Example 4.34. (Airplane Reliability). Consider the four-engine airplane of Example 4.17. Suppose that in a testing experiment the airplane takes off with four properly functioning engines and keeps on flying until it crashes. Compute the expected time of the crash.

Let $\{X(t), t \geq 0\}$ be the CTMC of Example 4.6 with $\lambda = .005$ per hour. The airplane crashes as soon as the CTMC visits the set $\{1, 2, 3, 4, 7\}$. Using this as A in (4.60), we get $m_i = 0$ for $i \in A$ and

$$\begin{aligned} .02m_9 &= 1 + .01m_8 + .01m_6, \\ .015m_8 &= 1 + .01m_5, \\ .015m_6 &= 1 + .01m_5, \\ .01m_5 &= 1. \end{aligned}$$

Solving backward recursively, we get

$$m_5 = 100, m_6 = \frac{400}{3}, m_8 = \frac{400}{3}, m_9 = \frac{550}{3}.$$

Thus the expected time to failure is $m_9 = 550/3 = 183$ hours and 20 minutes. ■