

Queueing Problems with Solutions

Collected by:

Dr. János Sztrik

University of Debrecen,

Faculty of Informatics

2021

We can now use Little's formula and (5.3.42) to show that

$$L_q = E[N_q] = \lambda E[q] = \lambda^2 E[s^2]/2(1 - \rho). \quad (5.3.46)$$

Little's formula applied to (5.3.44) yields (5.3.23), as expected. Thus equations (5.3.23) and equations (5.3.42)–(5.3.46) together show that, if we know the first three moments of service time, we can calculate both the expected value and the standard deviation for the random variables q and w . (We can do the same for N_q and N , by Exercises 20 and 21.) However, if we know only the first and second moment of service time, we must be content with average values, only, for the random variables of interest.

In many cases knowledge of average values, only, will not enable us to make the kind of probability calculations we desire. It is especially valuable to be able to compute percentile values, such as we did for the random variables q and w in the M/M/1 queueing system. There is no such general formula for the M/G/1 queueing system but J. Martin [7] gives the estimates

$$\pi_w(90) = E[w] + 1.3\sigma_w, \quad (5.3.47)$$

and

$$\pi_w(95) = E[w] + 2\sigma_w. \quad (5.3.48)$$

(Actually Martin gives only the second estimate, (5.3.48), but the reasoning he gives to justify it yields (5.3.47), also.)

Another approach to estimating the percentile values of w and q is to calculate the C^2 values and use Table 5.3.1 or Table 5.3.2. This is equivalent to approximating the random variable of interest with a gamma random variable having the same mean and standard deviation.

Example 5.3.1 Four communication lines are connected to one central computer; each has an average transmission time per message of 2.4 seconds and operates at 80% line utilization. However, the message transmission time has a different distribution for each line. The transmission time for the

TABLE 5.3.1
Percentile Values of the Erlang- k Distribution
(Special Case of Gamma Distribution)

$r/100$	$\pi_x(r)/E[X]$									
	$k = 1/C_x^2$									
.	1	2	3	4	5	10	20	40	100	
0.90	2.30	1.94	1.77	1.67	1.60	1.42	1.30	1.21	1.13	
0.95	3.00	2.37	2.09	1.94	1.83	1.57	1.39	1.27	1.17	
0.99	4.61	3.32	2.80	2.51	2.32	1.88	1.59	1.40	1.25	

TABLE 5.3.2
Percentile Values of the Gamma Distribution

$r/100$	$\pi_X(r)/E[X]$										
	C_x^2										
1.25	1.5	1.75	2.0	2.5	3.0	4.0	6.0	8.0	10.0	20.0	
0.90	2.43	2.54	2.63	2.71	2.82	2.91	3.00	3.00	2.86	2.66	1.53
0.95	3.24	3.46	3.66	3.84	4.16	4.42	4.84	5.38	5.68	5.80	5.25
0.99	5.16	5.68	6.17	6.63	7.50	8.30	9.74	12.16	14.17	15.88	11.09

first line is hyperexponential with $\alpha_1 = 0.4$, $\alpha_2 = 0.6$, $1/\mu_1 = 4.8$ seconds, and $1/\mu_2 = 0.8$ seconds; the distribution on the second line is exponential; it is Erlang-3 on the third line; and constant on the fourth line. Find W_q and W for each line. Then estimate $\pi_w(90)$ by Martin's estimate and by using Tables 5.3.1 and 5.3.2 (for the nonexponential cases).

Solution The M/G/1 model applies for each line. For the hyperexponential service time

$$E[s] = 0.4 \times 4.8 + 0.6 \times 0.8 = 2.4 \text{ seconds.}$$

and

$$E[s^2] = 2 \times (0.4 \times 4.8^2 + 0.6 \times 0.8^2) = 19.2 \text{ seconds}^2$$

(by (3.2.59) and (3.2.60), respectively). Therefore

$$\text{Var}[s] = E[s^2] - E[s]^2 = 19.2 - 2.4^2 = 13.44,$$

and

$$C_s^2 = 13.44/2.4^2 = 2.33.$$

Using the formula

$$E[s^3] = 6 \sum_{i=1}^k \frac{\alpha_i}{\mu_i^3} \quad (\text{see Exercise 22})$$

we calculate

$$E[s^3] = 6(0.4 \times 4.8^3 + 0.6 \times 0.8^3) = 267.264.$$

No special computations are necessary for the second line because it fits the M/M/1 model. For the third line with an Erlang-3 service time distribution we can use the formulas (see Exercise 25) valid for an Erlang- k distribution:

$$E[s^2] = \frac{(k+1)E[s]^2}{k}, \quad E[s^3] = \frac{(k+1)(k+2)E[s]^3}{k^2}.$$

This gives

$$E[s^2] = 7.68, \quad C_s^2 = 1/k = 1/3, \quad \text{and} \quad E[s^3] = 30.72.$$

(The formulas for the M/E_k/1 queueing system are summarized in Table 14 of Appendix C.)

Finally, for the fourth line, we have

$$E[s^2] = E[s]^2 = 5.76, \quad C_s^2 = 0, \quad \text{and} \quad E[s^3] = E[s]^3 = 13.824.$$

(The formulas for the M/D/1 queueing system are given in Table 15 of Appendix C.) Substituting these results into the M/G/1 equations of Table 12 in Appendix C yields the number in Table 5.3.3. We illustrate the calculation of $\pi_w(90)$ by the “table” method for line 3. We have $C_w^2 = (\sigma_w/W)^2 = 0.7736$. Since this number does not appear in Table 5.3.2, we take the reciprocal $1/C_w^2$, which is 1.2927, so, by Table 5.3.1,

$$\pi_w(90)/E[w] = 2.195 \quad \text{or} \quad \pi_w(90) = 2.195 \times 8.8 = 19.32 \text{ seconds.}$$

This example dramatically demonstrates the inimical effect of “irregularity” in service time, as measured by C_s^2 . (We have all been conditioned by television ads to recognize the deleterious effects of irregularity in our personal lives.) The average waiting time in the system, W , and the 90th percentile value of w is about one and a half times as large for the hyperexponential service time as for exponential service time; exponential service time yields significantly poorer performance than Erlang-3 or constant service time.

TABLE 5.3.3
Summary of Results in Example 5.3.1

Line	Distribution (line time)	$\pi_w(90)$					
		$E[s^2]$	$E[s^3]$	W_q	W	σ_w	Martin
1	two-stage						
	hyperexponential	19.2	267.264	16.00	18.40	20.44	44.98
2	exponential	11.52	82.944	9.60	12.00	12.00	27.60
3	Erlang-3	7.68	30.720	6.40	8.80	7.74	18.87
4	constant	5.760	13.824	4.80	7.20	5.54	14.40
							14.77

Example 5.3.2 (Example 5.2.2 Revisited) We saw, in Example 5.2.2, that, if a large computer system could be modeled as an M/M/1 queueing model, then replacing this system by n M/M/1 systems, that is by n smaller computers, each with $1/n$ of the capacity and traffic, then the average time in the queue, W_q , and the average time in the system, W , would each increase n -fold. We now can show that the same holds true for an M/G/1 system, if

the service time provided by the smaller machines is ns , where s is the service time for the large computer. To see this, let λ be the present arrival rate. Then

$$E[q]_{\text{proposed}} = \frac{\lambda}{n} \frac{E[(ns)^2]}{2(1-\rho)} = \frac{\lambda n^2 E[s^2]}{2n(1-\rho)} = nE[q]_{\text{present system}}$$

and

$$\begin{aligned} E[w]_{\text{proposed}} &= E[q]_{\text{proposed}} + E[s]_{\text{proposed}} \\ &= nE[q]_{\text{present system}} + nE[s]_{\text{present system}} \\ &= nE[w]_{\text{present system}}. \end{aligned}$$

The M/G/1 queueing model is quite a useful one because random arrival patterns are quite common, although random service time is not. The model is often used as we used it in Example 5.3.1, that is, to calculate means and estimated percentile values rather than attempting to invert the Pollaczek–Khintchine transform equations. The interested readers can find the details of inverting these transforms for the M/H₂/1 queueing system in Kleinrock [10, Ch. 5]. The reader may find Tables 13–15 of Appendix C useful for models similar to those used in Example 5.3.1.

5.3.2 The GI/M/1 Queueing System

The GI/M/1 queueing system is another important model for which the embedded Markov chain technique enables us to obtain useful results. For this model we assume that the interarrival times are independent identically distributed random variables. (Such an arrival pattern is called a renewal process.) We represent the system state by the number of customers in the system at the instant of a customer arrival. This yields a stochastic process {X_n}, where X_n is the number of customers in the system when the nth customer arrives. By proceeding much as we did in Section 5.3.1, it can be shown that {X_n} is a Markov chain, and that, if $\rho < 1$, then a steady state probability distribution {π_n} exists where

$$\pi_n = P[\text{an arrival finds } n \text{ customers in the system}], \quad n = 0, 1, 2, \dots$$

For the details see Kleinrock [10] or Gross and Harris [6]. It is also shown in the above references that

$$\pi_n = \pi_0(1 - \pi_0)^n, \quad n = 0, 1, 2, \dots, \quad (5.3.49)$$

where, of course, π₀ is the probability that an arriving customer finds the system empty. Furthermore, π₀ is the unique solution of the equation

$$1 - \pi_0 = A^*(\mu\pi_0), \quad (5.3.50)$$

We give the equations for the most common types of M/G/1 queueing systems, called nonpreemptive (HOL), and preemptive resume. In both cases each of the priority classes has a Poisson arrival pattern with average arrival rate λ_i , and a general independent service time distribution with average value $E[s_i] = 1/\mu_i$. Thus, by Section 3.1.4, the total arrival rate to the system has a Poisson distribution with average rate

$$\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \quad (5.4.1)$$

By the law of total expectation, Theorem 2.8.1,

$$E[s] = \frac{\lambda_1}{\lambda} E[s_1] + \frac{\lambda_2}{\lambda} E[s_2] + \cdots + \frac{\lambda_n}{\lambda} E[s_n], \quad (5.4.2)$$

and, by the law of total moments,

$$E[s^2] = \frac{\lambda_1}{\lambda} E[s_1^2] + \frac{\lambda_2}{\lambda} E[s_2^2] + \cdots + \frac{\lambda_n}{\lambda} E[s_n^2], \quad (5.4.3)$$

for both queueing systems. The remainder of the equations for the HOL queueing system are given in Table 18, Appendix C, while those for the preemptive resume queueing system are given in Table 19.

We illustrate the effects of priority queueing by the following example.

Example 5.4.1 An on-line inquiry system receives two types of inquiries. Type 1 inquiries arrive in a Poisson pattern at an average arrival rate of 0.9 per second. The time required for the system to respond is nearly constant with an average value of 0.4 seconds. The Type 2 inquiries arrive at an average rate of 1 every 10 seconds. The system response time for Type 2 inquiries has a two-stage hyperexponential distribution with $\alpha_1 = 0.4$, $\alpha_2 = 0.6$, $1/\mu_1 = 10$ seconds, $1/\mu_2 = 5/3$ seconds; so the average system response time for Type 2 inquiries is 5 seconds, with a second moment of 83.33 seconds². Contrast the operation of the system with (a) no priorities, (b) with an HOL priority system that gives priority to Type 1 inquiries, and (c) with preemptive-resume priority given to Type 1 inquiries.

Solution (a) For a nonpriority system the average service time

$$E[s] = 0.9 \times 0.4 + 0.1 \times 5 = 0.36 + 0.5 = 0.86 \text{ seconds}$$

$$E[s^2] = 0.9 \times 0.4^2 + .1 \times 83.33 = 8.477 \text{ seconds}^2$$

$$\rho = \lambda E[s] = 0.86$$

$$W_q = E[q] = \frac{\lambda E[s^2]}{2(1 - \rho)} = 30.275 \text{ seconds.}$$

Average time in the system for Type 1 inquiries is

$$W_1 = 30.275 + 0.4 = 30.675 \text{ seconds.}$$

Average time in the system for Type 2 inquiries is

$$W_2 = 30.275 + 5 = 35.275 \text{ seconds.}$$

Overall average waiting time in the system is

$$W = W_q + E[s] = 31.135 \text{ seconds.}$$

- (b) For an HOL queueing system with Type 1 inquiries having non-preemptive priority over Type 2 inquiries

$$u_1 = 0.9 \times 0.4 = 0.36 \text{ seconds, } u_2 = 0.36 + 0.1 \times 5 = 0.86 \text{ seconds.}$$

The average queueing times are

$$W_{q_1} = \frac{\lambda E[s^2]}{2(1 - u_1)} = 6.6227 \text{ seconds}$$

$$W_{q_2} = \frac{\lambda E[s^2]}{2(1 - u_1)(1 - u_2)} = 47.3047 \text{ seconds.}$$

The average times in the system are

$$W_1 = W_{q_1} + E[s_1] = 7.0227 \text{ seconds,}$$

$$W_2 = W_{q_2} + E[s_2] = 52.3047 \text{ seconds.}$$

The overall average queueing time

$$W_q = 0.9 \times 6.6227 + 0.1 \times 47.3047 = 10.6909 \text{ seconds.}$$

The overall average waiting time in the system is

$$W = W_q + E[s] = 11.5509 \text{ seconds.}$$

- (c) For a priority queueing system with Type 1 inquiries receiving preemptive-repeat priority over Type 2 inquiries, using u_1 and u_2 from (b), yields the waiting times in the system

$$W_1 = E[s_1] + \frac{\lambda_1 E[s_1^2]}{2(1 - u_1)} = 0.4 + \frac{0.9 \times 0.16}{2(1 - 0.36)} = 0.5125 \text{ seconds}$$

$$\begin{aligned} W_2 &= \frac{1}{(1 - u_1)} \left[E[s_2] + \frac{\lambda_1 E[s_1^2] + \lambda_2 E[s_2^2]}{2(1 - u_2)} \right] \\ &= \frac{1}{1 - 0.36} \left[5 + \frac{0.9 \times 0.16 + 0.1 \times 83.33}{2(1 - 0.86)} \right] = 55.1172 \text{ seconds.} \end{aligned}$$

The corresponding average queueing times for the two inquiry types are

$$W_{q_1} = W_1 - E[s_1] = 0.1125 \text{ seconds},$$

$$W_{q_2} = W_2 - E[s_2] = 50.1172 \text{ seconds}.$$

The overall average queueing time

$$W_q = 0.9 \times 0.1125 + 0.1 \times 50.1172 = 5.1130 \text{ seconds},$$

and the overall average waiting time in the system

$$W = 0.9 \times 0.5125 + 0.1 \times 55.1172 = 5.9224 \text{ seconds}.$$

We summarize this data from this example in Table 5.4.1.

TABLE 5.4.1
Results of Example 5.4.1^a

	No priority	HOL priority	Preemptive-resume priority
W_{q_1} (type 1)	30.275	6.6227	0.1125
W_{q_2} (type 2)	30.275	47.3047	50.1172
W_1 (type 1)	30.675	7.0227	0.5125
W_2 (type 2)	35.275	52.3047	55.1172
W_q	30.275	10.6909	5.1130
W	31.124	11.5509	5.9724

^a All times in seconds.

The results shown in Table 5.4.1 illustrate how a priority system can dramatically improve the performance of a queueing system.

The average queueing time for a Type 1 inquiry drops from 30.275 seconds for a nonpriority system to 6.6227 seconds for an HOL queueing system; for a preemptive-resume system, it is only 0.1125 seconds! The overall average queueing time drops from 30.275 seconds to 10.6909 seconds, and then to 5.113 seconds; the improvement in average system time is similar. The performance of the system for Type 2 inquiries suffers, but not severely.

The reader is asked to show in Exercise 37 that, if the Type 2 inquiries were given priority over Type 1 inquiries, the overall average queueing and system times would be larger for the priority systems than for the original nonpriority system.

so that

$$S = 1/p_0 = \sum_{n=0}^{\infty} u^n/n! = e^u.$$

Hence,

$$p_n = e^{-u}(u^n/n!), \quad n = 0, 1, 2, \dots, \quad (5.2.80)$$

that is, N has a Poisson distribution! It can be shown (see Gross and Harris [6]) that (5.2.80) is also true for an $M/G/\infty$ queueing system. The fact that p_n has a Poisson distribution tells us that $L = E[N] = u$ is the average number of busy servers, with $\text{Var}[N] = u$. The $M/M/\infty$ queueing model can be used to estimate the number of lines in use in a large communication network or as a gross estimate of values in an $M/M/c$ or $M/M/c/c$ queueing system for large values of c . In Example 5.2.8, u was 7 erlangs which was close to the average number of servers in use for the $M/M/15/15$ queueing system. Also

$$p_{15} = 0.00332 \approx e^{-7} \frac{7^{15}}{15!} = 0.00331.$$

Example 5.2.9 Calls in a telephone system arrive randomly at an exchange at the rate of 140 per hour. If there are a very large number of lines available to handle the calls which last an average of 3 minutes, what is the average number of lines in use? Estimate the 90th and 95th percentile of number of lines in use.

Solution The $M/M/\infty$ model can be used to estimate the requested values. For this example

$$u = \lambda E[s] = \frac{140}{60} \frac{\text{calls}}{\text{minute}} \times \frac{3 \text{ minutes}}{\text{call}} = 7 \text{ erlangs.}$$

Hence, the average number of lines in use is 7. We can use the normal approximation as a first estimate of percentile values. The 90th percentile value of the normal distribution is the mean plus 1.28 standard deviations; the 95th percentile value is the mean plus 1.645 standard deviations. Thus the 90th percentile value of number of lines is $7 + 1.28\sqrt{7} = 10.38$ or 11 lines; the 95th percentile value is $7 + 1.645\sqrt{7} = 11.35$ or 12 lines.

5.2.6 The $M/M/1/K/K$ Queueing System (Machine Repair with One Repairman)

This model, a limited source model in which there are only K customers, is variously called the machine repair model, the machine interference

Similarly, using the distribution function of q , we calculate

$$\pi_q(r) = \frac{E[s]}{1 - \rho} \ln \left(\frac{100\rho}{100 - r} \right) = \frac{E[q]}{\rho} \ln \left(\frac{100\rho}{100 - r} \right). \quad (5.2.33)$$

A number of formulas for an M/M/1 queueing system are shown in Table 3 of Appendix C and can be evaluated by the APL function $M\Delta M\Delta 1$ of Appendix B.

Example 5.2.1 For a small batch computing system the processing time per job is exponentially distributed with an average time of 3 minutes. Jobs arrive randomly at an average rate of one job every 4 minutes and are processed on a first-come-first-served basis. The manager of the installation has the following concerns.

(a) What is the probability that an arriving job will require more than 20 minutes to be processed (the job turn-around time exceeds 20 minutes)?

(b) A queue of jobs waiting to be processed will form, occasionally. What is the average number of jobs waiting in this queue?

(c) It is decided that, when the work load increases to the level such that the average time in the system reaches 30 minutes, the computer system capacity will be increased. What is the average arrival rate of jobs per hour at which this will occur? What is the percentage increase over the present job load? What is the average number of jobs in the system at this time?

(d) Suppose the criterion for upgrading the computer capacity is that not more than 10% of all jobs have a time in the system (turn-around time) exceeding 40 minutes. At the arrival rate at which this criterion is reached, what is the average number of jobs waiting to be processed?

Solution (a) $E[\tau] = 4$ minutes, so

$$\lambda = 1/E[\tau] = 0.25 \text{ jobs/minute},$$

and

$$\rho = \lambda E[s] = 0.25 \times 3 = 0.75.$$

The average time in the system, $W = E[s]/(1 - \rho) = 12$ minutes, so, by (5.2.27),

$$W(t) = P[w \leq t] = 1 - e^{-t/12} \quad \text{or} \quad P[w > t] = e^{-t/12}.$$

Therefore, the probability that w exceeds 20 minutes is $e^{-20/12} = e^{-5/3} = 0.1889$.

(b) If we assume a job queue has not formed unless there is a job in it, we use the formula

$$E[N_q | N_q > 0] = 1/(1 - \rho) = 4 \text{ jobs}$$

(see Exercise 3).

If the question is interpreted to mean the average job queue length, including queues of length zero, then we calculate

$$L_q = E[N_q] = \rho^2 / (1 - \rho) = (0.75)^2 / 0.25 = 2.25 \text{ jobs.}$$

The most reasonable answer to the question, as stated, is 4 jobs.

(c) When $W = 30$ minutes the system is to be upgraded, assuming the current $E[s]$ is 3 minutes. We solve the equation

$$30 = W = \frac{E[s]}{1 - \lambda E[s]} = \frac{3}{1 - 3\lambda}$$

or

$$\lambda = 27/90 = 3/10 \text{ jobs/minute} = 18 \text{ jobs/hour.}$$

The percentage increase is

$$100 \times (18 - 15)/15 = 100/5 = 20\%.$$

When $\lambda = 18$ jobs/hour = 3/10 jobs/minute, the average number of jobs in the system

$$L = \rho / (1 - \rho) = 0.9 / (1 - 0.9) = 9 \text{ jobs.}$$

(d) The criterion is that $\pi_w(90)$ reaches 40 minutes. We solve the equation

$$40 = \pi_w(90) = 2.3W = \frac{2.3 \times E[s]}{1 - \lambda E[s]} = \frac{2.3 \times 3}{1 - 3\lambda},$$

to obtain

$$\lambda = \frac{33.1}{120} \text{ jobs/minute} = 60 \times \frac{33.1}{120} = 16.55 \text{ jobs/hour.}$$

That is only a $[(16.55 - 15)/15] \times 100 = 10.3\%$ increase over the present arrival rate. At this arrival rate $\rho = \lambda E[s] = 0.8275$ and the average number of jobs in the queue is

$$L_q = E[N_q] = \rho^2 / (1 - \rho) = 3.97.$$

This is an increase over the current value of 2.25 jobs. The average time in the system at this increased arrival rate is 17.39 minutes; it is only 12 minutes at the current arrival rate.

In part (c) of the above example we see that increasing the arrival rate by 20% increased the average time a job would spend in the system from 12 minutes to 30 minutes—a 150% increase! The reason for this phenomenon is shown graphically in Fig. 5.2.2. The curve of $E[w]/E[s]$ rises sharply as ρ approaches the value 1.

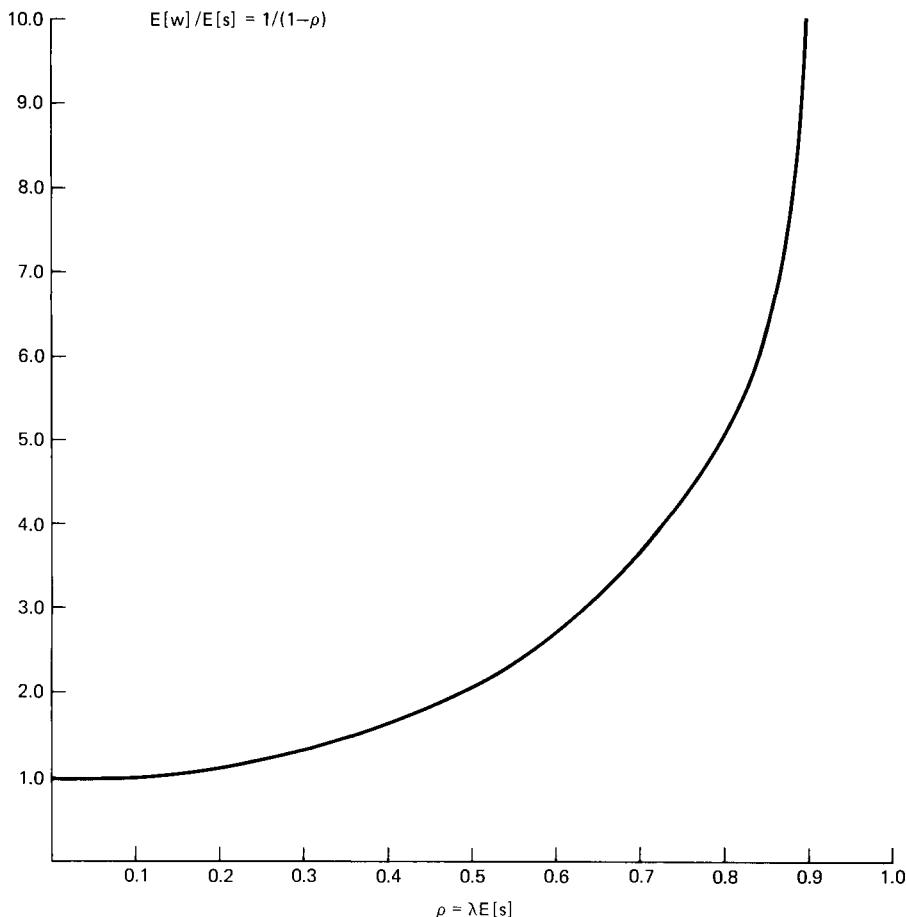


Fig. 5.2.2 Normalized average time in the system, $E[w]/E[s]$, for M/M/1 queueing system.

That is, the slope of the curve increases rapidly as ρ grows beyond about 0.8. Since

$$dW/d\rho = E[s](1 - \lambda E[s])^{-2},$$

a small change in ρ (due to a small change in λ , assuming $E[s]$ is fixed) causes a change in W given approximately by

$$(dW/d\rho)\Delta\rho = (dW/d\rho)E[s]\Delta\lambda = E[s]^2(1 - \lambda E[s])^{-2}\Delta\lambda.$$

Thus, if $\rho = 0.5$, a change $\Delta\lambda$ in λ will cause a change in W of about $4E[s]^2\Delta\lambda$, while, if $\rho = 0.9$, the change in W will be about $100E[s]^2\Delta\lambda$, or 25 times the size of the change that occurred for $\rho = 0.5$!

That is, when the system is operating at 90% server utilization, a small change in the system load (arrival rate) will cause 25 times as great an increase in the average system time as the same increase in load would cause if the system were operating at 50% utilization! This illustrates the danger of designing a system to operate at a high utilization level—a small increase in the load can have disastrous effects on the system performance.

Example 5.2.2 A computing facility has a large computer dedicated to a certain type of on-line application for users who are scattered about the country. The arrival pattern of requests to the central machine is random (Poisson), and the service time provided is random (exponential) also, so the system is an M/M/1 queueing system. A proposal is made that the workload be divided equally among n smaller machines—each with $1/n$ times the processing power of the original machine. It is claimed that the response time (time a request is in the system) will not change but the users will have a local computer. Are these claims justified?

Solution Let λ, μ be the average arrival and service rates, respectively, of the current system so that $\rho = \lambda/\mu$ is the computer utilization. For each of the proposed new systems the average arrival rate is λ/n and the average service rate is μ/n , so the server utilization is $(\lambda/n)/(\mu/n) = \lambda/\mu = \rho$, the same value as the present system. If we assume the small computers also provide random service, then

$$\frac{W_{\text{proposed}}}{W_{\text{current}}} = \left(\frac{n/\mu}{(1 - \rho)} \right) / \left(\frac{1/\mu}{1 - \rho} \right) = n,$$

and

$$\frac{W_{q_{\text{proposed}}}}{W_{q_{\text{current}}}} = \left(\frac{\rho n/\mu}{1 - \rho} \right) / \left(\frac{\rho/\mu}{1 - \rho} \right) = n.$$

Thus, the average time in the system and the average time in the queue would *increase n-fold* rather than remain the same! Of course the n new computer systems, together, process the same number of requests per hour as before, but each individual request requires n times as long to be processed, on the average, as in the present system. Thus, if the present system has an average service time of 2 seconds with a utilization of 0.7, then it has an average response time of 6.67 seconds; a proposed system of 10 computers, each providing 20-second service time, would yield a response time of 66.7 seconds! The effect discussed in this example is called the “scaling effect” and is discussed more fully by Streeter [4]. The result can be used to show that centralizing a computing facility can improve the response time while providing more computing capability for less money (economy of scale).

Example 5.2.3 A branch office of a large engineering firm has one on-line terminal connected to a central computer system for 16 hours each day. Engineers, who work throughout the city, drive to the branch office to use the terminal for making routine calculations. The arrival pattern of engineers is random (Poisson) with an average of 20 persons per day using the terminal. The distribution of time spent by an engineer at the terminal is exponential with an average time of 30 minutes. Thus the terminal is $5/8$ utilized ($20 \times 1/2 = 10$ hours out of 16 hours available). The branch manager receives complaints from the staff about the length of time many of them have to wait to use the terminal. It does not seem reasonable to the manager to procure another terminal when the present one is only used five-eighths of the time, on the average. How can queueing theory help this manager?

Solution The M/M/1 queueing system is a reasonable model with $\rho = 5/8$, as we computed above. The M/M/1 formulas give the following.

$$W = E[w] = E[s]/(1 - \rho) = 80 \text{ minutes.}$$

Average time an engineer spends at the branch office.

$$L_q = \rho^2/(1 - \rho) = 1.0417.$$

Average number of engineers waiting in the queue.

$$E[N_q | N_q > 0] = 1/(1 - \rho) = 8/3.$$

Average number of engineers in nonempty queues.

$$W_q = E[q] = \rho E[s]/(1 - \rho) = 50 \text{ minutes.}$$

Average waiting time in queue.

$$E[q | q > 0] = E[w] = 80 \text{ minutes.}$$

Average waiting time of those who must wait.

$$\pi_q(90) = W \ln(10\rho) = 146.61 \text{ minutes.}$$

90th percentile of time in the queue.

$$\pi_w(90) \approx 2.3W = 184 \text{ minutes.}$$

90th percentile time in the branch office.

Since $\rho = 5/8$, only three-eighths of the engineers who use the terminal need not wait. For those who must wait, the average wait for the terminal is 80 minutes—quite a long wait, by most standards! Ten percent of the engineers spend over 3 hours (actually 184 minutes) in the office to do an average of 30 minutes of computing. The probability of waiting more than an hour to use the terminal is

$$P[q > 60] = \frac{5}{8}e^{-60/80} = 0.295229,$$

or almost 30%.

These results may seem a little startling to those not acquainted with

queueing theory. It might seem, intuitively, that adding another terminal would cut the average waiting time in half—from 50 minutes to 25 minutes (to 40 minutes for those who must wait). We shall see, in Example 5.2.6, that the improvement is much more dramatic than this. The queueing theory we have presented so far should suffice to convince the manager that an improvement is needed.

Example 5.2.4 Traffic to a message switching center for one of the outgoing communication lines arrives in a random pattern at an average rate of 240 messages per minute. The line has a transmission rate of 800 characters per second. The message length distribution (including control characters) is approximately exponential with an average length of 176 characters. Calculate the principal statistical measures of system performance assuming that a very large number of message buffers are provided. What is the probability that 10 or more messages are waiting to be transmitted?

Solution The average service time is the average time to transmit a message or

$$E[s] = \frac{\text{average message length}}{\text{line speed}} = \frac{176 \text{ characters}}{800 \text{ characters/second}} = 0.22 \text{ seconds.}$$

Hence, since the average arrival rate

$$\lambda = 240 \text{ messages/minute} = 4 \text{ messages/second},$$

the server utilization

$$\rho = \lambda E[s] = 4 \times 0.22 = 0.88,$$

that is, the communication line is transmitting outgoing messages 88% of the time. Using the M/M/1 formulas of Table 3, Appendix C we calculate the following.

$$L = E[N] = \rho / (1 - \rho) = 7.33 \text{ messages.}$$

Average number of messages in the system.

$$L_q = E[N_q] = \rho^2 / (1 - \rho) = 6.45 \text{ messages.}$$

Average number of messages in the queue waiting to be transmitted.

$$W = E[w] = E[s] / (1 - \rho) = 1.83 \text{ seconds.}$$

Average time a message spends in the system.

$$W_q = E[q] = \rho E[s] / (1 - \rho) = 1.61 \text{ seconds.}$$

Average time a message waits for transmission.

$$\pi_w(90) = 2.3W = 4.209 \text{ seconds.}$$

90th percentile time in the system.

$$\pi_q(90) = W \ln(10\rho) = 3.98 \text{ seconds.}$$

90th percentile waiting time in queue (90% of the messages wait no longer than 3.98 seconds.)

Since 10 or more messages are waiting if and only if 11 or more messages are in the system, the required probability is

$$P[11 \text{ or more messages in the system}] = \rho^{11} = 0.245.$$

Our discussion of the M/M/1 model has been more complete than it will be for many queueing models because it is an important but simple model. It is also a pleasant model to study because the probability distributions of the random variables w , q , N , and N_q can be calculated; for some queueing models only the averages W , W_q , L , and L_q can be computed, and these only with difficulty. A number of systems can be modeled, at least in a limiting sense, as an M/M/1 queueing system.

5.2.2 The M/M/1/K Queueing System

Example 5.2.4 was somewhat unrealistic in the sense that no message switching system can have an unlimited number of buffers. The M/M/1/ K system is a more accurate model of this type of system in which a limit of K customers is allowed in the system. When the system contains K customers, arriving customers are turned away. Figure 5.2.3 is the state-transition diagram for this model. Thus, as a birth-and-death process, the coefficients are

$$\lambda_n = \begin{cases} \lambda & \text{for } n = 0, 1, 2, \dots, K-1 \\ 0 & \text{for } n \geq K, \end{cases} \quad (5.2.34)$$

and

$$\mu_n = \begin{cases} \mu & \text{for } n = 1, 2, \dots, K \\ 0 & \text{for } n > K. \end{cases} \quad (5.2.35)$$

This gives the steady state probabilities

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 = u^n p_0 \quad \text{for } n = 0, 1, 2, \dots, K, \quad (5.2.36)$$

where

$$u = \lambda E[s] = \lambda/\mu.$$

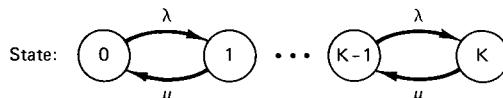


Fig. 5.2.3 State-transition diagram for the M/M/1/ K queueing system.

Proceeding as we did in deriving (5.2.27) for the M/M/1 model, we calculate (see Exercise 9)

$$\begin{aligned}
 W(t) &= P[w \leq t] \\
 &= \sum_{n=0}^{K-1} P[w \leq t | N_a = n] P[N_a = n] \\
 &= \sum_{n=0}^{K-1} \left[\int_0^t \frac{\mu(\mu x)^n}{n!} e^{-\mu x} dx \right] q_n \\
 &= 1 - \sum_{n=0}^{K-1} q_n P[\mu t, n], \tag{5.2.48}
 \end{aligned}$$

where

$$P[\mu t, n] = \sum_{k=0}^n e^{-\mu t} \frac{(\mu t)^k}{k!}$$

is the Poisson distribution function and N_a is the random variable which counts the number of customers in an M/M/1/ K queueing system just before a customer arrives to enter the system. (Thus N_a assumes the values 0, 1, 2, ..., $K - 1$ and $P[N_a = n] = q_n$.) $W(t)$ can be calculated with the aid of tables of the Poisson distribution function or by using an APL function such as POISSONADIST. The APL function DMAMDA1ΔK computes the values $W(t)$ and $W_q(t)$ for the M/M/1/ K model. The same reasoning that led to (5.2.48) shows that $W_q(t)$ is given by

$$\begin{aligned}
 W_q(t) &= P[q \leq t] = W_q(0) + \sum_{n=1}^{K-1} P[q \leq t | N_q = n] q_n \\
 &= q_0 + \sum_{n=1}^{K-1} q_n \int_0^t \frac{\mu(\mu x)^{n-1}}{(n-1)!} e^{-\mu x} dx = 1 - \sum_{n=0}^{K-2} q_{n+1} P[\mu t, n], \tag{5.2.49}
 \end{aligned}$$

where

$$P[\mu t, n] = \sum_{k=0}^n \frac{e^{-\mu t} (\mu t)^k}{k!}.$$

Example 5.2.5 Consider Example 5.2.4. Suppose we have the same arrival pattern, message length distribution, and line speed as described in the example. Suppose, however, that it is desired to provide only enough message buffers so that the probability is less than one-half of a percent (probability < 0.005) that all the buffers are filled at any particular time. How many buffers should be provided? For the required number of buffers calculate L , L_q , W , and W_q . What is the probability that the time an arriving message spends in the system does not exceed 2.5 seconds? What is the

probability that the queueing time of a message before transmission is begun does not exceed 2.5 seconds?

Solution The M/M/1/K model fits this system with $u = \lambda E[s] = 0.88$ erlangs. The probability that all the buffers are filled, given $K - 1$ buffers are provided, is

$$p_K = \frac{(1-u)u^K}{1-u^{K+1}} \quad \text{where } u = 0.88. \quad (5.2.50)$$

APL can be used to ease the burden of calculating the values of p_K . The number of buffers required is expected to be large. Let us try $K = 20$ (19 buffers). Using (5.2.50), we obtain

$$p_{20} = \frac{(1-0.88)(0.88)^{20}}{1-(0.88)^{21}} = 0.00998936.$$

This is almost 1% so we try $K = 25$, which yields $p_{25} = 0.005095$; still not enough buffers.

Since $p_{26} = 0.004464 < 0.005$, we need 25 buffers, which allows 26 messages in the system (counting the one being transmitted).

Using the APL function $\text{M}\Delta\text{M}\Delta1\Delta K$, which makes the calculations for the M/M/1/K model (shown in Table 4 of Appendix C), we obtain the following.

$L = E[N] = 6.449$ messages.	Average number of messages in the system.
$\sigma_N = 6.0387$ messages.	Standard deviation of the number of messages in the system.
$L_q = E[N_q] = 5.573$ messages.	Average number of messages queued for the line.
$\sigma_{N_q} = 5.914$ messages.	Standard deviation of number of messages queued for the line.
$W = E[w] = 1.62$ seconds.	Average time a message spends in the system (queueing for the line and being transmitted).
$W_q = E[q] = 1.40$ seconds.	Average time a message queues for the line.
$E[q q > 0] = 1.60$ seconds.	Average time in the queue for those messages delayed.

All of these numbers are smaller than the numbers for the M/M/1 model with the same λ and $E[s]$.

Using the APL function $\text{DM}\Delta\text{M}\Delta\text{1}\Delta\text{K}$ we calculate the probability an arriving message is in the system for not more than 2.5 seconds is

$$W(2.5) = 0.77208, \quad \text{while} \quad W_q(2.5) = P[q \leq 2.5] = 0.8039.$$

We also calculate the probability the system is empty, p_0 , is 0.123928, while the server utilization, ρ , is 0.87607.

For the M/M/1 system of Example 5.2.4 with unlimited queue length, the 90th percentile of waiting time in the system, $\pi_w(90)$ is 4.216667 seconds, but, for the corresponding M/M/1/26 system, $P[w \leq 4.216667] = 0.9328$; it appears that all the performance statistics for the M/M/1/26 system are superior to those for the system with unlimited queue length. The penalty for this improved performance, however, is that $100 \times p_K = 0.4464\%$ of the messages are refused and must be resent at a later time.

5.2.3 The M/M/c Queueing System

For this model we assume random (exponential) interarrival and service times with c identical servers. This system can be modeled as a birth-and-death process with the coefficients

$$\lambda_n = \lambda, \quad n = 0, 1, 2, \dots, \quad (5.2.51)$$

and

$$\mu_n = \begin{cases} n\mu, & n = 1, 2, \dots, c \\ c\mu, & n \geq c. \end{cases} \quad (5.2.52)$$

The state-transition diagram is shown in Fig. 5.2.4.

Thus, by (5.2.2), with $u = \lambda/\mu$ and $\rho = u/c$,

$$C_n = \begin{cases} \frac{u^n}{n!}, & n = 1, 2, 3, \dots, c, \\ \frac{u^c}{c!} \left(\frac{u}{c}\right)^{n-c}, & n = c, c+1, \dots \end{cases} \quad (5.2.53)$$

Hence, if $\rho < 1$ so that the steady state exists, then

$$\begin{aligned} S = \frac{1}{p_0} &= 1 + u + \frac{u^2}{2!} + \cdots + \frac{u^{c-1}}{(c-1)!} + \frac{u^c}{c!} \left(1 + \frac{u}{c} + \left(\frac{u}{c}\right)^2 + \cdots\right) \\ &= \sum_{n=0}^{c-1} \frac{u^n}{n!} + \frac{u^c}{c!} \sum_{n=0}^{\infty} \rho^n = \sum_{n=0}^{c-1} \frac{u^n}{n!} + \frac{u^c}{c! (1-\rho)}. \end{aligned} \quad (5.2.54)$$

Hence

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{u^n}{n!} + \frac{u^c}{c! (1-\rho)} \right]^{-1} \quad (5.2.55)$$

which user programs were swapped in and out of main memory with only one complete program in memory at a time. Since there was no overlap of program execution and swapping, Scherr used the sum of program execution time and swapping time as the CPU service time. The machine repair analytic model gave results that were very close to those for a simulation model and to actual measured values.

For this model, since the operating time for machines corresponds to think time, with average think time $1/\alpha$, we have, by (5.2.89), the mean response time

$$W = (N/\lambda) - (1/\alpha). \quad (6.3.1)$$

But $\lambda = \rho/E[s]$, and, by (5.2.83),

$$\rho = 1 - p_0. \quad (6.3.2)$$

Therefore, the mean response time can be written as

$$W = \frac{NE[s]}{1 - p_0} - \frac{1}{\alpha}, \quad (6.3.3)$$

where, by (5.2.82),

$$p_0 = \left[\sum_{n=0}^N \frac{N!}{(N-n)!} \left(\frac{\alpha}{\mu} \right)^n \right]^{-1}. \quad (6.3.4)$$

Example 6.3.1 SLOBOVIAN SCIENTIFIC has an interactive time-sharing system of 20 active terminals which can be studied by the machine repair model. The average CPU service time, including swapping, is 2 seconds, while the mean think time is 20 seconds. Find p_0 , ρ , λ , and the average response time W . Note that λ is the average throughput in interactions per second. What would be the effect of adding five terminals?

Solution For 20 terminals

$$p_0 = \left[\sum_{n=0}^{20} \frac{20!}{(20-n)!} \left(\frac{2}{20} \right)^n \right]^{-1} = 0.001869.$$

Then,

$$\rho = 1 - p_0 = 0.998131, \quad \lambda = 0.49907 \text{ interactions/second},$$

and

$$W = (20/0.49907) - 20 = 20.075 \text{ seconds.}$$

With 25 terminals

$$p_0 = 0.00002927, \quad \rho = 1 - p_0 = 0.99997073,$$

$$\lambda = \rho/E[s] = 0.499985365, \quad \text{and} \quad W = (25/\lambda) - 20 = 30 \text{ seconds.}$$

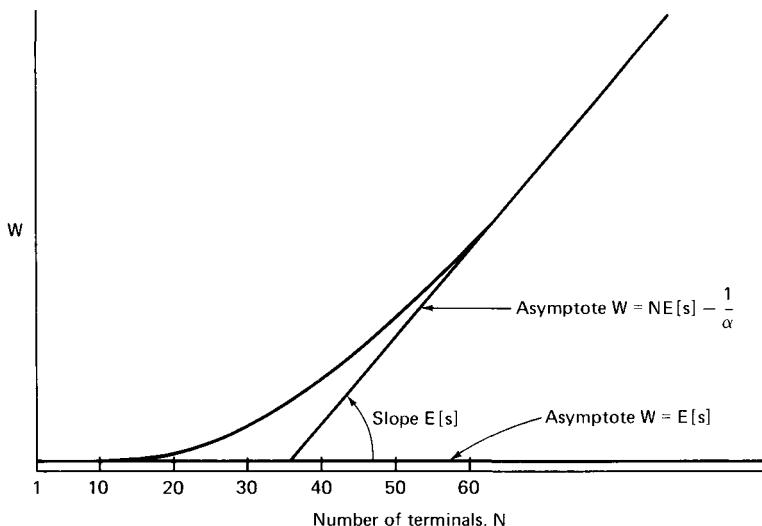


Fig. 6.3.2 Mean response time W versus N , the number of terminals, for the machine repair model.

Thus the addition of 5 terminals has increased the average throughput by only 0.18% while increasing the mean response time by 49.44%.

This example illustrates the concept of system saturation. Consider Fig. 6.3.2, the graph of the mean response time $W = \{NE[s]/(1 - p_0)\} - (1/\alpha)$, versus N . For $N = 1$ there is no queueing so $W = E[s]$. For small values of N , the customers interfere with each other very little; that is, when one person wants a CPU interaction the others are usually in think mode so little queueing occurs. Thus the curve is asymptotic at $N = 1$ to the line $W = E[s]$. As $N \rightarrow \infty$, $p_0 \rightarrow 0$ since the likelihood of the CPU being idle must go to 0. Hence the curve is asymptotic to the line $NE[s] - (1/\alpha)$ as $N \rightarrow \infty$. Clearly the two asymptotes intersect where

$$N = N^* = \frac{E[s] + (1/\alpha)}{E[s]} = \frac{E[s] + E[t]}{E[s]}.$$

Kleinrock [4] calls N^* the *system saturation point*. He points out that, if each interaction required exactly $E[s]$ units of CPU service time and exactly $E[t]$ units of think time, then N^* is the maximum number of terminals that could be scheduled in such a way as to cause no mutual interference. For $N \ll N^*$ there is almost no mutual interference, and W is approximately $E[s]$. For $N \gg N^*$ users "totally interfere" with each other; that is the addition of a terminal raises everyone's average response time by $E[s]$. In Example 6.3.1, $N^* = 22/2 = 11$ terminals, and the increase in W due to the change from 20

terminals to 25 terminals was close to $5 \times 2 = 10$ seconds (actually it was 9.963 seconds).

Lassettre and Scherr [14] have also successfully used the machine repair model to develop the OS/360 time-sharing option (TSO).

Kobayashi [25] shows equations (6.3.1)–(6.3.3) hold for any system like that shown in Fig. 6.3.1 providing (a) the equilibrium or steady state of the system exists and (b) the queue discipline is “work conserving” in the sense that the service times of the individual requests are not affected by it. Thus these equations have a simple form independent of the distributional form of t and s . Of course the value of p_0 may depend upon the distributional form of t and s as well as on the queue discipline; we shall see this in two special cases of this model described in Sections 6.3.2 and 6.3.3.

6.3.2 Finite Processor-Sharing Model

This model is Fig. 6.3.1 with a single CPU operating with the processor-sharing queue discipline; that is, the CPU operates as an M/G/1 processor-sharing system but with the finite input shown in Fig. 6.3.1. Kleinrock [4] shows that if the CPU service time is only restricted to the extent that the Laplace-Stieltjes transform $W_s^*(\theta)$ is rational (the ratio of two polynomials in θ), and similarly for the think time, then exactly the same equations hold as we found in the last section for the exponential CPU service time with FCFS queue discipline. That is

$$W = \frac{NE[s]}{1 - p_0} - \frac{1}{\alpha}, \quad (6.3.5)$$

where $1/\alpha$ is the average think time and

$$p_0 = \left[\sum_{n=0}^N \frac{N!}{(N-n)!} \left(\frac{\alpha}{\mu} \right)^n \right]^{-1}. \quad (6.3.6)$$

Also

$$\rho = 1 - p_0, \quad (6.3.7)$$

and

$$\lambda = \rho/E[s]. \quad (6.3.8)$$

Of course all the other equations for the machine repair queueing system (M/M/1/K/K) hold as well, with K replaced by N and O by think time t .

Example 6.3.2 Let us consider the example on page 317 of Reiser [6], which he solves using a very sophisticated APL program called QNET4. He

considers a finite processor-sharing model with 20 active terminals, $1/\alpha = 3$ seconds, a CPU average service rate of 500,000 instructions/second and an average interaction requirement of 100,000 instructions. Thus $E[s] = 100,000/500,000 = 0.2$ seconds. Hence $E[s]/E[t] = 0.2/3 = 1/15$, and

$$\begin{aligned} p_0 &= \left[\sum_{n=0}^{20} \frac{20!}{(20-n)!} \left(\frac{1}{15}\right)^n \right]^{-1} = \left[20! \sum_{n=0}^{20} \frac{1}{(20-n)!} \left(\frac{1}{15}\right)^n \right]^{-1} \\ &= 0.045593216. \end{aligned}$$

The mean response time

$$W = \frac{NE[s]}{1 - p_0} - \frac{1}{\alpha} = 1.191 \text{ seconds}$$

agrees with Reiser's solution as does the average throughput

$$\lambda_T = \rho/E[s] = (1 - p_0)/E[s] = 4.772 \text{ interactions/second.}$$

The average number in the central processor system

$$\lambda_T W = 5.6835, \quad \text{and} \quad \rho = 1 - p_0 = 0.954407$$

also agrees with Reiser's solution.

Note that in the simplified form of Reiser's example considered in Example 6.2.1 we assumed the value of λ_T , whereas in this example we actually had to calculate it. Note also that in the infinite processor-sharing model used in Example 6.2.1 we computed $W = 4.348$ seconds and the average number in the CPU system as 20.74! This agrees with the results of Buzen and Goldberg [1] who show the infinite source approximation is poor for high server utilization.

6.3.3 The Straightforward Model of Boyse and Warn

Boyse and Warn [12] have developed a computer performance prediction model that is very useful for certain kinds of computer systems. The model, which is shown in Fig. 6.3.3, is essentially that of Fig. 6.3.1 with the "central processor system" box replaced by the system inside the dashed lines. The assumptions made in the Boyse and Warn model are as follows.

- (1) A fixed multiprogramming level K , which implies the system is heavily loaded so the queue for main memory, is never empty.
- (2) Multiple CPUs or a single CPU with each treated as a single server.
- (3) There are K parallel I/O servers, which implies there is no queueing for I/O service.

The formulas for the $M/M/c$ system are given in Table 5 of Appendix C. Table 6 gives the formulas for the special case that $c = 2$. Figure 1 is a graph of $C(c, u)$ as a function of traffic intensity u .

Example 5.2.6 In Example 5.2.3 the branch manager was mystified by the complaints of the staff concerning the availability of the computer terminal; it was idle $3/8$ of the time yet the engineers were complaining about excessive waiting times. One of the engineers with an understanding of queueing theory explained to the manager that the theory showed that nearly 30% of the personnel using the terminal would have to wait more than an hour to gain access to it; that the average waiting time for the $5/8$ of the users who must wait was 80 minutes. A committee of senior engineers met with the branch manager and jointly decided that the situation was intolerable. Moreover, it could not be solved by scheduling terminal usage. It was decided that enough terminals should be provided to ensure that the average queueing time should not exceed 10 minutes, that 90% of all engineers should queue less than 15 minutes, and that not more than 5% of all terminal users should have to queue more than 1 hour. After the specifications were agreed upon the manager had second thoughts. If the average value of queueing time is 50 minutes with one terminal, it seemed that it would require 5 terminals to reduce this value to 10 minutes; worse yet, to drop the 90th percentile value of waiting time in queue from 184 minutes to 15 minutes would require 13 terminals! How many terminals are actually needed?

Solution If more terminals are added at the branch office the $M/M/c$ model applies. (We assume that adding a few terminals to the corporate on-line system will not affect the terminal response time of each terminal.) Let us first try two terminals and use the formulas of Table 6, Appendix C (we could also use the APL function $M\Delta M\Delta C$ of Appendix B). With two terminals ($c = 2$) the server utilization ρ is $5/16$. The average waiting time in the queue, W_q , is $\rho^2 E[s]/(1 - \rho^2) = 3.247$ minutes. The 90th percentile waiting time in the queue is

$$\pi_q(90) = \frac{E[s]}{2(1 - \rho)} \ln \left(\frac{20\rho^2}{1 + \rho} \right) = 8.673 \text{ minutes.}$$

The probability that the waiting time in queue exceeds 1 hour is

$$P[q > 60] = \frac{2\rho^2}{1 + \rho} e^{-2 \times 60(1 - \rho)/30} = 0.00951$$

or 0.951%. Thus all the specifications are satisfied with one additional terminal in the branch office and the branch manager relaxes. However, it would save commuting time and gasoline if the second terminal were placed

in a convenient location across town so that half the engineering force could use it, thus having two M/M/1 systems, each with $u = 5/16$ erlang. Using the formulas for M/M/1 from Table 3, Appendix C, as implemented by the APL function $M\Delta M\Delta 1$ of Appendix B, we calculate the numbers in the third row of Table 5.2.1 and see that not one of the criteria is met. The fourth row shows that even with 4 terminals, at 4 different locations, the specification on 90th percentile queueing time fails. Thus the branch manager is right; it requires 5 terminals, if they are placed in 5 separate locations, to meet all the criteria. (It is not nice to fool your branch manager!)

TABLE 5.2.1
Summary of Calculations for Example 5.2.6^a

System	ρ	W_q	$E[q \mid q > 0]$	$\pi_q(90)$	$P[q > 60]$
1 M/M/1	5/8	50	80	146.61	0.29523
1 M/M/2	5/16	3.25	21.82	8.67	0.00951
2 M/M/1's	5/16	13.64	43.64	49.72	0.07902
4 M/M/1's	5/32	5.56	35.56	15.87	0.02891
5 M/M/1's	1/8	4.29	34.29	7.65	0.02172

^a All times in minutes.

Providing a second terminal does not cut the waiting time for service in half, as intuition might suggest, but rather to one-fourth when the new terminal is installed at a different location, and to one-sixteenth when the new terminal is placed in close proximity to the previous one. A cynic may note that the average queueing time for those who must queue is only cut in half (from 80 minutes to 44 minutes) if the second terminal is remotely located and to one-fourth if it is in proximity to the original terminal. However, it is difficult to improve this queueing time because of the relatively long service time (time using the terminal) which is exponentially distributed. Thus an engineer arriving at a one-terminal facility which is in use has to queue an average of at least one 30-minute service time, even if no one else is waiting, because of the memoryless property of the exponential distribution.

The manager, armed with the information in Table 5.2.1, is in a position to make an informed decision as to how many terminals to provide and where to put them. For example, suppose the average driving time to the branch office is now 30 minutes, but each terminal could be reached in 20 minutes if a new terminal is located across town. Then the average round trip to do some computing when both terminals are in the branch office (1 M/M/2 system) is $30 + 3.25 + 30 + 30 = 93.25$ minutes. With two

separately located terminals (2 M/M/1 systems) it is $20 + 13.64 + 30 + 20 = 83.64$ minutes. However, round trips for engineers delayed, average about 112 minutes for the first system and 114 minutes for the second system; the latter system has a larger 90th percentile queueing time, and a larger (25.28% versus 9.2%) probability that an engineer must spend more than 1 hour in the office. On balance, the single M/M/2 system seems better.

Example 5.2.7 KAMAKAZY Airlines is planning a telephone reservations office. Each agent will have a reservations terminal and can service a typical caller, on the average, in 5 minutes, the time being exponentially distributed. Calls arrive randomly and the system will hold calls that arrive when no agent is free. Thirty-six calls per hour are expected during the peak period of the day, on the average. The three design criteria for the new office follow.

- (1) The probability a caller will find all agents busy should not exceed 0.1 (10%).
- (2) The average waiting time for those who must wait should be no greater than one minute.
- (3) Less than 5% of all callers should have to wait more than one minute for an agent.

How many agents (and terminals) should be provided? How will this system perform if the number of callers per hour is 10% higher than anticipated?

Solution The expected peak period average arrival rate, λ , is 36 calls per hour or $36/60 = 0.6$ calls per minute. Hence the traffic intensity is $\lambda E[s] = 0.6 \times 5 = 3$ erlangs. Thus a minimum of 4 agents (servers) are required to keep up with the inquiries. We seek the minimum c such that $C(c, 3) \leq 0.1$. Using Fig. 1 of Appendix C (the reader is warned that the use of such graphs induces vertigo in some people, including the author), we see that $C(6, 3)$ is very close to 0.1. Direct calculation using the formula for $C(c, u)$ in Table 5 of Appendix C or the APL function ERLANGAC of Appendix B, shows that $C(6, 3) = 0.0991$ while $C(5, 3) = 0.236$; six agents are required to satisfy the first design criterion. The formulas of Table 5 as implemented by the APL function MΔMΔC show that, for six agents, the average queueing time for callers delayed is 1.67 minutes. Thus six agents are not enough. Actually eight agents are required, since for seven servers, $E[q | q > 0] = 1.25$ minutes; for eight agents it is exactly 1 minute. With eight servers

$$P[q > 1] = C(8, 3)e^{-8(1-\rho)/5} = 0.00476,$$

so the final design criterion is also satisfied. If the peak traffic is 10% higher than anticipated, the probability that all eight agents are busy is 0.022, the average queueing time for callers delayed is 1.06 minutes, and 97.8% of them

will not have to wait at all. Thus the proposed system looks good, even if the traffic is slightly higher than estimated. Of course each agent is busy only three-eighths of the time during the peak hour—such is the price of good service. As shown by the figures in Table 5.2.2, with six agents only one of the design criteria is met and with 4 agents the performance is deplorable. Eight agents looks like a good choice. (We have shown that eight agents should be on duty during the peak period. More than this number may be needed to provide for coffee breaks, I/O breaks, etc., so that eight agents are available for duty.)

TABLE 5.2.2
Summary of Calculations for Example 5.2.7^a

c	ρ	$C(c, 3)$	$E[q]$	$E[q \mid q > 0]$	$P[q > 1]$
8	0.3750	0.0129	0.013	1.00	0.00476
7	0.4286	0.0376	0.047	1.25	0.01692
6	0.5000	0.0991	0.165	1.67	0.05441
5	0.6000	0.2362	0.5904	2.50	0.15830
4	0.7500	0.5094	2.5472	5.00	0.41709

^a All times in minutes.

Parzen [5] has suggested that an appropriate measure of effectiveness of a queueing system is the *customer loss ratio*, R , defined by

$$R = \frac{\text{average time spent by a customer waiting for service}}{\text{average time spent by a customer being served}}$$

$$= \frac{E[q]}{E[s]} = \frac{W_q}{W_s}. \quad (5.2.70)$$

For an M/M/ c system

$$R = \frac{C(c, u)}{c(1 - \rho)}. \quad (5.2.71)$$

Thus, for Example 5.2.7, if eight agents are provided, $R = 0.0026$; with six agents R increases to 0.0330, and for four agents R reaches the value of 0.509! The customer loss ratio for the original system of Example 5.2.6 (that is the system of Example 5.2.3) was 167% while the suggested M/M/2 system of Example 5.2.6 has an R value of 10.8%.

5.2.4 The M/M/ c/c Queueing System (M/M/ c Loss System)

This system is sometimes called the M/M/ c loss system because customers who arrive when all the servers are busy are not allowed to wait for

course, which becomes $W(t) = P[w \leq t] = P[s \leq t]$) are also true for the M/G/c/c queueing system, that is, only the average value of the service time is important. (Such queueing systems are called “robust” systems.) For a proof see Gross and Harris [6].

Example 5.2.8 The Sad Sack Clothing Company has decided to install a tie-line telephone system between its east coast and west coast facilities. A caller receives a busy signal if the call is dialed when all the lines are in use. An average of 105 calls per hour with an average length of 4 minutes are expected. Enough lines are to be provided to ensure that the probability of getting a busy signal will not exceed 0.005. How many lines should be provided? With this number of lines, how many will be in use, on the average, during the peak period? How many lines are required if the probability of a busy signal is not to exceed 0.01? What would the performance be with 10 lines?

Solution The traffic intensity u is $(105/60) \times 4 = 7$ erlangs. By Fig. 2 of Appendix C, it appears that 15 tie lines are required. Using the APL function BCU we find that $B(15, 7) = 0.00332$ while $B(14, 7) = 0.00714$, so 15 lines are required. With 15 lines, the average number in use, L , is

$$7(1 - 0.00332) = 6.97675.$$

The smallest c such that $B(c, 7) \leq 0.01$ is 14, so we save only one line if we double the allowed probability of a busy signal. With 14 lines the average number in use is 6.95. If only 10 tie-lines are provided, the probability of a busy signal is $B(10, 7) = 0.07874$ and the average number in use is $7(1 - 0.07874) = 6.4488$.

The formulas for the M/M/c/K queueing system can be derived much as they were for the M/M/c/c system and are given in Table 8 of Appendix C.

5.2.5 M/M/ ∞ Queueing System

No real life queueing system can have an infinite number of servers; what is meant, here, is that a server is immediately provided for each arriving customer. The state-transition diagram for this model is shown in Fig. 5.2.6. We can read off from the figure that

$$C_n = u^n/n!, \quad n = 1, 2, 3, \dots,$$

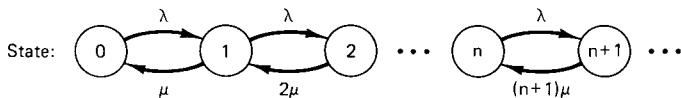


Fig. 5.2.6 State-transition diagram for M/M/ ∞ queueing system.

Then we apply Little's formula to obtain

$$W_q = L_q/\lambda = (E[O] + W_q + E[s])(L_q/K). \quad (5.2.96)$$

Solving (5.2.96) for W_q yields

$$W_q = L_q(E[O] + E[s])/(K - L_q). \quad (5.2.97)$$

The formulas for the $(M/M/c/K/K)$ queueing system are summarized in Table 11 of Appendix C. The formulas of this table apply to the single repairman model, also. In fact, they are arranged better for computation than those of Table 10 for the single repairman case. The calculations for the machine repair model using the formulas of Table 11 are implemented by the APL function MACHΔREP. In Table 25 of Appendix C we give the formulas for the machine repair model $D/D/c/K/K$; that is, with constant operating time and constant repair time.

Example 5.2.11 A company has 6 computer systems at each of its three computer centers for a total of 18 systems. The operating time of each computer is exponentially distributed with a mean time between failure of 30 hours. All computers are in continuous operation when not undergoing repairs. A customer engineering team is always on duty at each site and can repair a computer system in an average time of 30 minutes, the repair time being exponentially distributed. The company is considering consolidating all computers at one location with two customer engineering teams to maintain the equipment. The maintenance requirements are that the probability of no customer engineering team being free when a system goes down should not exceed 0.1 (10%), that the average down time per machine should not exceed 35 minutes, and the average waiting time for repairs to begin, for those systems with a delay, should not exceed 27 minutes. Will two customer engineering teams suffice?

Solution The calculation of the probabilities p_n , $n = 0, 1, \dots, 6$ are summarized in Table 5.2.3. The values displayed were calculated by the APL function MACHΔREP using the formulas from Table 11 and are shown accurate to five decimal places. As an example of how the calculations could be made by hand or with a desk calculator

$$\frac{p_2}{p_0} = 2! \binom{6}{2} \left(\frac{30}{1800}\right)^2 = \frac{2 \times 6!}{2! 4!} \left(\frac{1}{60}\right)^2 = \frac{30}{3600} = \frac{1}{120} = 0.0083333.$$

The probabilities from Table 5.2.3 can be used to make other calculations using the formulas of Table 11. We summarize the statistics for each of the 3 $M/M/1/6/6$ systems and the one $M/M/2/18/18$ system in Table 5.2.4.

It is clear that all three of the criteria are met by the proposed consol-

TABLE 5.2.3
Calculation of Probabilities for
Example 5.2.11 with One
Customer Engineering Team for
Six Computer Systems

<i>n</i>	p_n/p_0	p_n
0	1.0	0.90178
1	0.1	0.090178
2	0.0083333	0.0075148
3	0.00055556	0.00050099
4	2.7778×10^{-5}	2.5049×10^{-5}
5	9.2593×10^{-7}	8.3498×10^{-7}
6	1.5432×10^{-8}	1.3916×10^{-8}

TABLE 5.2.4
Summary of Calculations for Example 5.2.11

	Present system (each site)	Proposed consolidated system
L_q	0.0085954	0.005465
L	0.10682	0.30046
W_q	2.6253 minutes	0.55578 minutes
$E[q q > 0]$	26.729 minutes	15.311 minutes
W	32.625 minutes	30.556 minutes
$P[\text{machine } n \text{ down}]$	0.017803	0.016692
$P[\text{waiting for repair}]$	0.09822	0.0363

idated system. It is superior to the present system on all the statistics. For example, with the present system $3 \times 0.10682 \times 24 = 7.69$ hours of computer time are lost due to down time each 24 hours. With the proposed system $0.30046 \times 24 = 7.21$ hours of computer time are lost each 24 hours.

5.3 EMBEDDED MARKOV CHAIN QUEUEING SYSTEMS

In Section 5.2 we showed how, in some cases, a queueing system can be modeled as a birth-and-death process. This makes it relatively easy to calculate the steady state distribution of number of customers in the system and other performance measures; these include average queueing time, average waiting time in the system, etc. One fact makes analysis of such systems straightforward; it is that the stochastic process $\{N(t), t \geq 0\}$ is Markov. This is true because of the memoryless property of the exponential service

Exercises

1. [HM18] (a) Show that, for an M/M/1 queueing system, the probability there are n or more customers in the system is ρ^n .
 (b) Use this result to find the value of μ such that, for given values of λ , n and α , with $0 < \alpha < 1$, the probability of n or more customers in the system is α . This value of μ should be found explicitly in terms of λ , n and α .
 (c) In particular find μ if $\lambda = 10$, $n = 3$, and $\alpha = 0.05$.
2. [HM20] Show that for an M/M/1 queueing system the conditional density function for waiting time in the queue given that a wait occurs ($q > 0$), say q' , is given by $(\mu - \lambda)e^{-(\mu - \lambda)t}$, $t > 0$. Use this formula to calculate the conditional distribution function

$$P[q \leq t | q > 0] = P[q' \leq t] = 1 - e^{-\mu(1-\rho)t}.$$

Thus q' has the same distribution as w .

[Hint: The conditional density function of q given that $q > 0$, that is, q' , is the density function of q divided by the probability that $q > 0$.]

3. [18] Show that for an M/M/1 queueing system $E[N_q | N_q > 0] = 1/(1 - \rho)$.
4. [HM20] Show that, for an M/M/1 queueing system,

$$\text{Var}[N_q] = \rho^2(1 + \rho - \rho^2)/(1 - \rho)^2.$$

[Hint: Apply Theorem 2.9.2.]

5. [18] The BRITE LITE company has machines which break down in a Poisson pattern at an average rate of three per hour during the 8-hour working day. BRITE LITE is considering the repair services of I. M. Slow and I. M. Fast. Slow repairs machines with an exponential repair time distribution at an average rate of four machines per hour for a service charge of \$10.00 per hour. Fast provides exponential repair time for \$18.00 per hour but at an average rate of six machines per hour. Which person should be hired on a daily basis if the cost of an idle machine is \$30.00 per hour?

6. [15] People arrive at a telephone booth in a random pattern, with an average interarrival time of 12 minutes. The length of phone calls from the booth, including the dialing time, wrong numbers, etc., is exponentially distributed with an average of 4 minutes.

(a) What is the probability that a person arriving at the booth will have to wait? Do not assume your mother-in-law is in the booth.

(b) What is the average length of the waiting lines that form from time to time; that is, the average of those that are not of zero length?

(c) What is the probability that an arrival will have to wait more than 10 minutes before the phone is available?

(d) The telephone company plans to install a second booth when convinced that an arriving customer would expect to have to wait at least five minutes to use the phone. At what average interarrival time will this occur?

7. [20] A clerk provides exponentially distributed service to customers who arrive randomly at the average rate of 15 per hour. What average service time must

the clerk provide in order that 90% of all customers will not have to queue for service longer than 12 minutes.

[Hint: A graphical or iterative technique is necessary.]

8. [15] Consider an M/M/1/K queueing system. Let q_n be the probability that there are n customers in the system just before a customer arrival who actually enters the system. Assume that $q_n = kp_n$ for some constant k . Prove that $q_n = p_n/(1 - p_K)$, $n = 0, 1, \dots, K - 1$.

9. [HM25] Carry out the details of (5.2.48), that is, show that

$$W(t) = P[w \leq t] = 1 - \sum_{n=0}^{K-1} q_n \left(\sum_{k=0}^n e^{-\mu t} \frac{(\mu t)^k}{k!} \right)$$

for the M/M/1/K queueing system, where $q_n = p_n/(1 - p_K)$.

[Hint: Write

$$\begin{aligned} W(t) &= \sum_{n=0}^{K-1} \left[\int_0^t \frac{\mu(\mu x)^n e^{-\mu x}}{n!} dx \right] q_n = \sum_{n=0}^{K-1} \left[1 - \int_t^\infty \frac{\mu(\mu x)^n e^{-\mu x}}{n!} dx \right] q_n \\ &= 1 - \sum_{n=0}^{K-1} q_n \int_t^\infty \frac{\mu(\mu x)^n}{n!} e^{-\mu x} dx. \end{aligned}$$

Then make the change of variable $y = x - t$ in each of the integrals. By recognizing the integral form of the gamma function (see formula (3.2.35)) and the property of the gamma function expressed in (3.2.36), deduce that

$$\int_t^\infty \frac{\mu(\mu x)^n}{n!} e^{-\mu x} dx = \sum_{k=0}^n e^{-\mu t} \frac{(\mu t)^k}{k!}, \quad n = 0, 1, \dots, K - 1.]$$

10. [HM18] Prove that, for an M/M/c queueing system, the average number of busy servers is $u = \lambda E[s]$.

11. [15] Martin [7, p. 461] claims that, for an M/G/c system, one can approximate the average waiting time in queue, $W_q = E[q]$, by

$$W_q = \frac{C(c, u)E[s]}{c(1 - \rho)} \left\{ \frac{1 + C_s^2}{2} \right\}$$

where, of course, $C_s^2 = \text{Var}[s]/E[s]^2$. W can then be approximated by $W_q + E[s]$. Consider Example 5.2.7. Suppose KAMAKAZY Airlines installs the new reservations office with 8 agents and the system performs even better than expected. Each agent has an Erlang-3 service time distribution with average value 5 minutes. The OR department estimates that just before the holidays the peak calling rate may go up to 64.8 calls per hour. Use Martin's estimate to calculate W_q and W for the increased traffic. Note that Martin's estimate is a special case of the Allen-Cunneen approximation.

12. [12] The data processing manager at a certain company provides three consultants to help open-shop programmers debug their programs. Programmers with "buggy" programs arrive randomly, at an average rate of 20 per 8-hour day. The amount of time that a consultant spends with a programmer has an exponential

distribution with an average value of 40 minutes. Programmers are assigned to consultants in the order of their arrival.

(a) What is the average number of hours, per 40 hour week, that each consultant spends with programmers seeking help?

(b) What is the average amount of time a programmer spends in the consulting facility?

13. [C20] The WEIRDOENGINEER Company of Examples 5.2.3 and 5.2.6 installs five terminals at customer locations about town and finds the distribution of users and average driving times as shown in Table E5.13. (The "turnpike effect" has caused the number of users to rise to an average of 30 per day.) Assuming that the average time at a terminal is 30 minutes calculate the following.

(a) W_q , W , and $\pi_q(90)$ for each terminal.

(b) The (weighted) average values of W_q , W , and $\pi_q(90)$ over all the terminals.

(c) The average value of time required for an engineer to drive to the assigned terminal, complete a work session, and return (the average total time, that is).

TABLE E5.13

Terminal number	Average number of users per day	Average driving time per user (minutes)
1	6	2
2	8	5
3	4	4
4	10	1
5	2	10

14. [25] Customers arrive randomly (during the evening hours) at the Kittenhouse, the local house of questionable services, at an average rate of five per hour. Service time is exponential with a mean of 20 minutes per customer. There are two servers on duty.

(a) What is the probability an arriving customer must queue?

(b) That one or both servers are idle?

(c) What is the average time a customer spends at the Kittenhouse?

(d) If the house is raided, how many customers will be caught, on the average?

(e) What is the probability that five or more customers will be captured in a raid?

(The data for this problem were conjectured by the author. Observed data from readers would be appreciated.)

15. [18] JETSET Airlines, a fierce competitor of KAMAKAZY Airlines (Example 5.2.7), also is planning a new telephone reservations office. Their agents provide customers who call with an exponential service time; the average time is 3 minutes. Like KAMAKAZY Airlines, calls that arrive when all agents are busy are held (with appropriate background music) until an agent is free. They expect a random pattern of customer calls with an average of 30 calls per hour during the peak period.

(a) If the two criteria are (1) that the average queueing time should not exceed 1 minute and (2) that 90% of all callers must wait less than 2 minutes for service to begin, how many agents are required?

(b) With the number of agents determined by part (a), what is the average queueing time, and the average number of customers waiting for service.

(c) What is the probability that all the agents are busy during the peak period? That all are idle?

16. [C15] YOUTOOLCOMPUTE has 10 portable computers available for rent. The average rental time is 2.5 days and is exponentially distributed. Customers arrive randomly at an average rate of two customers per day. If a computer is not available, a customer will go to another store.

(a) What fraction of arriving customers will be lost?

(b) What is the average number of computers in use?

(c) What fraction of customers will be lost if one of the computers is out of service for an extended period?

17. [15] The SUPERCOMPUTER Company offers computer service bureau services to drive-in customers. Twelve customer parking spaces are provided for customers who arrive randomly at the average rate of 14 per hour; those who arrive when all spaces are in use take their business across the street to the SUPER-DUPERCOMPUTER Company. Each parking space is occupied for an average of 30 minutes, occupancy time having an exponential distribution. Find the following.

(a) The effective average arrival rate.

(b) The fraction of arriving customers turned away.

(c) The average number of spaces in use.

18. [HM25] Use the Cauchy–Schwartz inequality (stated below) to show that, if X has a k -stage hyperexponential distribution, then $C_X^2 \geq 1$. The Cauchy–Schwartz inequality asserts that, if a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n are real, then

$$\left(\sum_{i=1}^n a_i b_i \right)^2 \leq \left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{i=1}^n b_i^2 \right).$$

19. [18] Use Theorem 5.3.1 and the Corollary to show that formulas (5.3.42)–(5.3.45) are true.

20. [20] Prove that, for the M/G/1 queueing system,

$$\sigma_N^2 = \frac{\lambda^3 E[s^3]}{3(1-\rho)} + \left(\frac{\lambda^2 E[s^2]}{2(1-\rho)} \right)^2 + \frac{\lambda^2 (3-2\rho) E[s^2]}{2(1-\rho)} + \rho(1-\rho).$$

[Hint: Use (5.3.29) to show that

$$E[N(N-1)] = E[N^2] - E[N] = \lambda^2 E[w^2].$$

Then calculate $E[N^2]$ from this formula, (5.3.45), and Little's formula.]

21. [HM20] Show that, for an M/M/c queueing system,

$$\sigma_{N_q}^2 = \frac{\rho C(c, u)[1 + \rho - \rho C(c, u)]}{(1-\rho)^2}.$$

[Hint: Apply Theorem 2.9.2.]

22. [HM15] Show that, if the service time, s , has a k -stage hyperexponential distribution, then

$$E[s^3] = 6 \sum_{i=1}^k \frac{\alpha_i}{\mu_i^3}.$$

23. [15] Repairing a small computer requires four steps in sequence. The time to complete each of these steps is exponentially distributed with a mean time of 3 minutes; the steps being independent of one another. If a facility has one person who can repair these computers and they break down in a Poisson pattern at an average rate of three per hour, what is the average down time of a computer?

24. [HM20] Suppose the service time has a gamma distribution with parameters α and β . (Since $E[s] = 1/\mu$, this means $\alpha/\beta = 1/\mu$ or $\alpha = \beta/\mu$.) Show that $E[s] = 1/\mu$, $E[s^2] = E[s]^2 + 1/\beta\mu$, $\text{Var}[s] = 1/\beta\mu$,

$$C_s^2 = \mu/\beta \quad \text{and} \quad E[s^3] = (\beta^2 + 3\beta\mu + 2\mu^2)/\beta^2\mu^3.$$

Then verify the equations of Table 13, Appendix C.

[Hint: By Table 2 of Appendix A, $\psi(\theta) = (\beta/\beta - \theta)^\alpha$, and, by Theorem 2.9.1,

$$E[s^k] = \left. \frac{d^k \psi}{d\theta^k} \right|_{\theta=0}.$$

25. [12] Apply the results of Exercise 24 to an Erlang- k service time, which is a special case of the gamma distribution with $\alpha = k$ and $\beta = k\mu$, to obtain

$$\begin{aligned} E[s] &= \frac{1}{\mu}, & E[s^2] &= \frac{(k+1)}{k} \frac{1}{\mu^2} = \frac{(k+1)}{k} E[s]^2, \\ \text{Var}[s] &= 1/k\mu^2 = E[s]^2/k, & C_s^2 &= 1/k, \\ E[s^3] &= \frac{(k+1)(k+2)}{k^2} E[s]^3. \end{aligned}$$

Then verify the equations of Table 14, Appendix C.

26. [C20] Suppose four communication lines are connected to one computer and that each line has an average incoming message transmission time of 2 seconds with a utilization of 0.6. The message transmission time is gamma with parameters $\alpha = 1/3$, $\beta = 1/6$ for the first line; it is exponential for the second line. The third line has an Erlang-3 service time, while it is constant for the fourth line. Calculate W_q , W , σ_w , L_q , and L for each line. Then estimate $\pi_w(90)$ by

- (a) Martin's estimate and
- (b) Tables 5.3.1 and 5.3.2.

27. [HM20] Show that, for the GI/M/1 queueing system, given that

$$P[q \leq t] = 1 - (1 - \pi_0)e^{-t/W} = 1 - (1 - \pi_0)e^{-\pi_0 t/E[s]},$$

then

- (a) $E[q] = (1 - \pi_0)E[s]/\pi_0$,
- (b) $E[q^2] = 2(1 - \pi_0)(E[s]/\pi_0)^2$,
- (c) $\text{Var}[q] = (1 - \pi_0^2)(E[s]/\pi_0)^2$.

28. [HM20] If $\pi(z)$ is given by (5.3.16), show that

$$L = \pi'(1) = \rho + \frac{K''(1)}{2(1 - \rho)}.$$

29. [12] Consider Example 5.2.10. Suppose that all the parameters are the same as given except that RPS is not used. Find the maximum number of inquiries per second that the channel can handle. Use a read time of 6 milliseconds.

30. [C18] Suppose 30 buffered terminals on one communication line are used for data entry to a computer system. (Buffered terminals are those in which entries are first keyed into a local memory and then transmitted as one message when the line is free to accept it.) The average time to key in an entry is 60 seconds; the keying time has an exponential distribution. The average system response time (line, both ways, plus computer time) is 2 seconds; the response time is also exponential.

(a) Find the line utilization, the mean rate entries can be processed (λ), and the average queueing time for an entry (time spent waiting for the line to become free). If you have APL available you should compare your solution using MACHΔREP or direct calculation to that obtained by Fig. 3 of Appendix C.

- (b) Repeat the calculations for 40 terminals on the line.
 (c) For 50 terminals on the line.

31. [HM10] Suppose the interarrival time has an Erlang- k distribution. Show that the Laplace-Stieltjes transform of τ , $A^*(\theta)$, is given by

$$A^*(\theta) = \left(\frac{k\lambda}{k\lambda + \theta} \right)^k, \quad \theta < k\lambda.$$

[Hint: Recall that an Erlang- k random variable with parameters λ can be represented as the sum of k independent, exponential random variables, each with parameter $k\lambda$.]

32. [HM15] Show that if the interarrival time in a GI/M/1 queueing system is uniformly distributed over the interval 0 to $2/\lambda$, then the Laplace-Stieltjes transform of τ , $A^*(\theta)$, is

$$A^*(\theta) = (\lambda/2\theta)(1 - e^{-2\theta/\lambda}).$$

33. [HM18] Consider the interarrival time distribution for which the π_0 values are shown in the next to last column of Table 17, Appendix C. That is, a two-stage hyperexponential distribution with (in the notation of Section 3.2.9) $\alpha_1 = 0.4$, $\alpha_2 = 0.6$, $\mu_1 = 0.5\lambda$, and $\mu_2 = 3\lambda$. Show that $E[\tau] = 1/\lambda$ and (5.3.50) for π_0 reduces to

$$\pi_0^2 + (3.5\rho - 1)\pi_0 + 1.5\rho(\rho - 1) = 0.$$

This implies that the unique value of π_0 between 0 and 1 is given by

$$\pi_0 = 0.5 - 1.75\rho + (1.5625\rho^2 - 0.25\rho + 0.25)^{1/2}.$$

34. [HM20] Show that, for a GI/M/1 queueing system

$$\text{Var}[N_q] = \frac{\rho(1 - \pi_0)\{2 - \pi_0 - \rho(1 - \pi_0)\}}{\pi_0^2}$$

[Hint: Use Theorem 2.9.2.]

35. [HM18] Show that, for a GI/M/1 queue,

$$\text{Var}[N] = \frac{\rho(2 - \pi_0 - \rho)}{\pi_0^2}.$$

36. [15] Suppose that in Exercise 6 the arrival pattern to the telephone booth is Erlang-2, rather than random, that the average interarrival time is 10 minutes, and that the length of phone calls is exponentially distributed with mean 3 minutes. Find (a) through (c) of Exercise 6.

37. [18] Consider Example 5.4.1. Recalculate the values of Table 5.4.1 under the assumption that, for both the HOL queueing system and the preemptive-priority queueing system, Type 2 inquiries will be given preference over Type 1 inquiries.

38. [15] Consider a communication line as a GI/G/1 queueing system in which $E[\tau] = 0.05$ seconds, $\text{Var}[\tau] = 0.003125$ seconds², $E[s] = 0.0475$ seconds, and $\text{Var}[s] = 0.001805$ seconds². Find $E[q]$, $E[w]$, $\pi_q(90)$, and the probability that the queueing time for a message will exceed 1.5 seconds. Check your results by approximating the system with a M/M/1 queueing system with the same average interarrival time and the same average service (transmission) time. Find the mean queueing time by the Allen–Cunneen approximation formula (5.5.6).

39. [12] Consider a D/M/1 queueing system with $E[\tau] = 2$ seconds and $E[s] = 1.6$ seconds. Apply Theorem 5.5.3 to compute upper and lower bounds for $E[q]$. Then calculate the exact value using Tables 16 and 17 of Appendix C.

40. [12] Consider an H₂/H₂/1 queueing system in which, for τ , $\alpha_1 = \alpha_2 = 0.5$, $1/\lambda_1 = 0.2$ seconds, and $1/\lambda_2 = 1.8$ seconds. Suppose, for s , $\alpha_1 = \alpha_2 = 0.5$ but $1/\mu_1 = 0.2$ seconds and $1/\mu_2 = 1.6$ seconds. Find upper and lower bounds for $E[q]$. Also compute the mean queueing time by the Allen–Cunneen approximation formula (5.5.6).

41. [10] Let $E[q_1]$ be the average queueing time for GI/G/1 queueing system and $E[q_2]$ the average queueing time for a GI/G/c queueing system where ρ , $E[s]$, C_s^2 , and C_τ^2 are the same for both systems. Show that, if ρ is less than but close to 1, so that the heavy traffic approximation applies, then

$$E[q_2] = E[q_1]/c.$$

42. [HM20] Suppose a two-stage hyperexponential distribution is generated using Algorithm 6.2.1 of Chapter 6. The distribution is to represent the interarrival time to a GI/M/1 queueing system with a given average interarrival time, $1/\lambda$, and a given squared coefficient of variation, $C^2 \geq 1$. Thus

$$\alpha_1 = \frac{1}{2} \left[1 - \left(\frac{C^2 - 1}{C^2 + 1} \right)^{1/2} \right],$$

$\alpha_2 = 1 - \alpha_1$, $\lambda_1 = 2\alpha_1\lambda$, and $\lambda_2 = 2\alpha_2\lambda$. Substitute these values into (5.3.74) and show that π_0 is given by the formula (5.3.76). Note that you must prove that, if $0 < \rho < 1$, then $0 < \pi_0 < 1$.

References

1. J. D. C. Little, A proof of the queueing formula: $L = \lambda W$, *Operations Res.* **9** (3), (1961), 383-387.
2. Lajos Takács, *Introduction to the Theory of Queues*. Oxford University Press, London and New York, 1962.
3. W. Feller, *An Introduction to Probability and Its Applications*, Vol. II, 2nd ed. Wiley, New York, 1971.
4. D. N. Streeter, Centralization or dispersion of computing facilities, *IBM Systems J.* **12** (3), (1973), 283-301.
5. E. Parzen, *Stochastic Processes*. Holden-Day, San Francisco, 1962.
6. D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. Wiley, New York, 1974.
7. J. Martin, *Systems Analysis for Data Transmission*. Prentice-Hall, Englewood Cliffs, New Jersey, 1972.
8. Analysis of some queueing models in real-time systems, IBM Report Number GF20-0007-1, IBM Data Processing Division, 1133 Westchester Avenue, White Plains, New York 10604.
9. S. Karlin, *A First Course in Stochastic Processes*. Academic Press, New York, 1969.
10. L. Kleinrock, *Queueing Systems, Volume I: Theory*. Wiley, New York, 1975.
11. L. Takács, A single-server queue with Poisson input, *Operations Res.* **10**, (1962), 388-397.
12. L. Kleinrock, *Queueing Systems, Volume II: Computer Applications*. Wiley, New York, 1976.
13. J. F. C. Kingman, On queues in heavy traffic, *J. Roy. Statist. Soc. Ser. B* **24**, (1962), 383-392.
14. J. Kölleström, Heavy traffic theory for queues with several servers. I, *J. Appl. Probability* **11**, (1974), 544-552.
15. W. G. Marchal, Some simple bounds and approximations in queueing, Technical Memorandum Serial T-294, Institute for Management Science and Engineering, The George Washington University, Washington, D.C., January 1974.
16. S. L. Brumelle, Bounds on the wait in a GI/M/k queue, *Management Sci.* **19** (7), (1973), 773-777.
17. T. Suzuki and Y. Yoshida, Inequalities for many-server queue and other queues, *J. Operations Res. Soc. Japan* **13**, (1970), 59-77.
18. J. F. C. Kingman, Inequalities in the theory of queues, *J. Roy. Statist. Soc. Ser. B* **32**, (1970), 102-110.
19. F. S. Hillier and F. D. Lo, Tables for multiserver queueing systems involving Erlang distributions, Tech. Rep. 31, December 28, 1971, Department of Operations Research, Stanford University, Stanford, California.
20. A. O. Allen, Elements of queueing theory for system design, *IBM Systems J.* **14** (2), (1975).
21. N. C. Wilhelm, A general model for the performance of disk systems, *J. ACM* **24** (1), (1977), 14-31.

Little's Law revisited

The Little's Law certainly applies to the M/M/1 queuing system and its components, the queue and the server. Assuming the system is functional ($r < 1$), all the jobs go through the entire system, and thus, each component is subject to the same arrival rate λ_A . The Little's Law then guarantees that

$$\begin{aligned}\lambda_A \mathbf{E}(R) &= \mathbf{E}(X), \\ \lambda_A \mathbf{E}(S) &= \mathbf{E}(X_s), \\ \lambda_A \mathbf{E}(W) &= \mathbf{E}(X_w).\end{aligned}$$

Using our results in (7.10), it wouldn't be a problem for you to verify all three equations, would it?

M/M/1:
main
performance
characteristics

$\mathbf{E}(R) = \frac{\mu_S}{1-r} = \frac{1}{\lambda_S(1-r)}$ $\mathbf{E}(W) = \frac{\mu_S r}{1-r} = \frac{r}{\lambda_S(1-r)}$ $\mathbf{E}(X) = \frac{r}{1-r}$ $\mathbf{E}(X_w) = \frac{r^2}{1-r}$ $\mathbf{P}\{\text{server is busy}\} = r$ $\mathbf{P}\{\text{server is idle}\} = 1-r$	(7.10)
---	--------

Example 7.4 (MESSAGE TRANSMISSION WITH A SINGLE CHANNEL). Messages arrive to a communication center at random times with an average of 5 messages per minute. They are transmitted through a single channel in the order they were received. On average, it takes 10 seconds to transmit a message. Conditions of an M/M/1 queue are satisfied. Compute the main performance characteristics for this center.

Solution. The arrival rate $\lambda_A = 5 \text{ min}^{-1}$ and the expected service time $\mu_S = 10 \text{ sec or } (1/6) \text{ min}$ are given. Then, the utilization is

$$r = \lambda_A / \lambda_S = \lambda_A \mu_S = \underline{5/6}.$$

This also represents the proportion of time when the channel is busy and the probability of a non-zero waiting time.

The average number of messages stored in the system at any time is

$$\mathbf{E}(X) = \frac{r}{1-r} = \underline{5}.$$

Out of these, an average of

$$\mathbf{E}(X_w) = \frac{r^2}{1-r} = \underline{4.17}$$

messages are waiting, and

$$\mathbf{E}(X_s) = r = \underline{0.83}$$

are being transmitted.

When a message arrives to the center, its waiting time until its transmission begins averages

$$\mathbf{E}(W) = \frac{\mu_S r}{1 - r} = \underline{50 \text{ seconds}},$$

whereas the total amount of time since its arrival until the end of its transmission has an average of

$$\mathbf{E}(R) = \frac{\mu_S}{1 - r} = \underline{1 \text{ minute}}.$$

◊

Example 7.5 (FORECAST). Let's continue Example 7.4. Suppose that next year the customer base of our transmission center is projected to increase by 10%, and thus, the intensity of incoming traffic λ_A increases by 10% too. How will this affect the center's performance?

Solution. Recompute the main performance characteristics under the new arrival rate

$$\lambda_A^{\text{NEW}} = (1.1)\lambda_A^{\text{OLD}} = 5.5 \text{ min}^{-1}.$$

Now the utilization equals $r = 11/12$, getting dangerously close to 1 where the system gets overloaded. For high values of r , various parameters of the system increase rapidly. A 10% increase in the arrival rate will result in rather significant changes in other variables. Using (7.10), we now get

$$\begin{aligned}\mathbf{E}(X) &= 11 \text{ jobs}, \\ \mathbf{E}(X_w) &= 10.08 \text{ jobs}, \\ \mathbf{E}(W) &= 110 \text{ seconds, and} \\ \mathbf{E}(R) &= 2 \text{ minutes.}\end{aligned}$$

We see that the response time, the waiting time, the average number of stored messages, and therefore, the average required amount of memory more than doubled when the number of customers increased by a mere 10%. ◊

When a system gets nearly overloaded

As we observed in Example 7.5, the system slowed down significantly as a result of a 10% increase in the intensity of incoming traffic, projected for the next year. One may try to forecast the two-year future of the system, assuming a 10% increase of a customer base each year. It will appear that during the second year the utilization will exceed 1, and the system will be unable to function.

What is a practical solution to this problem? Another channel or two may be added to the center to help the existing channel handle all the arriving messages! The new system will then have more than one channel-server, and it will be able to process more arriving jobs.

Such systems are analyzed in the next section.

$$F_n(x) = p_n e^{-\lambda x}, \quad n = 1, 2, 3, \dots \quad (4.2.36)$$

Thus we get

$$\begin{aligned} F(x) &= \sum_{n=0}^{\infty} p_n e^{-\lambda x} \\ &= e^{-\lambda x}, \end{aligned} \quad (4.2.37)$$

which is the same as the distribution of the interarrival times. Since $\{p_n\}$ is also the distribution of the number of customers in the system at departure points, equation (4.2.36) also confirms the independence of the distribution of T from the queue length distribution at departure points. Note that here we are talking about the independence of distribution of two random variables and not any relationship between their specific values. For a more exhaustive treatment of this problem, see Burke (1956), who has considered this problem for the multiserver $M/M/s$ queue.

The important result from this analysis states that the departure process of the $M/M/1$ queue in equilibrium is the same Poisson as the arrival process. Consequently, the expected number of customers served during a length of time t when the system is in equilibrium is given by λt .

Example 4.2.1. An airport has a single runway. Airplanes have been found to arrive at the rate of 15 per hour. It is estimated that each landing takes 3 minutes. Assuming a Poisson process for arrivals and an exponential distribution for landing times, use an $M/M/1$ model to determine the following performance measures.

(a) Runway utilization:

$$\begin{aligned} \text{arrival rate} &= 15/\text{hour } (\lambda), \\ \text{service rate} &= \frac{60}{3}/\text{hour} = 20/\text{hour } (\mu), \\ \text{utilization} &= \rho = \frac{\lambda}{\mu} = \frac{3}{4}. \end{aligned} \quad \text{ANSWER}$$

(b) Expected number of airplanes waiting to land:

$$L_q = \frac{\rho^2}{1 - \rho} = \frac{(0.75)^2}{0.25} = 2.25. \quad \text{ANSWER}$$

(c) Expected waiting time:

$$E(W_q) = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{15}{20(20 - 15)} = \frac{3}{20} \text{ hour} = 9 \text{ minutes.} \quad \text{ANSWER}$$

(d) Probability that the waiting will be more than 5 minutes? 10 minutes? No waiting?

$$P(\text{no waiting}) = P(T_q = 0) = 1 - \rho = .25, \quad \text{ANSWER}$$

$$P(T_q > t) = \rho e^{-\mu(1-\rho)t},$$

$$\begin{aligned} P(T_q > 5 \text{ minutes}) &= \frac{3}{4} e^{-20(1-\frac{3}{4})5/60} \\ &= \frac{3}{4} e^{-\frac{25}{60}} = 0.4944, \end{aligned} \quad \text{ANSWER}$$

$$P(T_q > 10 \text{ minutes}) = \frac{3}{4} e^{-\frac{50}{60}} = 0.3259. \quad \text{ANSWER}$$

$$(e) \text{ Expected number of landings in a 20-minute period} = \frac{15}{60} \times 20 = 5. \quad \text{ANSWER}$$

4.3 The Queue $M/M/s$

The multiserver queue $M/M/s$ is the model used most in analyzing service stations with more than one server such as banks, checkout counters in stores, check-in counters in airports, and the like. The arrival of customers is assumed to follow a Poisson process, and service times are assumed to have an exponential distribution. We will let the number of servers be s , providing service independently of each other. We also assume that the arriving customers form a single queue and the one at the head of the waiting line enters into service as soon as a server is free. No server stays idle as long as there are customers to serve.

Let λ be the arrival rate and μ the service rate. (This means that the interarrival times and service times have exponential distributions with densities $\lambda e^{-\lambda x}$ ($x > 0$) and $\mu e^{-\mu x}$ ($x > 0$), respectively.) Note that the service rate μ is the same for all servers. In order to use the birth-and-death model introduced earlier, we have to establish values for λ_n and μ_n , when there are n customers in the system. Clearly, the arrival rate does not change with the number of customers in the system (i.e., λ is the constant arrival rate). What about μ_n , and how does it change?

Suppose n ($n = 1, 2, \dots, s$) servers are busy at time t . Then during $(t, t + \Delta t]$, the event that a busy server will complete service has the probability $\mu \Delta t + o(\Delta t)$. Since there are n busy servers at t , the probability that any one of the n busy servers will complete service during $(t, t + \Delta t]$ can be determined using the binomial probability distribution as

$$\begin{aligned} &= \binom{n}{1} [\mu \Delta t + o(\Delta t)] [1 - \mu \Delta t + o(\Delta t)]^{n-1} \\ &= n \mu \Delta t + o(\Delta t). \end{aligned} \quad (4.3.1)$$

Note that $\frac{o(\Delta t)}{\Delta t} \rightarrow 0$ as $\Delta t \rightarrow 0$.

In a similar manner the probability that a number $r (> 1)$ of the busy servers will complete service during $(t, t + \Delta t]$ can be given as

$$\begin{aligned} &= \binom{n}{r} [\mu \Delta t + o(\Delta t)]^r [1 - \mu \Delta t + o(\Delta t)]^{n-r} \\ &= o(\Delta t). \end{aligned}$$

$$\begin{aligned} W_q &= \frac{1}{\mu} \left[\frac{\rho}{1-\rho} - \frac{K\rho^K}{1-\rho^K} \right], & \rho &\neq 1, \\ &= \frac{1}{2\mu}(K-1), & \rho &= 1, \end{aligned} \quad (4.4.13)$$

$$\begin{aligned} W &= \frac{1}{\mu} \left[\frac{1}{1-\rho} - \frac{K\rho^K}{1-\rho^K} \right], & \rho &\neq 1, \\ &= \frac{1}{2\mu}(K+1), & \rho &= 1, \end{aligned} \quad (4.4.14)$$

$$\begin{aligned} L_q &= \frac{\rho}{1-\rho} - \frac{\rho(1+K\rho^K)}{1-\rho^{K+1}}, & \rho &\neq 1, \\ &= \frac{K(K-1)}{2(K+1)}, & \rho &= 1, \end{aligned} \quad (4.4.15)$$

$$\begin{aligned} L &= \frac{\rho(1-\rho^K)}{(1-\rho)(1-\rho^{K+1})} - \frac{K\rho^{K+1}}{1-\rho^{K+1}}, & \rho &\neq 1, \\ &= \frac{K}{2}, & \rho &= 1. \end{aligned} \quad (4.4.16)$$

Note that in the simplifications leading to some of the results given above, we have used the formula

$$\sum_{n=1}^{K-1} n\rho^{n-1} = \frac{d}{d\rho} \left(\frac{1-\rho^K}{1-\rho} \right).$$

Example 4.4.1. A small mail-order business has one telephone line and a facility for call waiting for two additional customers. Orders arrive at the rate of one per minute and each order requires 2 minutes and 30 seconds to take down the particulars. Model this system as an $M/M/1/3$ queue and answer the following questions:

- (a) What is the expected number of calls waiting in the queue? What is the mean wait in queue?

Assuming that the arrivals are in a Poisson process with rate 1 per minute (λ) and the service times are exponential with mean 2.5 minutes ($1/\mu$), we have $\rho = 2.5$. Also, $K = 3$. Using the first result from (4.4.15), we get

$$\begin{aligned} L_q &= \frac{2.5}{1-2.5} - \frac{(2.5)[1+3(2.5)^3]}{1-(2.5)^4} \\ &= 1.4778. \end{aligned}$$

ANSWER

Since $\lambda = 1$, the mean waiting time in queue is

$$W_q = 1.4778 \text{ minutes.}$$

ANSWER

- (b) What is the probability that the call has to wait for more than 1.5 minutes before being served?

We use the formula for $1 - F_q(t)$ from (4.4.12) with $t = 1.5$, $1/\mu = 2.5$, and $\rho = 2.5$. We get

$$P(\text{wait in queue} > 1.5 \text{ minutes})$$

$$= \frac{1 - 2.5}{1 - (2.5)^3} \sum_{n=1}^{3-1} (2.5)^n \sum_{r=0}^{n-1} e^{-\frac{1.5}{2.5}} \frac{(1.5/2.5)^r}{r!}$$

$$= 0.7036.$$

ANSWER

- (c) Because of the excessive waiting time, the business decides to use two telephone lines instead of one, keeping the same total capacity for the number in the system, namely 3. What improvements result in the performance measures considered under (a) and (b)?

With two lines, now $s = 2$ and we have an $M/M/2/3$ system. Accordingly, in (4.4.3) we have $\alpha = 2.5$, $\rho = 1.25$, and $s = 2$ and $K = 3$. We get

$$\begin{aligned} p_0 &= 0.0950, & p_1 &= 0.2374, \\ p_2 &= 0.2969, & p_3 &= 0.3711. \end{aligned}$$

Using these results in (4.4.6), (4.4.9), and (4.4.5), we get

$$W_q = 0.5902 \text{ minute};$$

ANSWER

$$L_q = \lambda(1 - p_3)W_q = 0.3712;$$

ANSWER

$P(\text{wait in queue} > 1.5 \text{ minutes}):$

$$1 - F_q(1.5) = 0.1422.$$

ANSWER

- (d) What is the impact of increasing the capacity to four customers in the system?

Now we have an $M/M/2/4$ queue. Using the formulas as in (c), we get

$$\begin{aligned} p_0 &= 0.0649, & p_1 &= 0.1622, \\ p_2 &= 0.2028, & p_3 &= 0.2535, \\ p_4 &= 0.3169, \\ W_q &= 1.2989 \text{ minutes}; & & \text{ANSWER} \\ L_q &= 0.8873; & & \text{ANSWER} \end{aligned}$$

$P(\text{wait in queue} > 1.5 \text{ minutes}):$

$$1 - F_q(1.5) = 0.3353.$$

ANSWER

It is instructive to note that the performance has not improved from the viewpoint of the customer, because the system now accepts more customers than before. But from the management perspective fewer customers are being denied access to the system ($p_4 = 0.3169$ vs. $p_3 = 0.3711$).

$$\begin{aligned} F(x) &= \sum_{n=0}^{\infty} p_n e^{-\lambda x} \\ &= e^{-\lambda x}. \end{aligned} \quad (4.3.22)$$

(See Burke (1956) for details; see also Reich (1965).)

Example 4.3.1. In the airport problem of Example 4.2.1, how would the performance measures change if there are two runways while assuming the same arrival and service rates?

(a) Runway utilization:

$$\begin{aligned} \text{arrival rate} &= 15/\text{hour } (\lambda), \\ \text{service rate} &= 20/\text{hour } (\mu), \\ \text{number of servers} &= 2 (s), \\ \text{utilization of each runway} &= \rho = \frac{\lambda}{s\mu} = \frac{3}{8}. \end{aligned} \quad \text{ANSWER}$$

(b) Expected number of airplanes waiting to land:

$$L_q = \frac{\rho \alpha^s p_0}{s!(1-\rho)^2}$$

(note that $\alpha = s\rho = \frac{3}{4}$),

$$\begin{aligned} p_0 &= \left[\sum_{r=0}^1 \frac{\alpha^r}{r!} + \frac{\alpha^s}{s!(1-\rho)} \right]^{-1} \\ &= \left[1 + \frac{3}{4} + \frac{(\frac{3}{4})^2}{2} \left(1 - \frac{3}{8} \right)^{-1} \right]^{-1} \\ &= 0.4545, \\ L_q &= \left[\left(\frac{3}{8} \right) \left(\frac{3}{4} \right)^2 (0.4545) \right] / 2 \left(\frac{5}{8} \right)^2 \\ &= 0.1227. \end{aligned} \quad \text{ANSWER}$$

(c) Expected waiting time:

$$\begin{aligned} W_q &= \frac{\alpha^s p_0}{s! s \mu (1-\rho)^2} \\ &= \left[\left(\frac{3}{4} \right)^2 (0.4545) \right] / 2 \times 2 \times 20 \left(1 - \frac{3}{8} \right)^2 \\ &= 0.00818 \text{ hour} = 0.49 \text{ minute.} \end{aligned} \quad \text{ANSWER}$$

(d) Probability that the waiting will be more than 5 minutes? 10 minutes? No waiting?

$$\begin{aligned} P(\text{no waiting}) &= F_q(0) = 1 - \frac{\alpha^s p_0}{s!(1-\rho)} \\ &= 1 - \frac{(\frac{3}{4})^2(0.4545)}{2(1-3/8)} \\ &= 0.7955; \end{aligned}$$
ANSWER

$$P(T_q > t) = \frac{\alpha^s p_0}{s!(1-\rho)} e^{-s\mu(1-\rho)t},$$

$$\begin{aligned} P(T_q > 5 \text{ minutes}) &= \frac{(\frac{3}{4})^2(0.4545)}{2(\frac{5}{8})} e^{-2(\frac{1}{3})(\frac{5}{8})5} \\ &= 0.1245; \end{aligned}$$
ANSWER

$$P(T_q > 10 \text{ minutes}) = 0.0155.$$
ANSWER

(e) Expected number of landings in a 20-minute period = $\frac{15}{60} \times 20 = 5$. ANSWER
 (The departure process is Poisson with parameter λ .)

Example 4.3.2. A bank has established two counters—one for commercial banking and the second for personal banking. Arrival and service rates at the commercial counter are 6 and 12 per hour, respectively. The corresponding numbers at the personal banking counter are 12 and 24, respectively. Assume that arrivals occur in Poisson processes and service times have exponential distributions.

(a) Assuming that the two counters operate independently of each other, determine the expected number of waiting customers and their mean waiting time at each counter. The results are listed in Table 4.3.1.

Table 4.3.1. Results from Example 4.3.2(a).

	<u>Commercial</u>	<u>Personal</u>	
λ	6/hour	12/hour	
μ	12/hour	24/hour	
$\rho = \frac{\lambda}{\mu}$	0.5	0.5	
$L_q = \frac{\rho^2}{1-\rho}$	0.5	0.5	ANSWER
$W_q = \frac{\rho}{\mu(1-\rho)}$	5 minutes	2.5 minutes	ANSWER

(b) What is the effect of operating the two queues as a two-server queue with arrival rate 18/hour and service rate 18/hour? What conclusion can you draw from this operation? See Table 4.3.2.

Table 4.3.2. Results from Example 4.3.2(b).

	Two-server queue	
λ	18/hour	
μ	18/hour	
Number of servers (s)	2	
$\rho = \frac{\lambda}{s\mu}$	0.5	
$\alpha = \frac{\lambda}{\mu}$	1	
$p_0 = \left[\sum_0^1 \frac{\alpha^r}{r!} + \frac{\alpha^2}{2(1-\rho)} \right]^{-1}$	0.4	
$L_q = \frac{\rho\alpha^2 p_0}{2(1-\rho)^2}$	0.4	ANSWER
$W_q = \frac{\alpha^2 p_0}{(2)2\mu(1-\rho)^2}$	1.33 minutes	ANSWER

Conclusion: The two-server queue operation is more efficient than the two single-server operations.

Incidentally, the efficiency of multiserver queues over single-server systems is the reason that multiserver service systems, whenever possible, use single waiting lines feeding multiple counters for service. Airline check-in counters and checkout counters in stores effectively operate this way because of jockeying among the waiting lines. (See Smith and Whitt (1981).)

4.4 The Finite Queue $M/M/s/K$

When the waiting room in a queueing system has a capacity limit we get a finite queue. In most situations, a finite queue occurs more naturally than a queue with a waiting room of infinite size. However, as the capacity limit gets larger, the behavior of the system approximates that of an infinite-capacity system, and in such cases we are justified in ignoring the size limit. A communication system with a finite buffer and several service channels is a good example of a finite queue.

Consider an s -server queueing system with Poisson arrivals, exponential service, and a capacity limit of K for the number in the system. Clearly, $K \geq s$. Assume that λ and μ are the arrival and service rates, respectively. These assumptions result in the following infinitesimal transition rates in the generalized birth-and-death queueing model:

$$\lambda_n = \lambda, \quad n = 0, 1, 2, \dots, K - 1,$$

4.8 Remarks

In this chapter, we have discussed only a few queueing systems for which generalized birth-and-death process models are suitable. We shall discuss a few more extended models in Chapters 6 and 7. There are many more examples in the queueing literature where such models have been effectively used. For instance, Syski (1960) has provided a large number of models for queueing systems applicable to the telephone industry. Further perusal of the telecommunication systems literature would reveal models developed since 1960.

There are other application areas, such as computer and manufacturing systems, where investigators use birth-and-death process models as a first line of attack in solving problems. The major advantages of these models are their Markovian structure (often leading to usable explicit results), and the ability to use numerical investigations without complex computational problems when explicit results are not forthcoming. After all, queueing models are approximate representations of real systems, and starting with a Markovian model provides a good starting point for an understanding of their approximate behavior.

4.9 Exercises

1. Compare the system idle time probability (p_0) in the three systems: (1) $M/M/s/s$, (2) $M/M/s$, and (3) $M/M/\infty$ and show that

$$p_0^{(1)} > p_0^{(2)} \quad \text{and} \quad p_0^{(3)} > p_0^{(2)}. \quad (4.9.1)$$

2. An airline employs two counters, one exclusively for first-class and business-class passengers and the other for coach-class passengers. The service times at both counters have been found to be exponential with mean 3 minutes. The coach-class passengers arrive at the rate of 18 per hour and the upper-class passengers arrive at the rate of 15 per hour. Is there any advantage in keeping the exclusivity of service in the counters? Answer this question using server utilization, mean number of customers in the system, and the mean waiting time, all in steady state.
3. A customer service counter has s telephone lines. Service requests arrive in a Poisson process with rate λ and the length of service is exponentially distributed with mean $1/\mu$. What is the probability that a request will encounter a busy system? What is the probability that a service request will arrive when the service center is busy?
4. Customer arrivals at a 7-Eleven is Poisson at the rate of 20 per hour. They can be assumed to spend an average of 12 minutes picking up merchandise, with the length of time having an exponential distribution. Two checkout counters provide service with a service rate of 15 per hour at each counter. We may also assume that the service times have an exponential distribution. Determine the limiting results for the following:

- (a) the distribution of the number of customers picking up merchandise and its mean;
 - (b) the mean length of time the customers wait at the counter for service;
 - (c) the mean total amount of time the customers spend in the store.
5. In a taxi stand there is space for only five taxicabs. Taxis arrive in a Poisson process with rate 12 per hour. If there is no waiting room, arriving taxis leave without passengers. Customers arrive at the taxi stand in a Poisson process once every 6 minutes on average.
- (a) Determine the limiting distribution of the number of customers waiting for taxis.
 - (b) What is the probability that there are taxis waiting for customers?
 - (c) Determine the mean waiting time for a customer.
6. An automobile service station has one station for oil and filter changes. On average the oil and filter change takes 7 minutes, the amount of time having an exponential distribution. Cars arrive in a Poisson process at the rate of 6 per hour. What is the probability that an arriving car has to wait more than 10 minutes to get served?
- What is the effect on the waiting time of adding another station with identical service characteristics? Determine the probability that the waiting time will be more than 5 minutes with two stations for oil and filter changes.
7. Customer arrivals to a service counter are in a Poisson process at the rate of 10 per hour. The service time distribution can be assumed to be exponential. Determine the minimum rate of service that would result in the customer waiting time being greater than 5 minutes with a probability of 0.10 or less.
8. In a manufacturing process production machines break down at the rate of 3 per hour. We may assume that the process of breakdowns is Poisson. The repair times of the machines can be assumed to have an exponential distribution. The repairs can be run at two rates: 4 per hour at a cost of \$20/hour and 5 per hour at a cost of \$30/hour. Considering the loss of productivity of the machines while they are either waiting for service or being in service, what is the minimum rate of productivity gain that would make it beneficial to provide service at the faster rate? You may assume an 8-hour workday in your calculations.
9. Customer arrivals to a store are in a Poisson process with a rate of 50 per hour. On average each customer spends 15 minutes in the store, and we assume that the time the customer spends in the store to have an exponential distribution. Currently, the store provides parking spaces for 15 cars. Overflow cars from the parking lot park elsewhere in the neighborhood. What is the probability that no parking space will be available if a customer were to arrive at some time? How many more spaces will be needed to make sure that the arriving customer will find parking space 99% of the time?
10. Suppose that the arrival and the service rates in Exercise 9 are changed to arrivals = 100 per hour and mean service time = 30 minutes. How many parking spaces

should be provided to make sure that the arriving customers will find parking space 99% of the time?

11. A single switchboard is used to direct calls coming to a doctor's office. The calls arrive in Poisson process at a rate of 15 per hour. Call holding times can be assumed to be exponential with a mean of 2 minutes. What is the probability that the calls will not have to wait for more than 2 minutes before getting to the receptionist?

Suppose it is decided to establish an upper limit K for the number of calls waiting such that the waiting time will be less than 2 minutes with a 90% probability. Determine K .

12. In the $M/M/s/s$ (loss system) show that in the long run,

$$L = \rho[1 - P_B], \quad (4.9.2)$$

where L = long-run expected number of customers in the system,

$$\rho = \frac{\text{arrival rate}}{\text{service rate}},$$

P_B = probability that an arriving customer is blocked from entering the system.

13. In a drugstore, customers arrive at the counter (with one server per counter) in a Poisson process at the rate of 48/hour. The service time can be assumed to be exponential with an average of 1 minute. The service is provided by one or more servers depending on the number of customers waiting or being served as follows:

0–4 customers	1 counter;
5–9 customers	2 counters;
10–14 customers	3 counters;
15 or more customers	4 counters.

Assume that this policy is used to increase or decrease the number of servers.

Determine the following:

- (a) What is the probability of system idleness?
- (b) How often would the store need more than one counter?
- (c) What is the average number of customers either waiting for service or being served?
- (d) What is the average waiting time in the queue?

14. The atmospheric quality at time t —denoted $A(t)$ —is measured by the number of pollutant units residing in the airshed at that time. These units are emitted from pollutant sources one unit at a time with rate α . The emission process can be assumed to be Poisson. Each unit thus emitted is diffused in an average time of length β . Also assume that the diffusion times are exponential random variables that are i.i.d. Obtain the mean and variance of $A(t)$ as $t \rightarrow \infty$.

15. (a) Writing $\beta = \frac{1}{\alpha} = \frac{\mu}{\lambda}$ in (4.6.4), show that, using s in place of M , p_0 from (4.6.4) can be expressed as

$$p_0 = (\beta^s / s!) / \left(\sum_{n=0}^s \frac{\beta^n}{n!} \right),$$

which is the probability of blocking in an $M/M/s/s$ system (see (4.4.20)).

- (b) Let λ^* be the effective arrival rate of machines for repair. Noting that λ^* can also be expressed as

$$\lambda^* = \frac{M}{(1/\lambda) + W_q + (1/\mu)}$$

show that the mean waiting time of a machine repair (waiting + service) is given by

$$W = \frac{M}{\lambda^*} - \frac{1}{\lambda}.$$

16. Ten terminals used for data entry in a hospital share a communication line. Terminals use the line on an FCFS basis and wait in a queue when the line is busy. It has been observed that the data entry job takes on average 100 seconds, and once the terminal is free, it is ready for the next job in 5 seconds on average. Determine the throughput rate (effective arrival rate λ^* of Exercise 15) and the mean response time W . (Total time for job completion = waiting + service.)
17. A computer system has s servers. Since each server can be accessed separately, each of the s servers can be considered a separate subsystem as well. The arrival of jobs to each server is Poisson with rate λ , and the service time is exponential with mean $1/\mu$. The main system operator would like to find out whether pooling resources would be advantageous in terms of response time (the amount of time the job spends in the system). With this objective consider the following three setups when $s = 3$:
- (a) Three separate systems.
 - (b) Arrivals are pooled into a single queue and processed separately as a multi-server queue.
 - (c) The arrivals are pooled as in (b). In addition, the servers are connected such that together they process jobs as a single server with rate 3μ .

Let W_i be the mean response time with the i th setup ($i = a, b, c$). Show that

$$W_a > W_b > W_c.$$

18. In a cyclic queue model of a single CPU and an I/O processor, the number of jobs in the system remains a constant N . After receiving service at the CPU, the job leaves the system with probability α and joins the I/O queue with probability $1 - \alpha$. Soon after a job leaves the system a new job is admitted to the CPU queue. The service times at the CPU and the I/O are exponential with means $1/\mu_1$ and

- $1/\mu_2$, respectively. Determine the limiting distribution of the number of jobs waiting and being served at the CPU queue. Also determine the mean time in system for a job (Coffman and Denning (1973)).
19. In a communication system, messages are transmitted through M identical channels. Messages are segmented for storage in fixed size buffers (bins). An individual message may require several buffers, but no buffer contains data from more than one message. When messages release the buffers from which they are transmitted, the buffers are ready for reuse.
- Assume that messages arrive in a Poisson process with rate λ . The messages are of length L , which is exponentially distributed with mean $1/\mu_L$. The transmission rate for the messages is R , so that the transmission time is exponential with mean $1/(R\mu_L)$.
- The data field size per buffer is b . Let N be the random variable representing the number of buffers in a message.
- Obtain the distribution of N .
 - Obtain the limiting probability that no message is present in the system.
 - Determine the distribution of the number of occupied buffers under statistical equilibrium and its mean and variance in terms of the limiting probability of no messages present in the system (Pedersen and Shah (1972)).
20. The following model describes a simplified representation of a multiprogramming system. Let the drum storage unit with a shortest-latency-time-first file drum, described in Exercise 10 of Chapter 1, be connected to a CPU with a fixed number of m tasks circulating in a closed system, alternately requesting service at the processor and the drum. Let μ_n be the service rate at the file drum unit as described in Exercise 10 of Chapter 1, and let λ be the service rate at the central processor. Let p_n , $n = 0, 1, 2, \dots, m$, be the limiting distribution of the queue length (including the one in service) at the file drum unit.
- Determine $\{p_n\}$ and the expected processor utilization for various values of m (which is known as the degree of multiprogramming) (Fuller (1980)).
21. A simplified model of the drum storage unit described in Exercise 20 assumes a Poisson arrival of requests for files with rate λ . Let the service rate μ_n be determined by the formula
- $$\frac{1}{\mu_n} = \frac{\tau}{n+1} + \frac{1}{\mu},$$
- where τ is the period of rotation and n is the number of requests in the system. Determine the mean waiting time of a request (Fuller (1980)).
22. In a time-shared computer system M terminals share a central processor. Let μ be the processing rate at the CPU, with the processing time having an exponential distribution. If a terminal is free at time t , the probability that it will initiate a job in the infinitesimal interval $(t, t + \Delta t]$ is $\lambda\Delta t + o(\Delta t)$, and it will continue to be free at $t + \Delta t$ with probability $1 - [\lambda\Delta t + o(\Delta t)]$.

- (a) Let $\{p_n\}$ be the probability distribution of the number of busy terminals as $t \rightarrow \infty$. Determine $p_n, n = 0, 1, 2, \dots, M$.
- (b) Show that in the long-run, the arrival rate at the CPU is given by

$$\frac{M\lambda}{1 + \lambda W},$$

where W is the mean response time (= mean waiting time of a job arriving at the terminal).

- (c) Equating the arrival rate with the departure rate from the processor show that the mean response time can be obtained as

$$\frac{M}{\mu(1 - p_0)} - \frac{1}{\lambda}$$

(Fuller (1980)).

23. Consider a two-server Markovian queue $M/M_i/2$, in which customer arrivals are in a Poisson process with parameter λ , and the service times of the two servers are distributed exponentially with rates $\mu_1 > \mu_2$. An arriving customer finding both servers free always chooses the faster server. But if there is only one server free when an arrival occurs, it enters service with the free server regardless of the service rate. If both servers are busy, the arriving customer waits in line for service in the order of arrival.

Determine the limiting distribution of the number of customers in the system.

Compare numerically the mean number of customers in the heterogeneous system $M/M_i/2$ with the corresponding homogeneous system $M/M/2$ when the service rate in the latter system is $(\mu_1 + \mu_2)/2$ (Singh (1970)).

24. Extend Exercise 23 to an $M/M_i/3$ heterogeneous queue and determine the limiting distribution of the number of customers in it. Also, carry out a numerical comparison of the mean number of customers in the systems between $M/M_i/3$ and $M/M/3$ when the service rate in the latter system is the average of the three heterogeneous rates (Singh (1971)).

3.1 Supplementary Problems

- 3-1 Messages arrive to a statistical multiplexing system according to a Poisson process having rate λ . Message lengths, denoted by \tilde{m} , are specified in octets, groups of 8 bits, and are drawn from an exponential distribution having mean $1/\mu$. Messages are multiplexed onto a single trunk having a transmission capacity of C bits per second according to a FCFS discipline.
- Let \tilde{x} denote the time required for transmission of a message over the trunk. Show that \tilde{x} has the exponential distribution with parameter $\mu C/8$.
 - Let $E[\tilde{m}] = 128$ octets and $C = 56$ kilobits per second (kb/s). Determine λ_{\max} , the maximum message-carrying capacity of the trunk.
 - Let \tilde{n} denote the number of messages in the system in stochastic equilibrium. Under the conditions of (b), determine $P\{\tilde{n} > n\}$ as a function of λ . Determine the maximum value of λ such that $P\{\tilde{n} > 50\} < 10^{-2}$.
 - For the value of λ determined in part (c), determine the minimum value of s such that $P\{\tilde{s} > s\} < 10^{-2}$, where \tilde{s} is the total amount of time a message spends in the system.
 - Using the value of λ obtained in part (c), determine the maximum value of K , the system capacity, such that $P_B(K) < 10^{-2}$.

Solution:

- (a) Since trunk capacity is C , the time to transmit a message of length m bytes is $x = 8m/C$. Therefore, $\tilde{x} = 8\tilde{m}/C$, or $\tilde{m} = \tilde{x}C/8$. Thus,

$$\begin{aligned} P\{\tilde{x} \leq x\} &= P\left\{\frac{8\tilde{m}}{C} \leq x\right\} \\ &= P\left\{\tilde{m} \leq \frac{xC}{8}\right\} \\ &= 1 - e^{\frac{-\mu C}{8}x}, \end{aligned}$$

where the last equation follows from the fact that \tilde{x} is exponential with parameter μ . Therefore, \tilde{m} is exponential with parameter $\mu C/8$.

- (b) Since $E[\tilde{m}] = 128$ octets, $\mu = \frac{1}{128}$. Therefore, \tilde{x} is exponentially distributed with parameter $\frac{\mu C}{8} = \frac{7}{128}$ Kbps, or $\frac{7}{128} \times 10^3$ bps. Then $E[\tilde{x}] = \frac{128}{7} \times 10^{-3}$ sec. Since $\lambda_{\max} E[\tilde{x}] = \rho < 1$,

$$\lambda < \frac{1}{E[\tilde{x}]}$$

$$\begin{aligned}
&= \frac{7}{128} \times 10^3 \\
&= 54.6875 \text{ msg/sec.}
\end{aligned}$$

- (c) We know from (3.10) that $P\{\tilde{n} > n\} = \rho^{n+1}$. Thus, $P\{\tilde{n} > 50\} = \rho^{51}$. Now, $\rho^{51} < 10^{-2}$ implies

$$51 \log_{10} \rho < -2.$$

i.e.,

$$\log_{10} \rho < -\frac{2}{51}.$$

Hence, $\rho = \lambda E[\tilde{x}] < 10^{-\frac{2}{51}} = 0.91366$, so that

$$\lambda < 49.966 \text{ msg/sec.}$$

- (d) Now, \tilde{s} is exponential with parameter $\mu(1 - \rho)$, so $P\{\tilde{s} > x\} = e^{-\mu(1-\rho)x}$. Then for $P\{\tilde{s} > s\} < 10^{-2}$, we must have

$$e^{-\mu(1-\rho)s} < 10^{-2}$$

or

$$-\mu [1 - \rho] s < -2 (\ln 10)$$

i.e.,

$$\begin{aligned}
s &> \frac{2 (\ln 10)}{\mu (1 - \rho)} \\
&= \frac{2 (\ln 10)}{\frac{7}{128} \times 10^3 (1 - 0.91366)} \\
&= \frac{4.60517}{4.7217} \\
&= 0.9753 \text{ sec} \\
&= 975.3 \text{ ms}
\end{aligned}$$

- (e) Recall the equation for $P_B(K)$:

$$P_B(K) = \left(\frac{1 - \rho}{1 - \rho^{K+1}} \right) \rho^K$$

We wish to find K such that $P_B(K) < 10^{-2}$. First, we set $P_B(K) = 10^{-2}$ and solve for K .

$$\left(\frac{1 - \rho}{1 - \rho^{K+1}} \right) \rho^K = 10^{-2}$$

$$\begin{aligned}
 (1 - \rho) \rho^K &= 10^{-2} (1 - \rho^{K+1}) \\
 &= 10^{-2} - 10^{-2} \rho^{K+1} \\
 (1 - \rho) \rho^K + 10^{-2} \rho^{K+1} &= 10^{-2} \\
 (1 - \rho + 10^{-2} \rho) \rho^K &= 10^{-2} \\
 \rho^K &= \frac{10^{-2}}{1 - 0.99\rho}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 K &= \frac{-2(\ln 10) - \ln(1 - 0.99\rho)}{\ln \rho} \\
 &= \frac{-[2(\ln 10) + \ln(1 - 0.99\rho)]}{\ln \rho}
 \end{aligned}$$

For $\rho = 0.91366$,

$$\begin{aligned}
 K &= \frac{-[4.60517 - 2.348874]}{-0.09029676} \\
 &= \frac{2.256296}{0.09029676} \\
 &= 25.799
 \end{aligned}$$

Therefore, for a blocking probability less than 10^{-2} , we need $K \geq 26$. From part (c), note that for the non-blocking system, the value of K such that $P\{\tilde{n} > K\} < 10^{-2}$ is 50 for this particular value of λ . Thus, it is seen that the buffer size required to achieve a given blocking probability cannot be obtained directly from the survivor function. In this case, approximating buffer requirements from the survivor function would have resulted in $K = 50$, but in reality, only 26 storage spots are needed.

- 3-2 A finite population, K , of users attached to a statistical multiplexing system operate in a continuous cycle of *think, wait, service*. During the think phase, the length of which is denoted by \tilde{t} , the user generates a message. The message then waits in a queue behind any other messages, if any, that may be awaiting transmission. Upon reaching the head of the queue, the user receives service and the corresponding message is transmitted over a communication channel. Message service times, \tilde{x} , and think times, \tilde{t} , are drawn from exponential distributions with rates μ and λ , respectively. Let the state of the system be defined as the total number of users waiting and in service and be denoted by \tilde{n} .

ANNOTATED REFERENCES

References [1] and [2] provide an introduction to queueing theory at a level slightly higher than that given here. Reference [2] is an invaluable source of classical queueing theory results in telephony problems. Reference [3] demonstrates the application of queueing theory to data communication networks. References [1–7] discuss techniques for simulating queueing systems and for analyzing the resulting data. [8–10] presents excellent discussions on reversible processes and $M/G/c/c$ and $M/G/\infty$.

1. L. Kleinrock, *Queueing Systems*, vol. 1, Wiley, New York, 1975.
2. R. B. Cooper, *Introduction to Queueing Theory*, 2nd ed., North Holland, 1981. Reprinted by CEE Press of the George Washington University.
3. D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
4. A. M. Law and W. D. Kelton, *Simulation, Modeling, and Analysis*, 2nd ed., McGraw-Hill, New York, 1999.
5. J. Banks, J. S. Carson II, and B. L. Nelson, *Discrete-Event System Simulation*, Prentice-Hall, Upper Saddle River, NJ, 1996.
6. G. S. Fishman, *Discrete-Event Simulation: Modeling, Programming, and Analysis*, Springer-Verlag, New York, 2001.
7. S. M. Ross, *Stochastic Processes*, Wiley, New York, 1983.
8. M. Reiser and S. S. Lavenberg, “Mean-value analysis of closed multichain queueing networks,” *J. Assoc. Comput. Mach.* 27: 313–322, 1980.
9. S. S. Lavenberg, *Computer Performance Modeling Handbook*, Academic Press, New York, 1983.
10. K. Pawlikowski, “Steady-state simulation of queueing processes: survey of problems and solutions,” *ACM Computing Surveys*, Vol. 22, No. 2, pp. 123–170, 1990.

PROBLEMS

Sections 12.1 and 12.2: The Elements of a Queueing Network and Little’s Formula

- 12.1.** Describe the following queueing systems: $M/M/1$, $M/D/1/K$, $M/G/3$, $D/M/2$, $G/D/1$, $D/D/2$.
- 12.2.** Suppose that a queueing system is empty at time $t = 0$, let the arrival times of the first six customers be 1, 3, 4, 7, 8, 15, and let their respective service times be 3.5, 4, 2, 1, 1.5, 4. Find S_i , τ_i , D_i , W_i , and T_i for $i = 1, \dots, 5$; sketch $N(t)$ versus t ; and check Little’s formula by computing $\langle N \rangle_t$, $\langle \lambda \rangle_t$, and $\langle T \rangle_t$ for each of the following three service disciplines:
 - (a) First come, first served.
 - (b) Last come, first served.
 - (c) Shortest job first (assume that the precise service time of each job is known before it enters service).
- 12.3.** A data communication line delivers a block of information every $10\ \mu s$. A decoder checks each block for errors and corrects the errors if necessary. It takes $1\ \mu s$ to determine whether a block has any errors. If the block has one error, it takes $5\ \mu s$ to correct it, and if it has more than one error it takes $20\ \mu s$ to correct the error. Blocks wait in a queue when the decoder falls behind. Suppose that the decoder is initially empty and that the numbers of errors in the first ten blocks are 0, 1, 3, 1, 0, 4, 0, 1, 0, 0.

- (a) Plot the number of blocks in the decoder as a function of time.
 (b) Find the mean number of blocks in the decoder.
 (c) What percentage of the time is the decoder empty?
- 12.4.** Three queues are arranged in a loop as shown in Fig. P12.1. Assume that the mean service time in queue i is $m_i = 1/\mu_i$.

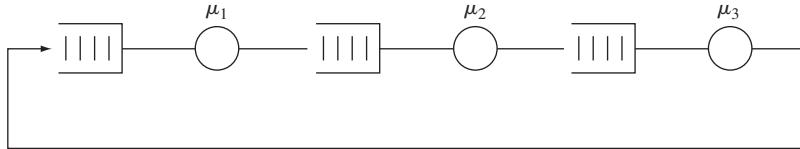


FIGURE P12.1

- (a) Suppose the queue has a single customer circulating in the loop. Find the mean time $E[T]$ it takes the customer to cycle around the loop. Deduce from $E[T]$ the mean arrival rate λ at each of the queues. Verify that Little's formula holds for these two quantities.
 (b) If there are N customers circulating in the loop, how are the mean arrival rate and the mean cycle time related?
- 12.5.** A very popular barbershop is always full. The shop has two barbers and three chairs for waiting, and as soon as a customer completes his service and leaves the shop, another enters the shop. Assume the mean service time is m .
 (a) Use Little's formula to relate the arrival rate and the mean time spent in the shop.
 (b) Use Little's formula to relate the arrival rate and the mean time spent in service.
 (c) Use the above formulas to find an expression for the mean time spent in the system in terms of the mean service time.
- 12.6.** In Problem 12.3, suppose that the probabilities of zero, one, and more than one errors are p_0 , p_1 , and p_2 , respectively. Use Little's formula to find the mean number of blocks in the decoder.
- 12.7.** A communication network receives messages from R sources with mean arrival rates $\lambda_1, \dots, \lambda_R$. On the average there are $E[N_i]$ messages from source i in the network.
 (a) Use Little's formula to find the average time $E[T_i]$ spent by type i customers in the network.
 (b) Let λ denote the total arrival rate into the network. Use Little's formula to find an expression for the mean time $E[T]$ spent by customers (of all types) in the network in terms of the $E[N_i]$.
 (c) Combine the results of part a and part b to obtain an expression for $E[T]$ in terms of $E[T_i]$. Derive the same expression using $A(t)$ the arrival processes for each type.

Section 12.3: The M/M/1 Queue

- 12.8.** (a) Find $P[N \geq n]$ for an M/M/1 system.
 (b) What is the maximum allowable arrival rate in a system with service rate μ , if we require that $P[N \geq 10] = 10^{-3}$?

- 12.9.** A decision to purchase one of two machines is to be made. Machine 1 has a processing rate of μ transactions/hour and it costs B dollars/hour to operate whether idle or not; machine 2 is twice as fast but costs twice as much to operate. Suppose that transactions arrive at the system according to a Poisson process of rate λ and that the transaction processing times are exponentially distributed. The total cost of the system is the operation cost plus a cost of A dollars for each hour a customer has to wait.
- Find expressions for the total cost per hour for each of the systems. Plot this cost versus the arrival rate.
 - If $A = B/10$, for what range of arrival rates is machine 1 cheaper? Repeat for $A = 10B$.
- 12.10.** Consider an M/M/1 queueing system in which each customer arrival brings in a profit of \$5 but in which each unit time of delay costs the system \$1. Find the range of arrival rates for which the system makes a net profit.
- 12.11.** Consider an M/M/1 queueing system with arrival rate λ customers/second.
- Find the service rate required so that the average queue is five customers (i.e., $E[N_q] = 5$).
 - Find the service rate required so that the queue that forms from time to time has mean 5 (i.e., $E[N_q | N_q > 0] = 5$).
 - Which of the two criteria, $E[N_q]$ or $E[N_q | N_q > 0]$, do you consider the more appropriate?
- 12.12.** Show that the p th percentile of the waiting time for an M/M/1 system is given by
- $$x = \frac{1/\mu}{1 - \rho} \ln\left(\frac{\rho}{1 - p}\right).$$
- 12.13.** Consider an M/M/1 queueing system with service rate two customers per second.
- Find the maximum allowable arrival rate if 90% of customers should not have a delay of more than 3 seconds.
 - Find the maximum allowable arrival rate if 90% of customers should not have to wait for service for more than 2 seconds. Hint: Use the result from Problem 12.12, and then find λ by trial and error.
- 12.14.** Verify Eq. (12.36) for the steady state pmf of an M/M/1/ K system.
- 12.15.** Consider an M/M/1/2 queueing system in which each customer accepted into the system brings in a profit of \$5 and each customer rejected results in a loss of \$1. Find the arrival rate at which the system breaks even.
- 12.16.** For an M/M/1/ K system show that

$$P[N = k | N < K] = \frac{P[N = k]}{1 - P[N = K]} \quad 0 \leq k < K.$$

Why does this probability represent the proportion of arriving customers who actually enter the system and find exactly k customers in the system?

- 12.17. (a)** Use the matrix exponential method of Eq. (11.72) to find the transient solution for the state pmfs for an M/M/1/5 queue under the following conditions:
- $\rho = 0.5$ and $N(0) = 0, N(0) = 2, N(0) = 5$;
 - $\rho = 1$ and $N(0) = 0, N(0) = 2, N(0) = 5$.
- (b)** Plot $E[N(t)]$ vs. t for the cases considered in part a.

- 12.18.** Suppose that two types of customers arrive at a queueing system according to independent Poisson process of rate $\lambda/2$. Both types of customers require exponentially distributed service times of rate μ . Type 1 customers are always accepted into the system, but type 2 customers are turned away when the total number of customers in the system exceeds K .
- Sketch the transition rate diagram for $N(t)$, the total number of customers in the system.
 - Find the steady state pmf of $N(t)$.
- 12.19.** Consider the queueing system in Problem 12.18 with $K = 5$ and with a maximum system occupancy of 10 customers. In this problem we use the matrix exponential method of Eq. (11.72) to explore how the system adjusts to sudden increases in load.
- Find the transient state pmf for the system with $\lambda = 1/2$ and $\mu = 1$, assuming that initially there are 5 customers in the system.
 - Suppose that at time 20, the λ increases to 1. Find the transient state pmf after this surge in traffic.

Section 12.4: Multiserver Systems: M/M/c, M/M/c/c, and M/M/ ∞

- 12.20.** Find $P[N \geq c + k]$ for an M/M/c system.
- 12.21.** Customers arrive at a shop according to a Poisson process of rate 12 customers per hour. The shop has two clerks to attend to the customers. Suppose that it takes a clerk an exponentially distributed amount of time with mean 5 minutes to service one customer.
- What is the probability that an arriving customer must wait to be served?
 - Find the mean number of customers in the system and the mean time spent in the system.
 - Find the probability that there are more than 4 customers in the system.
- 12.22.** Little's formula applied to the servers implies that the mean number of busy servers is $\lambda E[\tau]$. Verify this by explicit calculation of the mean number of busy servers in an M/M/c system.
- 12.23.** Inquiries arrive at an information center according to a Poisson process of rate 10 inquiries per second. It takes a server 1/2 second to answer each query.
- How many servers are needed if we require that the mean total delay for each inquiry should not exceed 4 seconds, and 90% of all queries should wait less than 8 seconds?
 - What is the resulting probability that all servers are busy? Idle?
- 12.24.** Consider a queueing system in which the maximum processing rate is $c\mu$ customers per second. Let k be the number of customers in the system. When $k \geq c$, c customers are served at a rate μ each. When $0 < k \leq c$, these k customers are served at a rate $c\mu/k$ each. Assume Poisson arrivals of rate λ and exponentially distributed times.
- Find the transition rate diagram for this system.
 - Find the steady state pmf for the number in the system.
 - Find $E[W]$ and $E[T]$.
 - For $c = 2$, compare $E[W]$ and $E[T]$ for this system to those of M/M/1 and M/M/2 systems of the same maximum processing rate.
- 12.25.** (a) Suppose that the queueing system in Problem 12.24 models a Web server where c is the maximum number of clients allowed to place queries at the same time. Discuss the impact of the choice of the parameter c on queueing and total delay performance.
- (b) Consider the fact that while connected to the Web server, clients spend their time in three states: sending the query, waiting for the response, and thinking after each response. How does this affect the choice of c ? Should the system impose a time-out limit on the customer's connection time?

- 12.26.** Show that the Erlang B formula satisfies the following recursive equation:

$$B(c, a) = \frac{aB(c - 1, a)}{c + aB(c - 1, a)},$$

where $a = \lambda E[\tau]$.

- 12.27.** Consider an M/M/5/5 system in which the arrival rate is 10 customers per minute and the mean service time is 1/2 minute.

- (a) Find the probability of blocking a customer. *Hint:* Use the result from the Problem 12.26.
- (b) How many more servers are required to reduce the blocking probability to 10%?

- 12.28.** A tool rental shop has four floor sanders. Customers for floor sanders arrive according to a Poisson process at a rate of one customer every two days. The average rental time is exponentially distributed with mean two days. If the shop has no floor sanders available, the customers go to the shop across the street.

- (a) Find the proportion of customers that go to the shop across the street.
- (b) What is the mean number of floor sanders rented out?
- (c) What is the increase in lost customers if one of the sanders breaks down and is not replaced?

- 12.29. (a)** Show that the Erlang C formula is related to the Erlang B formula by

$$C(c, a) = \frac{cB(c, a)}{c - a\{1 - B(c, a)\}} \quad \text{for } c > a.$$

- (b) Show that this implies that $C(c, a) > B(c, a)$.

- 12.30.** Suppose that department A in a certain company has three private videoconference lines connecting two sites. Calls arrive according to a Poisson process of rate 1 call/hour, and have an exponentially distributed holding time of 2 hours. Calls that arrive when the three lines are busy are automatically redirected to public video lines. Suppose that department B also has three private videoconference lines connecting the same sites, and that it has the same arrival and service statistics.

- (a) Find the proportion of calls that are redirected to public lines.
- (b) Suppose we consolidate the videoconference traffic from the two departments and allow all calls to share the six lines. What proportion of calls are redirected to public lines?

- 12.31.** A $c = 10$ server blocking system handles two streams of customers that each arrive at rate $\lambda/2$. Type 1 customers have a mean service time of 1 time unit, and Type 2 customers have a service time of 3 time units. Compare the blocking performance of a system that allows customers to access any available server against one that allocates half the servers to each class. Does scale matter? Does the answer change if $c = 100$?

- 12.32.** Suppose we use $P[N = c]$ from an M/M/ ∞ system to approximate $B(c, a)$ in selecting the number of servers in an M/M/c/c system. Is the resulting design optimistic or pessimistic?

- 12.33.** During the evening rush hour, users log onto a peer-to-peer network at a rate of 10 users per second. Each user stays connected to the network an average of 1 hour.

- (a) What is the steady state pmf for the number of customers logged onto the peer-to-peer network?
- (b) Is steady state ever achieved?
- (c) Is it reasonable to assume a Gaussian distribution for the number of customers in the system?

Section 12.5: Finite-Source Queueing Systems

- 12.34.** A computer is shared by 15 users as shown in Fig. 12.14(b). Suppose that the mean service time is 2 seconds and the mean think time is 30 seconds, and that both of these times are exponentially distributed.
- Find the mean delay and mean throughput of the system.
 - What is the system saturation point K^* for this system?
 - Repeat part a if 5 users are added to the system.
- 12.35.** A Web server that has the maximum number of clients connected is modeled by the system in Figure 12.14(b). Suppose that the system can handle a query in 10 milliseconds and the users click new queries at a rate of 1 every 5 seconds.
- Find the value of K^* for this system.
 - Find the pmf for the number of requests found in queue by arriving queries.
- 12.36.** Find the transition rate diagram and steady state pmf for a two-server finite-source queueing system.
- 12.37.** Verify that Eqs. (12.84) and (12.81) give $E[T]$ as given in Eq. (12.72).
- 12.38.** Consider a c -server, finite-source queueing system that allows no queueing for service. Requests that arrive when all servers are busy are turned away, and the corresponding source immediately returns to the “think” state, and spends another exponentially distributed think time before submitting another request for service.
- Find the transition rate diagram and show that the steady state pmf for the state of the system is

$$P_K[N = j] = \frac{\binom{K}{j} p^j (1-p)^{K-j}}{\sum_{i=0}^c \binom{K}{i} p^i (1-p)^{K-i}} \quad i = 0, \dots, c,$$

where c is the number of servers, K is the number of sources, and

$$p = \frac{\alpha/\mu}{1 + \alpha/\mu}.$$

- Find the probability that all servers are busy.
 - Use the fact that arriving customers “see” the steady state pmf of a system with one less source to show that the fraction of arrivals that are turned away is given by $P_{K-1}(c)$. The resulting expression is called the Engset formula.
- 12.39.** A video-on-demand system is modeled as a $c = 10$ server system that handles video chunk requests from K clients. Suppose that the system is modeled by the Engset system from Problem 12.38. Suppose that users generate requests at a rate of one per second and each server can meet the request within 100 ms. Find the number of clients that can be connected if the probability of turning away a request is 10%? 1%?

Section 12.6: M/G/1 Queueing Systems

- 12.40.** Find the mean waiting time and mean delay in an M/G/1 system in which the service time is a k -Erlang random variable (see Table 4.1) with mean $1/\mu$. Compare the results to M/M/1 and M/D/1 systems.

- 12.41.** A $k = 2$ hyperexponential random variable is obtained by selecting a service time at random from one of two exponential random variables as shown in Fig. P12.2. Find the mean delay in an M/G/1 system with this hyperexponential service time distribution.

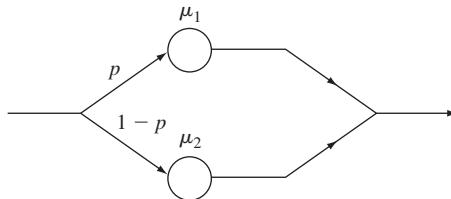


FIGURE P12.2

- 12.42.** Customers arrive at a queueing system according to a Poisson process of rate λ . A fraction α of the customers require a fixed service time d , and a fraction $1 - \alpha$ require an exponential service time of mean $1/\mu$. Find the mean waiting time and mean delay in the resulting M/G/1 system.
- 12.43.** Find the mean waiting time and mean delay in an M/G/1 system in which the service time consists of a fixed time d plus an exponentially distributed time of mean $1/\mu$.
- 12.44.** Fixed-length messages arrive at a transmitter according to a Poisson process of rate λ . The time required to transmit a message and to receive an acknowledgment is d seconds. If a message is acknowledged as having been received correctly, then the transmitter proceeds with the next message. If the message is acknowledged as having been received in error, the transmitter retransmits the message. Assume that a message undergoes errors in transmission with probability p , and that transmission errors are independent.
- Find the mean and variance of the effective message service time.
 - Find the mean message delay.
- 12.45.** Packets at a router with a 1 Gigabit/second transmission line arrive at a rate of λ packets per second. Suppose that half the packets are 40 bytes long and half the packets are 1500 bytes long. Find the mean packet delay as a function of λ .
- 12.46.** A file server receives requests at a rate of λ requests per second. The server can transmit files at a rate of 12.5 Megabytes per second. Suppose that file lengths have a Pareto distribution with mean 1 Megabyte.
- Find the average delay in meeting a file request.
 - Discuss the effect of the Pareto distribution parameter on system performance.
- 12.47.** Jobs arrive at a machine according to a Poisson process of rate λ . The service times for the jobs are exponentially distributed with mean $1/\mu$. The machine has a tendency to break down while it is serving customers; if a particular service time is t , then the probability that it will break down k times during this service time is a Poisson random variable with mean αt . It takes an exponentially distributed time with mean $1/\beta$ to repair the machine. Assume a machine is always working when it begins a job.
- Find the mean and variance of the total time required to complete a job. Hint: Use conditional expectation.
 - Find the mean job delay for this system.

- 12.48.** Consider a two-class nonpreemptive priority queueing system, and suppose that the lower-priority class is saturated (i.e., $\lambda_1 E[\tau_1] + \lambda_2 E[\tau_2] > 1$).

- (a) Show that the rate of low-priority customers served by the system is $\lambda'_2 = (1 - \lambda_1 E[\tau_1]) / E[\tau_2]$. Hint: What proportion of time is the server busy with class two customers?
- (b) Show that the mean waiting time for class 1 customers is

$$E[W_1] = \frac{(1/2)\lambda_1 E[\tau_1^2]}{1 - \lambda_1 E[\tau_1]} + \frac{E[\tau_2^2]}{2E[\tau_2]}.$$

- 12.49.** Consider an M/G/1 system in which the server goes on vacations (becomes unavailable) whenever it empties the queue. If upon returning from vacation the system is still empty, the server takes another vacation, and so on until it finds customers in the system. Suppose that vacation times are independent of each other and of the other variables in the system. Show that the mean waiting time for customers in this system is

$$E[W] = \frac{(1/2)\lambda E[\tau^2]}{1 - \lambda E[\tau]} + \frac{E[V^2]}{2E[V]},$$

where V is the vacation time. Hint: Show that this system is equivalent to a nonpreemptive priority system and use the result of Problem 12.48.

- 12.50.** Fixed-length packets arrive at a concentrator that feeds a synchronous transmission system. The packets arrive according to a Poisson process of rate λ , but the transmission system will only begin packet transmissions at times id , $i = 1, 2, \dots$, where d is the transmission time for a single packet. Find the mean packet waiting time. Hint: Show that this is an M/D/1 queue with vacations as in Problem 12.49.
- 12.51.** A queueing system handles two types of traffic. Type i traffic arrives according to a Poisson process and has exponentially distributed service times with mean $1/\mu_i$ for $i = 1, 2$. Suppose that type 1 customers are given nonpreemptive priority. Plot the overall and per-class mean waiting time versus λ if $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = 1$, $\mu_2 = 1/10$.
- 12.52.** Consider a two-class priority M/G/1 system in which high-priority customer arrivals preempt low-priority customers who are found in service. Preempted low-priority customers are placed at the head of their queue, and they resume service when the server again becomes available to low-priority customers.

- (a) What is the mean waiting time and the mean delay for the high-priority customers?
- (b) Show that the time required to service all customers found by a type 2 arrival to the system is

$$\frac{R_2}{1 - \rho_1 - \rho_2},$$

where $\rho_j = \lambda_j E[\tau_j]$, and

$$R_2 = \frac{1}{2} \sum_{j=1}^2 \lambda_j E[\tau_j^2].$$

- (c) Show that the time required to service all type 1 customers who arrive during the time a type 2 customer spends in the system is $\rho_1 E[T_2]$.

- (d) Use parts b and c to show that

$$E[T_2] = \frac{(1 - \rho_1 - \rho_2)/\mu_2 + R_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}.$$

- 12.53.** Evaluate and plot the formulas developed in Problem 12.52 using the two traffic classes described in Problem 12.51.

Section 12.7: M/G/1 Analysis Using Embedded Markov Chain

- 12.54.** The service time in an M/G/1 system has a $k = 2$ Erlang distribution with mean $1/\mu$ and $\lambda = \mu/2$.

- (a) Find $G_N(z)$ and $P[N = j]$.
 (b) Find $\hat{W}(s)$ and $\hat{T}(s)$ and the corresponding pdf's.

- 12.55. (a)** In Problem 12.47, show that the Laplace transform of the pdf for the total time τ required to complete the service of a customer is

$$\hat{\tau}(s) = \frac{\mu(s + \beta)}{(s + \beta)(s + \mu) + \alpha s}.$$

Hint: Use conditional expectation in evaluating $E[e^{-s\tau}]$, and note that the number of breakdowns depends on the service time of the customer.

- (b) Find $\hat{W}(s)$ and $\hat{T}(s)$ and the corresponding pdf's.

- 12.56. (a)** Show that Eqs. (12.110a) and (12.110b) can be written as

$$N_j = N_{j-1} - U(N_{j-1}) + M_j, \quad (12.186)$$

where

$$U(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0. \end{cases}$$

- (b) Take the expected value of both sides of Eq. (12.186) to obtain an expression for $P[N > 0]$.

- (c) Square both sides of Eq. (12.186) and take the expected value to obtain the Pollaczek–Khinchin formula for $E[N]$.

- 12.57. (a)** Show that for an M/D/1 system,

$$G_N(z) = \frac{(1 - \rho)(1 - z)}{1 - ze^{\rho(1-z)}}.$$

- (b) Expand the denominator in a geometric series, and then identify the coefficient of z^k to obtain

$$P[N = k] = (1 - \rho) \sum_{j=0}^k \frac{(-j\rho)^{k-j-1} (-j\rho - k + 1) e^{j\rho}}{(k - j)!}.$$

- 12.58. (a)** Show that Eq. (12.130) can be rewritten as

$$\hat{W}(s) = \frac{1 - \rho}{1 - \rho \hat{R}(s)}, \quad (12.87)$$

where

$$\hat{R}(s) = \frac{1 - \hat{\tau}(s)}{sE[\tau]}$$

is the Laplace transform of the pdf of the residual service time.

- (b) Expand the denominator of Eq. (12.187) in a geometric series and invert the resulting transform expression to show that

$$f_W(x) = \sum_{k=0}^{\infty} (1 - \rho)\rho^k f^{(k)}(x), \quad (12.188)$$

where $f^{(k)}(x)$ is the k th-order convolution of the residual service time.

- 12.59.** Approximate $f_W(x)$ for an M/D/1 system using the $k = 0, 1, 2$ terms of Eq. (12.188). Sketch the resulting pdf for $\rho = 1/2$.

Section 12.8: Burke's Theorem: Departures from M/M/c Systems

- 12.60.** Consider the interdeparture times from a stable M/M/1 system in steady state.

- (a) Show that if a departure leaves the system nonempty, then the time to the next departure is an exponential random variable with mean $1/\mu$.
- (b) Show that if a departure leaves the system empty, then the time to the next departure is the sum of two independent exponential random variables of means $1/\lambda$ and $1/\mu$, respectively.
- (c) Combine the results of parts a and b to show that the interdeparture times are exponential random variables with mean $1/\lambda$.

- 12.61.** Find the joint pmf for the number of customers in the queues in the network shown in Fig. P12.3.

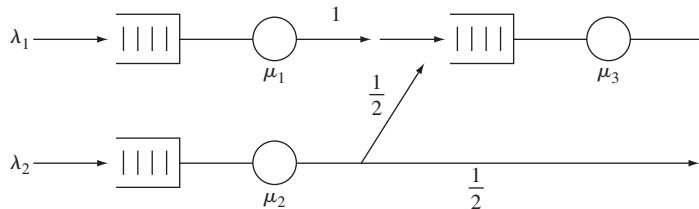


FIGURE P12.3

- 12.62.** Write the balance equations for the feedforward network shown in Fig. P12.4 and verify that the joint state pmf is of product form.

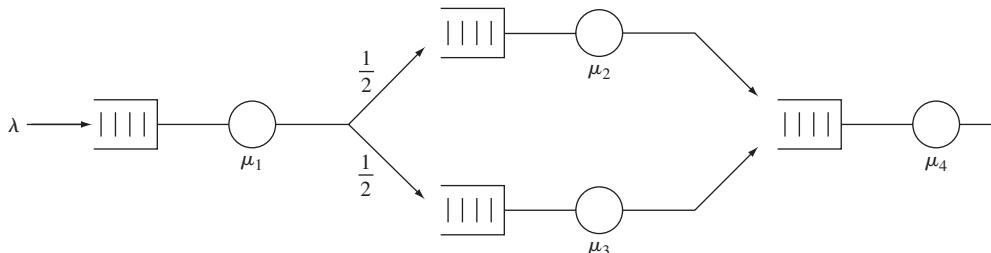


FIGURE P12.4

12.63. Verify that Eqs. (12.137) through (12.139) satisfy Eq. (12.135).

Section 12.9: Networks of Queues: Jackson's Theorem

12.64. Find the joint state pmf for the open network of queues shown in Fig. P12.5.

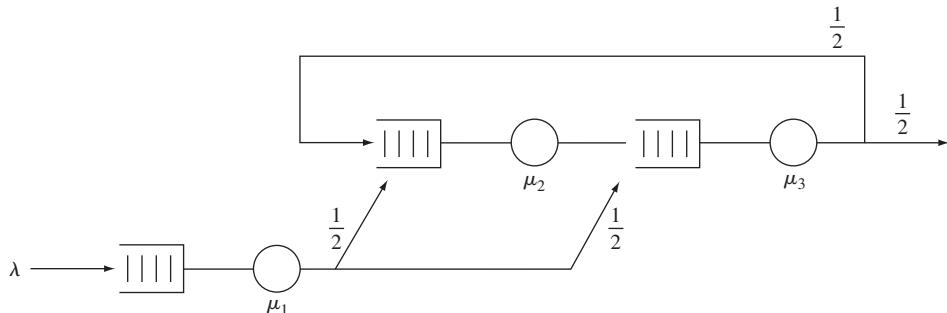


FIGURE P12.5

12.65. A computer system model has three programs circulating in the network of queues shown in Fig. P12.6.

- (a) Find the joint state pmf of the system.
- (b) Find the average program completion rate.

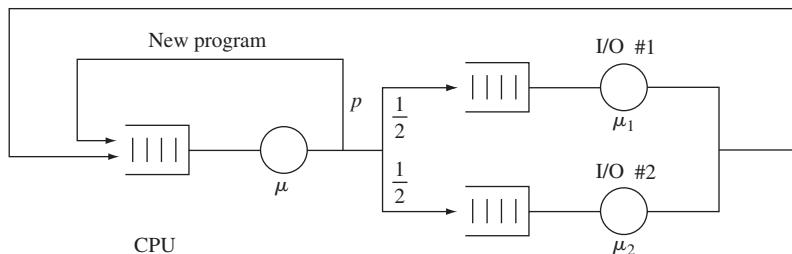


FIGURE P12.6

12.66. Use the mean value analysis algorithm to answer Problem 12.65, part b.

Section 12.10: Simulation and Data Analysis of Queueing Systems

- 12.67. (a)** Repeat the experiment in Example 12.28 for an M/M/1 system with $\rho = 0.5, 0.7$, and 0.9 . Use sample means for $N(t)$ based on 25 replications to characterize the transient behavior. Try out smoothing the sample means using a moving average filter over time. Give an estimate of the time to reach steady state in each of these systems.
- (b)** Now investigate the effect of initial condition on the duration of the transient phase. For each of the utilizations above compare the transient duration when the initial condition is: $N(0) = 0; N(0) = 5; N(0) = 10$.

- 12.68.** For the experiment in Problem 12.67, calculate the sample covariance for each realization and then average over the 25 replications. Find the number of lags required for each value of r until the correlation drops to zero. Comment on the implications for the size of the batches if a method of batch means approach is to be used.
- 12.69.** The correlation of $N(t)$ for an M/M/1 system has the following geometric upper bound [Fishman]:

$$\rho_j \leq \left[\frac{4\rho}{(1 + \rho)^2} \right]^j \quad \text{for } j = 0, 1, 2, \dots$$

Evaluate the ratio of the variance of the sample mean estimator for this process to that of an iid process when $\rho = 0.5, 0.75, 0.9, 0.99$.

- 12.70.** Run the simulation for the experiment in Example 12.29 50 times. For each simulation produce a confidence interval using the method of batch means. Determine the fraction of the confidence intervals that covered the actual mean $E[N]$. Comment on the accuracy of the confidence intervals given by Eq. (12.168).
- 12.71.** Develop a simulation model for an M/M/3 system with $\lambda = 2$ customers per second and $\mu = 1$ customer per second. Use the method of batch means as in Example 12.29 to estimate the probability that an arriving customer has to queue for service. Provide appropriate confidence intervals.
- 12.72.** (a) Consider the simulation in Example 12.30 where the embedded Markov chain approach is used to estimate the steady state pmf. For $\rho = 0.5$ and $\rho = 0.9$, use different warm-up periods to investigate the effect of the initial transient on the pmf estimates.
(b) Double the number of replications and observe the impact on the confidence intervals.
- 12.73.** Develop a simulation for an M/D/1 system with $\rho = 0.7$ using the embedded Markov chain in Eq. (12.172). Design the simulation to estimate the pmf for the number of customers in the system as well as the mean number in the system.
(a) Discuss what transient effects can be expected in this approach.
(b) Use the method of batch means to develop estimates for the mean number of customers in the system. Discuss the choice of batch size and warm-up period. Evaluate the confidence intervals produced by several realizations.
- 12.74.** Use Lindley's recursion to estimate the waiting-time distribution for customers in an M/D/1 system with $\rho = 0.5$ and $\rho = 0.7$. Is there anything peculiar about the distribution?
- 12.75.** Use Lindley's recursion to estimate the waiting-time distribution for customers in a D/M/1 system with $\rho = 0.5$ and $\rho = 0.7$.
- 12.76.** Use Lindley's recursion to estimate the waiting-time distribution for customers in an M/G/1 system with $\rho = 0.5$ and $\rho = 0.7$ where the service-time distribution is Pareto with parameter $\alpha = 2.5$. Try a simulation with $\alpha = 1.5$. Does anything peculiar happen?
- 12.77.** Repeat the experiment in Example 12.33, but use the method of batch means to provide confidence intervals for the mean waiting time.
- 12.78.** Explain why the estimator in Eq. (12.183) will converge to the expected value of the waiting time.
- 12.79.** Use the regenerative method to estimate the mean number in the system and the probability that the system is empty in an M/D/1 system. Evaluate the confidence interval provided by Eq. (12.185).

Problems Requiring Cumulative Knowledge

- 12.80.** Consider an M/M/2/2 system in which one server is twice as fast as the other server.
- What definition of “state” of the system results in a continuous-time Markov chain?
 - Find the steady state pmf for the system if customers arriving at an empty system are always routed to the faster server.
 - Find the steady state pmf for the system if customers arriving at an empty system are equally likely to be routed to either server.
- 12.81. (a)** Find the transient pmf, $P[N(t) = j]$, for an M/M/1/2 system which is in the empty state at time 0.
- (b)** Repeat part a if the system is full at time 0.
- 12.82. (a)** In an M/G/1 system, why are the set of times when customers arrive to an empty system renewal instants?
- (b)** How would you apply the results from renewal theory in Section 7.5 to estimate the pmf for the number of customers in the system?
- (c)** How would you obtain a confidence interval for $P[N(t) = j]$?
- 12.83.** Let $N(t)$ be a Poisson random process with parameter λ . Suppose that each time an event occurs, a coin is flipped and the outcome is recorded. Assume that the probability of heads depends on the time of the arrival and is denoted by $p(t)$. Let $N_1(t)$ and $N_2(t)$ denote the number of heads and tails recorded up to time t , respectively.
- Show that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with rates $p\lambda$ and $(1 - p)\lambda$, where

$$p = \frac{1}{t} \int_0^t p(t') dt'.$$

- Are $N_1(t)$ and $N_2(t)$ independent Poisson random processes? If so, how would you show this?
- 12.84.** Consider an M/G/ ∞ system in which customers arrive at rate λ and in which the customer service times have distribution $F_X(x)$. Suppose that the system is empty at time 0. Let $N_1(t)$ be the number of customers who have completed their service by time t , and let $N_2(t)$ be the number of customers still in the system at time t .
- Use the result of Problem 12.83 to find the joint pmf of $N_1(t)$ and $N_2(t)$.
 - What is the steady state pmf for the number of customers in an M/G/ ∞ system?
 - Apply Little’s formula to compute the average number of customers in the system. Is the result consistent with your result in part b?

Chapter 12: Introduction to Queueing Theory

12.1 & 12.2 The Elements of a Queueing Network and Little's Formula

12.1

9.1 a) M/M/1 Poisson arrivals, exponential service time, single server, no limit on number of customers

M/D/1/K Poisson arrivals, constant service time, single server, at most K customers allowed in system

M/G/3 Poisson arrivals, iid general service time, 3 servers, no limit on number of customers

D/M/2 Constant interarrival times, exponential service times, two servers, no limit on number of customers

G/D/1 Arrivals according to a general process, fixed constant service times, single server, no limit on number of customers

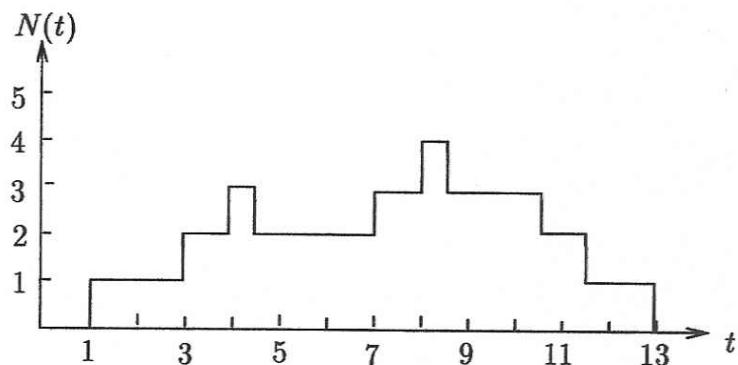
D/D/2 Constant interarrival times, constant service times, two servers, no limit on number of customers in the system

12.2 $\{S_i\} = \{1, 3, 4, 7, 8, 15\}$
 $\{\tau_i\} = \{3.5, 4, 2, 1, 1.5, 4\}$

a) FCFS

i	S_i	τ_i	D_i	W_i	T_i
1	1	3.5	4.5	0	3.5
2	3	4	8.5	1.5	5.5
3	4	2	10.5	4.5	6.5
4	7	1	11.5	3.5	4.5
5	8	1.5	13.0	3.5	5.0
6	15	4			

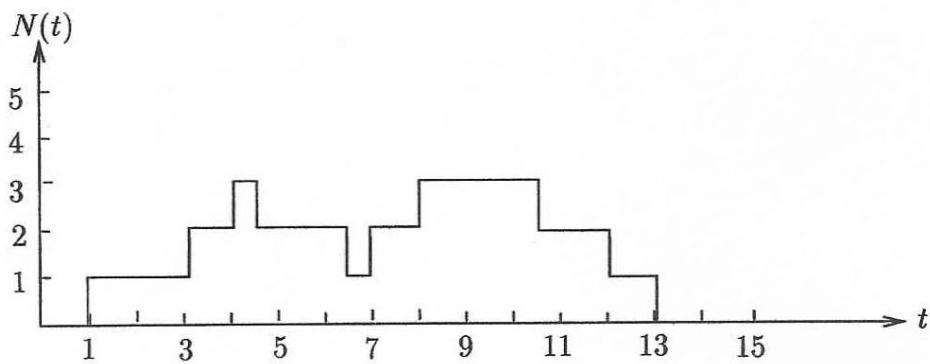
where $W_i = D_{i-1} - S_i = T_i - \tau_i$ and $T_i = D_i - S_i = W_i + \tau_i$



$$\begin{aligned} \langle N \rangle_{13} &= \frac{1}{13} \sum_{i=1}^{13} T_i = \frac{25}{13} \\ \langle \lambda \rangle_{13} &= \frac{A_{13}}{13} = \frac{5}{13} \\ \langle T \rangle_{13} &= \frac{1}{A_{13}} \sum_{i=1}^{13} T_i = \frac{25}{5} \\ \langle N \rangle_{13} &= \frac{25}{13} = \langle \lambda \rangle_{13} \langle T \rangle_{13} = \frac{5}{13} \frac{25}{5} \quad \checkmark \end{aligned}$$

b) LCFS

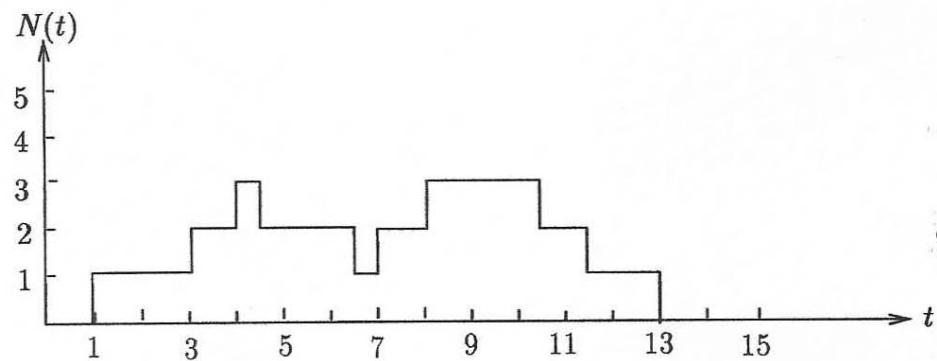
i	S_i	τ_i	D_i	$W_i = T_i - \tau_i$	$T_i = D_i - S_i$
1	1	3.5	4.5	0	3.5
2	3	4	10.5	3.5	7.5
3	4	2	6.5	0.5	2.5
4	7	1	13.0	5.0	6.0
5	8	1.5	12.0	2.5	4.0



$$\begin{aligned} \langle N \rangle_{13} &= \frac{23.5}{13} & \langle \lambda \rangle_{13} &= \frac{5}{13} & \langle T \rangle_{13} &= \frac{23.5}{5} \\ \langle N \rangle_{13} &= \langle \lambda \rangle_{13} \langle T \rangle_{13} \end{aligned}$$

c) Shortest Job First:

i	S_i	τ_i	D_i	$W_i = T_i - \tau_i$	$T_i = D_i - S_i$
1	1	3.5	4.5	0	3.5
2	3	4	10.5	3.5	7.5
3	4	2	6.5	0.5	2.5
4	7	1	11.5	3.5	4.5
5	8	1.5	13.0	3.5	5.0



$$\begin{aligned} \langle N \rangle_{13} &= \frac{23}{13} & \langle \lambda \rangle_{13} &= \frac{5}{13} & \langle T \rangle_{13} &= \frac{23}{5} \\ \langle N \rangle_{13} &= \langle \lambda \rangle_{13} \langle T \rangle_{13} \end{aligned}$$

12.3

1) Interarrivals are constant with interarrival times = 10 μ sec

2) Service time

if 0 error 1 μ sec1 error 1+5 μ sec> 1 error 1+20 μ sec

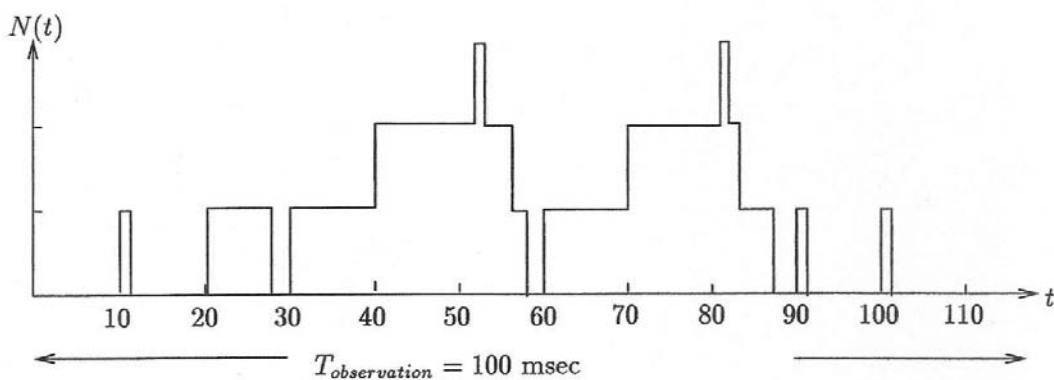
arrival time 10 20 30 40 50 60 70 80 90 100

errors 0 1 3 1 0 4 0 1 0 0

service time 1 6 21 6 1 21 1 6 1 1

dep. time 11 26 51 57 58 81 82 88 91 101

a)



$$\text{b) } \frac{1}{100} \int_{10}^{110} N(t) dt$$

$$= \frac{1}{100} [1 + 6 + 10 + 20 + 3 + 12 + 1 + 10 + 20 + 3 + 2 + 6 + 1 + 1]$$

$$= 0.96$$

c) Server is working during 65 msec = Σ service times

$$\text{proportion idle time} = 1 - \frac{65}{100} = 0.35$$

12.4

9.4 a) One customer \Rightarrow no waiting

$$\Rightarrow \mathcal{E}[T] = m_1 + m_2 + m_3 \Rightarrow \lambda = \frac{1}{m_1 + m_2 + m_3}$$

Little's formula \Rightarrow

$$\begin{aligned}\mathcal{E}[N_i] &= \lambda m_i = \frac{m_i}{m_1 + m_2 + m_3} = \% \text{ time customer in queue } i \\ \sum_{i=1}^3 \mathcal{E}[N_i] &= \sum_{i=1}^3 \frac{m_i}{m_1 + m_2 + m_3} = 1 \Rightarrow \text{one customer in system}\end{aligned}$$

b) Let $\mathcal{E}[T]$ = mean cycle time per customer, then

$$\frac{\text{total \#}}{\text{in system}} = N = \lambda \mathcal{E}[T] \quad \text{by Little's formula}$$

12.5

9.5 a) $\lambda T = 5$ b) $\lambda m = 2$

$$\text{c) } T = \frac{5}{\lambda} = 5 \left(\frac{m}{2} \right) = \frac{5}{2} m$$

12.6

9.6 Let τ be the service time, then

$$\begin{aligned}P[\tau = 1] &= p_0 \quad [\tau = 1 + 5] = p_1 \quad P[\tau = 1 + 20] = p_2 \\ \mathcal{E}[\tau] &= 1 \cdot p_0 + 6p_1 + 21p_2 \quad 10^{-6} \text{ sec} \\ \lambda &= 1 \text{ arrival every } 10 \mu\text{s} = \frac{1}{10^{-5}} = 10^5 \\ \mathcal{E}[N_d] &= \lambda \mathcal{E}[\tau] = \frac{p_0 + 6p_1 + 21p_2}{10}\end{aligned}$$

12.7 a) $\mathcal{E}[T_i] = \frac{1}{\lambda_i} \mathcal{E}[N_i]$

b) $\mathcal{E}[T] = \frac{1}{\lambda} \mathcal{E}[N] = \frac{1}{\lambda} \sum_i \mathcal{E}[N_i]$

c) $\mathcal{E}[T] = \frac{1}{\lambda} \sum_i \lambda_i \mathcal{E}[T_i] = \sum_i \frac{\lambda_i}{\lambda} \mathcal{E}[T_i]$

Let

$$A_i(t) = \# \text{ type } i \text{ arrivals during } [0, t]$$

$$A(t) = \sum_i A_i(t) = \text{total } \# \text{ arrivals}$$

$$\langle T \rangle = \frac{1}{A(t)} \sum_i^{A(t)} T_i \quad \text{average time in system}$$

$$= \frac{1}{A(t)} \left[\sum_{i_1}^{A_1(t)} T_{i_1} + \sum_{i_2}^{A_2(t)} T_{i_2} + \dots + \sum_{i_n}^{A_n(t)} T_{i_n} \right]$$

$$= \frac{1}{A(t)} \left[\frac{A_1(t)}{A_1(t)} \sum_{i_1}^{A_1(t)} T_{i_1} + \dots + \frac{A_n(t)}{A_n(t)} \sum_{i_n}^{A_n(t)} T_{i_n} \right]$$

$$= \underbrace{\frac{A_1(t)}{A(t)}}_{\frac{\lambda_1}{\lambda}} \underbrace{\left(\frac{1}{A_1(t)} \sum_{i_1}^{A_1(t)} T_{i_1} \right)}_{\mathcal{E}[T_1]} + \dots + \underbrace{\frac{A_n(t)}{A(t)}}_{\frac{\lambda_n}{\lambda}} \underbrace{\left(\frac{1}{A_n(t)} \sum_{i_n}^{A_n(t)} T_{i_n} \right)}_{\mathcal{E}[T_n]}$$

as $t \rightarrow \infty$

\Rightarrow same result as above.

12.3 The M/M/1 Queue

12.8

a) $P[N \geq n] = (1 - \rho) \sum_{j=n}^{\infty} \rho^j = (1 - \rho) \frac{\rho^n}{1 - \rho} = \rho^n$

b) $P[N \geq 10] = \rho^{10} = 10^{-3} \Rightarrow \rho = 10^{-0.3} \approx \frac{1}{2}$

$$\Rightarrow \lambda \approx \frac{1}{2} \mu$$

12.9

3.9 a) Machine 1 μ transactions/hr B \$/hr operating costMachine 2 2μ transactions/hr $2B$ \$/hr operating cost

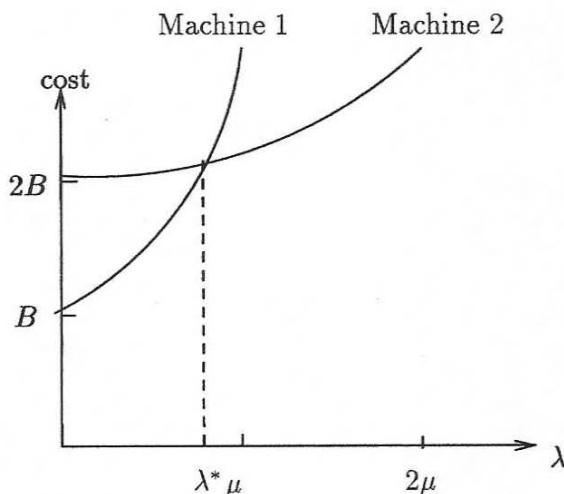
We assume "operating" cost is incurred regardless of whether machine is idle.

$$\text{Cost for } \#1 = B + \lambda \frac{\text{cost}}{\text{hr}} \cdot \bar{W} \frac{\text{waiting hrs.}}{\text{cust}} \cdot A \frac{\$}{\text{hr}}$$

$$= B + \lambda \left(\frac{\rho}{1 - \rho} \frac{1}{\mu} \right) A \quad \text{where } \rho = \frac{\lambda}{\mu}$$

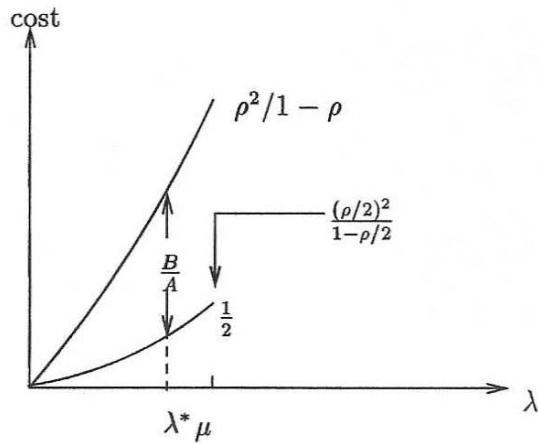
$$= B + A \frac{\rho^2}{1 - \rho}$$

$$\text{Cost for } \#2 = 2B + \lambda \left(\frac{\frac{\rho}{2}}{1 - \frac{\rho}{2}} \right) \left(\frac{1}{2\mu} \right) A = 2B + A \frac{\left(\frac{\rho}{2}\right)^2}{1 - \frac{\rho}{2}}$$

for $\lambda < \lambda^*$ machine 1 is less costly. Let $\rho^* = \lambda^*/\mu$

$$B + A \frac{\rho^{*2}}{1 - \rho^{*2}} = 2B + A \frac{\left(\frac{\rho^*}{2}\right)^2}{1 - \frac{\rho^*}{2}}$$

$$\Leftrightarrow \frac{\rho^{*2}}{1 - \rho^*} - \frac{\left(\frac{\rho^*}{2}\right)^2}{1 - \frac{\rho^*}{2}} = \frac{B}{A}$$



This requires finally the root of a cubic polynomial but we can estimate the root from the figure.

For $\frac{B}{A} \gg 1$

$$\frac{\left(\frac{\rho^*}{2}\right)^2}{1 - \frac{\rho^*}{2}} \approx \frac{1}{2}$$

so we solve the quadratic equation associated with

$$\frac{\rho^{*2}}{1 - \rho^*} - \frac{1}{2} = \frac{B}{A}$$

In particular if $\frac{B}{A} = 10$ then $\rho^* \approx 0.91$.

For $\frac{B}{A} \ll 1$

$$\frac{\left(\frac{\rho^*}{2}\right)^2}{1 - \frac{\rho^*}{2}} \approx 0$$

so we solve

$$\frac{\rho^{*2}}{1 - \rho^*} = \frac{B}{A}$$

In particular if $\frac{B}{A} = \frac{1}{10}$ then $\rho^* \approx 0.27$.

NOTE: If “operating” cost is incurred only when a machine is in use then:

$$\text{Cost of Machine 1} = \underbrace{B\rho}_{\substack{\text{prop. of time} \\ \text{machine 1 in use}}} + \frac{\rho^2}{1 - \rho}$$

$$\text{Cost of Machine 2} = 2B\left(\frac{\rho}{2}\right) + \frac{\left(\frac{\rho}{2}\right)^2}{1 - \frac{\rho}{2}} = B\rho + \frac{\left(\frac{\rho}{2}\right)^2}{1 - \frac{\rho}{2}}$$

In this case machine 2 is less costly than machine 1 for all ρ .

12.10

9.10 A net profit is made if

$$5 > \mathcal{E}[T] = \frac{\frac{1}{\mu}}{1 - \rho}$$

$$\begin{aligned} \Rightarrow 1 - \frac{\lambda}{\mu} &> \frac{1}{5\mu} \Rightarrow 5\mu - 5\lambda > 1 \\ \Rightarrow 0 < \lambda &< \mu - \frac{1}{5} \end{aligned}$$

12.11

a) $\mathcal{E}[N_q] = \frac{\rho^2}{1 - \rho} = 5 \Rightarrow \rho^2 - 5\rho - 5 = 0$

$$\Rightarrow \rho = \frac{-5 + \sqrt{25 - 4(-5)}}{2} = \frac{\sqrt{45} - 5}{2} = 0.854$$

$$\begin{aligned} \text{b)} \quad P[N_q = j | N_q > 0] &= \frac{P[N_q = j]}{P[N_q > 0]} \quad j > 0 \\ &= \frac{(1 - \rho)\rho^j}{\rho} = (1 - \rho)\rho^{j-1} \\ \mathcal{E}[N_q | N_q > 0] &= \sum_{j=1}^{\infty} j(1 - \rho)\rho^{j-1} = \frac{1}{1 - \rho} = 5 \\ \Rightarrow \rho &= \frac{4}{5} = 0.8 \end{aligned}$$

c) It depends on whether one is concerned with average queue over all time (Part a) or on the average queue when one forms (Part b).

12.12

$$\begin{aligned} p &= P[W \leq x] = \int_0^x ((1 - \rho)\delta(t') + \lambda(1 - \rho)e^{-\mu(1-\rho)t'})dt' \\ &= (1 - \rho) + \left| -\frac{\lambda}{\mu} e^{-\mu(1-\rho)t'} \right|_0^x \\ &= 1 - \rho e^{-\mu(1-\rho)x} \\ \Rightarrow \frac{1 - p}{\rho} &= e^{-\mu(1-\rho)x} \\ \Rightarrow x &= \frac{1}{\mu(1 - \rho)} \ln \frac{\rho}{1 - p} \quad \checkmark \end{aligned}$$

$$12.13 \quad \frac{1}{\mu} = \frac{1}{2}$$

a) From Example 12.5

$$\begin{aligned} x &= \frac{1}{\mu - \lambda} \ln \frac{1}{1-p} \\ \Rightarrow \mu - \lambda &= \frac{1}{x} \ln \frac{1}{1-p} \\ \lambda &= \mu - \frac{1}{x} \ln \frac{1}{1-p} = 2 - \frac{1}{3} \ln \frac{1}{1-.9} = 1.232 \end{aligned}$$

b) From Problem 12.12

$$\begin{aligned} x &= \frac{\frac{1}{\mu}}{1-\rho} \ln \frac{\rho}{1-\rho} = \frac{1}{\mu - \lambda} \ln \frac{\lambda}{\mu(1-p)} \\ 2 &= \frac{1}{2-\lambda} \ln 5\lambda \Rightarrow \lambda = 2 - \frac{1}{2} \ln 5\lambda \\ &\Rightarrow \lambda = 1.13 \end{aligned}$$

12.14

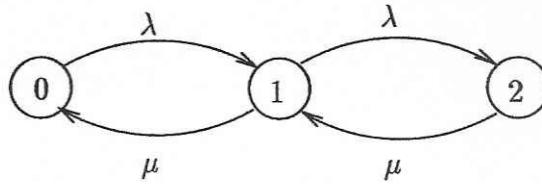
From Example 11.40 we have

$$\begin{aligned} P_j &= \frac{\lambda}{\mu} P_{j-1} = \left(\frac{\lambda}{\mu}\right)^j P_0 \quad 1 \leq j \leq K \\ 1 &= \sum_{j=0}^K \left(\frac{\lambda}{\mu}\right)^j P_0 \Rightarrow P_0 = \frac{1}{\sum_{j=0}^K \left(\frac{\lambda}{\mu}\right)^j} = \frac{1-\rho}{1-\rho^{K+1}} \end{aligned}$$

$$\text{where } \rho = \frac{\lambda}{\mu}$$

$$\therefore P_j = \frac{(1-\rho)}{1-\rho^{K+1}} \rho^j \quad 0 \leq j \leq K$$

12.15



$$\begin{aligned} P_0 &= \frac{1 - \rho}{1 - \rho^3} = \frac{1}{1 + \rho + \rho^2} \\ P_1 &= \frac{\rho}{1 + \rho + \rho^2} \quad P_2 = \frac{\rho^2}{1 + \rho + \rho^2} \end{aligned}$$

In a very long time interval of length T

$$\text{Profit} = \# \text{ accepted into system} \times \$5 - \# \text{ blocked} \times 1$$

$$= \lambda T \times (1 - P_B) \times 5 - \lambda T \times P_B \times 1$$

$$\text{Profit} = 0 \quad \text{if} \quad 5\lambda T(1 - P_B) = \lambda T P_B$$

$$\begin{aligned} &\Leftrightarrow 5 = 6P_B \\ &\Leftrightarrow P_B = \frac{\rho^2}{1 + \rho + \rho^2} = \frac{5}{6} \\ &\Leftrightarrow \rho^2 - 5\rho - 5 = 6 \\ &\Rightarrow \rho = \frac{5 + \sqrt{25 + 20}}{2} = 5.854 \\ &\Rightarrow \lambda = 5.854\mu \end{aligned}$$

$$12.16 \quad P[N = k | N < K] = \frac{P[N = k, N < K]}{P[N < K]} = \begin{cases} 0 & k \geq K \\ \frac{P[N=k]}{P[N < K]} & 0 \leq k < K \end{cases}$$

\therefore for $0 \leq k < K$

$$P[N = k | N < K] = \frac{P[N = k]}{P[N < K]} = \frac{P[N = k]}{1 - P[N = K]}$$

Arriving customers are allowed into the system only when $N < K$.

$\therefore P[N = k | N < K]$ represents the proportion of time when there are k in the system and arriving customers are allowed in. Since Poisson arrivals pick their arrival times at random, then $P[N = k | N < K]$ is the proportion of customers that see k in system upon being admitted in.

(12.17) a) M/M/1/5 \Rightarrow 6 states

$$(i) \quad \Gamma = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 \\ \mu - (\mu + \lambda) & \lambda & 0 & 0 & 0 & 0 \\ 0 & \mu - (\mu + \lambda) & \lambda & 0 & 0 & 0 \\ 0 & 0 & \mu - (\mu + \lambda) & \lambda & 0 & 0 \\ 0 & 0 & 0 & \mu - (\mu + \lambda) & \lambda & 0 \\ 0 & 0 & 0 & 0 & \mu - \mu & 0 \end{pmatrix} = \begin{pmatrix} -0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 1 & -1.5 & 0.5 & 0 & 0 & 0 \\ 0 & 1 & -1.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & -1.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & -1.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$$P(t)_2 e^{\Gamma t} = \bar{E}[e^{\Lambda t}] \bar{E}^{-1}$$

$$\bar{E} = \begin{pmatrix} -0.2644 & -0.4487 & 0.6172 & 0.8661 & -0.7487 & -0.8319 \\ 0.5882 & 0.7661 & -0.6172 & 0.4331 & 0.2193 & -0.1876 \\ -0.5882 & -0.3173 & -0.3086 & 0.2165 & 0.5194 & 0.1870 \\ 0.4263 & -0.1587 & 0.3086 & 0.1083 & 0.2647 & 0.3225 \\ -0.2280 & 0.2708 & 0.1543 & 0.0541 & -0.0775 & 0.3014 \\ 0.0661 & -0.1122 & -0.1543 & 0.0271 & -0.1872 & 0.2080 \end{pmatrix}$$

$$[e^{\Lambda t}] = \begin{pmatrix} e^{-2.72t} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-2.21t} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{+50t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-0.79t} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-0.27t} \end{pmatrix}$$

$$\bar{E}^{-1} = \begin{pmatrix} -0.0578 & 0.2573 & -0.5147 & 0.7460 & -0.7980 & 0.4627 \\ -0.1262 & 0.4309 & -0.3570 & -0.3570 & 1.2188 & -1.0097 \\ 0.1800 & -0.3600 & -0.3600 & -0.7201 & 0.7201 & -1.4402 \\ 0.5864 & 0.5864 & 0.5864 & 0.5864 & 0.5864 & 0.5864 \\ -0.2106 & 0.1233 & 0.5956 & 0.5956 & -0.3489 & -1.6846 \\ -0.1820 & -0.0818 & 0.5643 & 0.5643 & 1.0551 & 1.4558 \end{pmatrix}$$

$$E[e^{At}] \bar{E}^{-1} = X \Rightarrow P(t) = P(0)X$$

$$N(0)=0 \Rightarrow P(0) = (1, 0, 0, 0, 0, 0) \Rightarrow P(t) = \text{first row of } X$$

$$\Rightarrow P(t) = \begin{pmatrix} -2.72t & & & & & t \\ 0.0153e^{-2.72t} + 0.0566e^{-2.21t} + 0.1111e^{-1.5t} + 0.5091e^{-0.591t} + 0.1577e^{-0.791t} + 0.1514e^{-0.271t} & & & & & \\ -0.3040e^{-2.72t} + (-0.0967)e^{-2.21t} + 0.1111e^{-1.5t} + 0.2340e^{-0.591t} + 0.0462e^{-0.791t} + 0.0390e^{-0.271t} & & & & & \\ 0.3040e^{-2.72t} + 0.04e^{-2.21t} - 0.0555e^{-1.5t} + 0.1270e^{-0.591t} - 0.1115e^{-0.791t} - 0.030e^{-0.271t} & & & & & \\ -0.0246e^{-2.72t} + 0.02e^{-2.21t} + 0.0555e^{-1.5t} + 0.0635e^{-0.591t} - 0.0557e^{-0.791t} - 0.0587e^{-0.271t} & & & & & \\ 0.0132e^{-2.72t} - 0.0342e^{-2.21t} + 0.0278e^{-1.5t} + 0.0317e^{-0.591t} + 0.0163e^{-0.791t} - 0.0549e^{-0.271t} & & & & & \\ -0.0582e^{-2.72t} + 0.0142e^{-2.21t} - 0.0278e^{-1.5t} + 0.0159e^{-0.591t} + 0.0394e^{-0.791t} - 0.0379e^{-0.271t} & & & & & \end{pmatrix}$$

For other initial conditions also we can obtain

$P(t)$ in the same way:

$$N(0)=2 \Rightarrow P(0) = (0, 0, 1, 0, 0, 0) \Rightarrow P(t) = P(0) E[e^{At}] \bar{E}^{-1} \Rightarrow \text{third row of } X$$

$$N(0)=5 \Rightarrow P(0) = (0, 0, 0, 0, 0, 1) \Rightarrow P(t) = P(0) E[e^{At}] \bar{E}^{-1} \Rightarrow \text{last row of } X$$

(ii) if $p_{21} \Rightarrow$

$$\Gamma_z = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ +1 & -2 & 1 & 0 & 0 & 0 \\ 0 & +1 & -2 & 1 & 0 & 0 \\ 0 & 0 & +1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

```
% Problem 12.17
% (i)
lambda=0.5;
mu=1;

v=-(lambda+mu);
a=lambda;
b=mu;
L=[-a a 0 0 0 0
    b v a 0 0 0
    0 b v a 0 0
    0 0 b v a 0
    0 0 0 b v a
    0 0 0 0 b -b];

[E D]=eig(L);
t=sym('t');
N0=1;
p0=zeros(1,6);
p0(N0)=1;
p=p0*expm(t*D)*inv(E);
f=inline('E*expm(t*D)*inv(E)');
ff=p0*f(E,t,D); % Example: p0*f(E,2,D) will compute the amount of p(2)
%Plot symbolic function
ezplot(mean(ff));

N0=3;
p0=zeros(1,6);
p0(N0)=1;
p=p0*expm(t*D)*inv(E);
f=inline('E*expm(t*D)*inv(E)');
ff=p0*f(E,t,D);
%Plot symbolic function
ezplot(mean(ff));
```

```

N0=6;
p0=zeros(1,6);
p0(N0)=1;
p=p0*E*exp(t*D)*inv(E);
f=inline('E*expm(t*D)*inv(E)');
ff=p0*f(E,t,D);
%Plot symbolic function
ezplot(mean(ff));

% for part (ii) we repeat the same process with lambda=mu=1

% Problem 12.17
% (i)
lambda=1.0;
mu=1;

v=- (lambda+mu);
a=lambda;
b=mu;
L=[-a a 0 0 0 0
    b v a 0 0 0
    0 b v a 0 0
    0 0 b v a 0
    0 0 0 b v a
    0 0 0 0 b -b];

[E D]=eig(L);
t=sym('t');
N0=6;
p0=zeros(1,6);
p0(N0)=1;
p=p0*E*expm(t*D)*inv(E);
f=inline('E*expm(t*D)*inv(E)');
ff=p0*f(D,E,t); % Example: p0*f(E,2,D) will compute the amount of p(2)
%Plot symbolic function
ezplot(mean(ff));

```

```
% Problem 12.17
% (i)

lambda=0.5;
mu=1;

v=-(lambda+mu);
a=lambda;
b=mu;
L=[-a 0 0 0
    b v a 0 0
    0 b v a 0
    0 0 b v a
    0 0 0 b v a
    0 0 0 0 b -b];

[E D]=eig(L);
t=sym('t');
N0=1;
p0=zeros(1,6);
p0(N0)=1;
p=p0*expm(t*D)*inv(E);
f=inline('E*expm(t*D)*inv(E)');
ff=p0*f(D,E,t); % Example: p0*f(E,2,D) will compute the amount of p(2)
%Plot symbolic function
ezplot(mean(ff));

L =
-0.5000 0.5000 0 0 0 0
1.0000 -1.5000 0.5000 0 0 0
0 1.0000 -1.5000 0.5000 0 0
0 0 1.0000 -1.5000 0.5000 0
0 0 0 1.0000 -1.5000 0.5000
0 0 0 0 1.0000 -1.0000

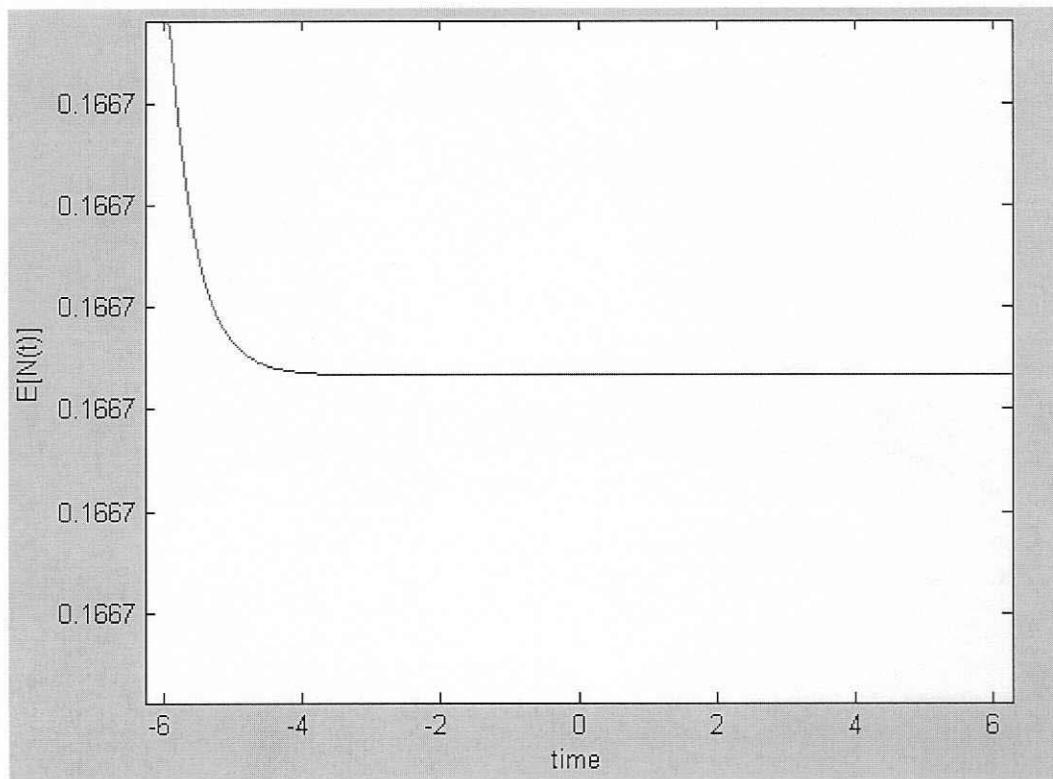
>> E

E =
0.0438 0.0733 -0.0976 -0.1091 -0.4082 -0.0957
-0.1950 -0.2504 0.1952 0.0639 -0.4082 -0.0430
0.3900 0.2074 0.1952 0.3084 -0.4082 0.0860
-0.5652 0.2074 -0.3904 0.3084 -0.4082 0.2967
0.6046 -0.7082 -0.3904 -0.1807 -0.4082 0.5547
-0.3506 0.5867 0.7807 -0.8724 -0.4082 0.7654

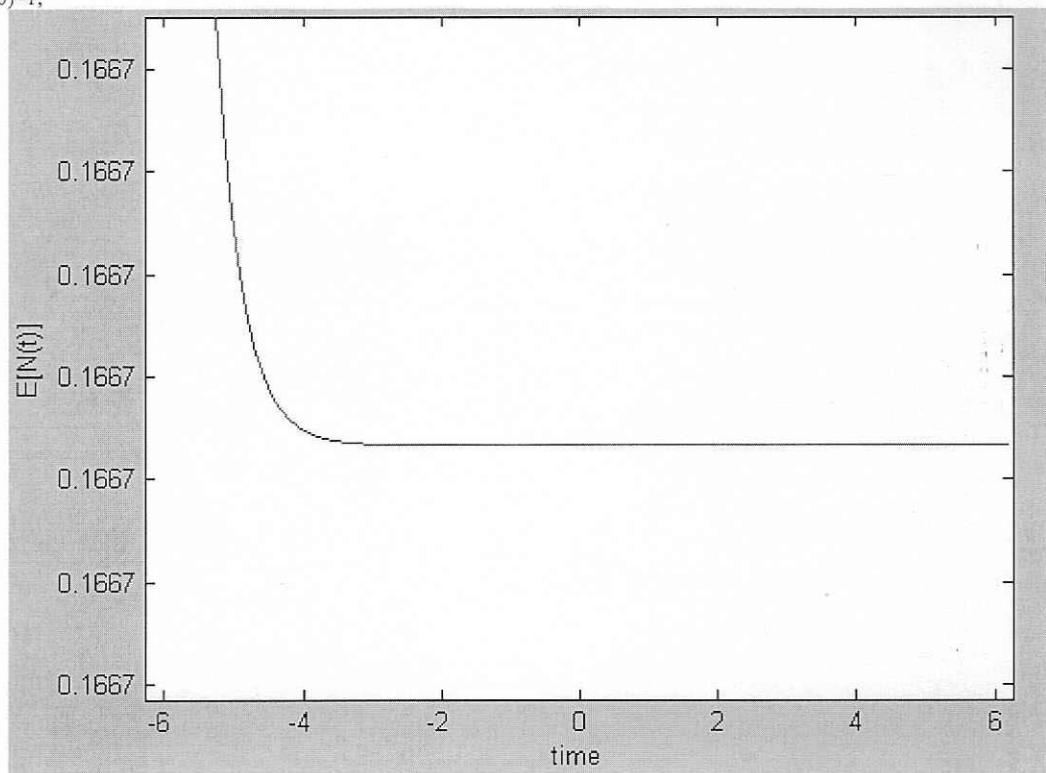
D =
-2.7247 0 0 0 0 0
0 -2.2071 0 0 0 0
0 0 -1.5000 0 0 0
0 0 0 -0.7929 0 0
0 0 0 0 -0.0000 0
0 0 0 0 0 -0.2753

>> inv(E)

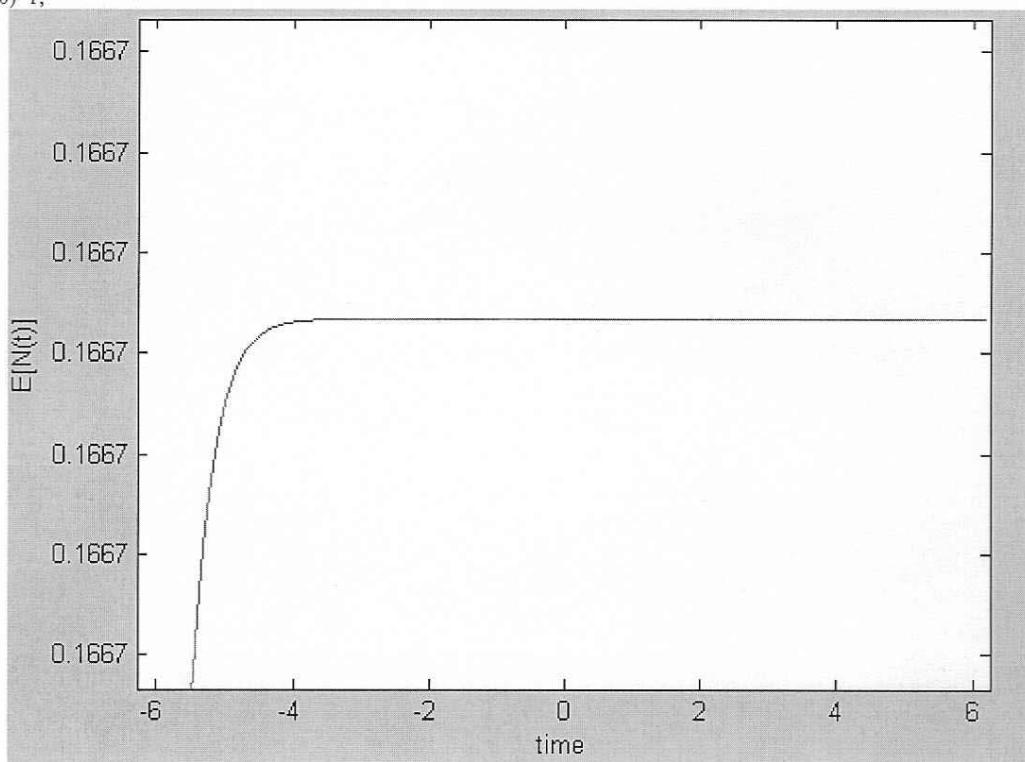
ans =
0.3490 -0.7764 0.7764 -0.5627 0.3009 -0.0872
0.7722 -1.3183 0.5460 0.2730 -0.4661 0.1931
-1.1386 1.1386 0.5693 -0.5693 -0.2846 0.2846
-1.4456 0.4234 1.0222 0.5111 -0.1497 -0.3614
-1.2442 -0.6221 -0.3110 -0.1555 -0.0778 -0.0389
-1.5822 -0.3556 0.3556 0.6133 0.5733 0.3955
```



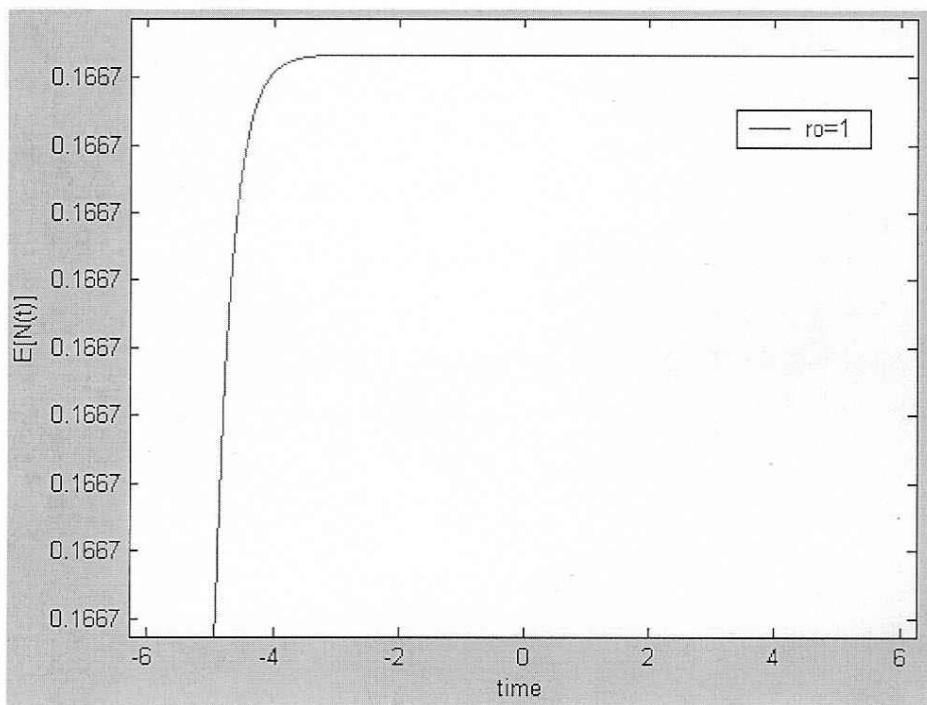
```
N0=3;  
p0=zeros(1,6);  
p0(N0)=1;
```



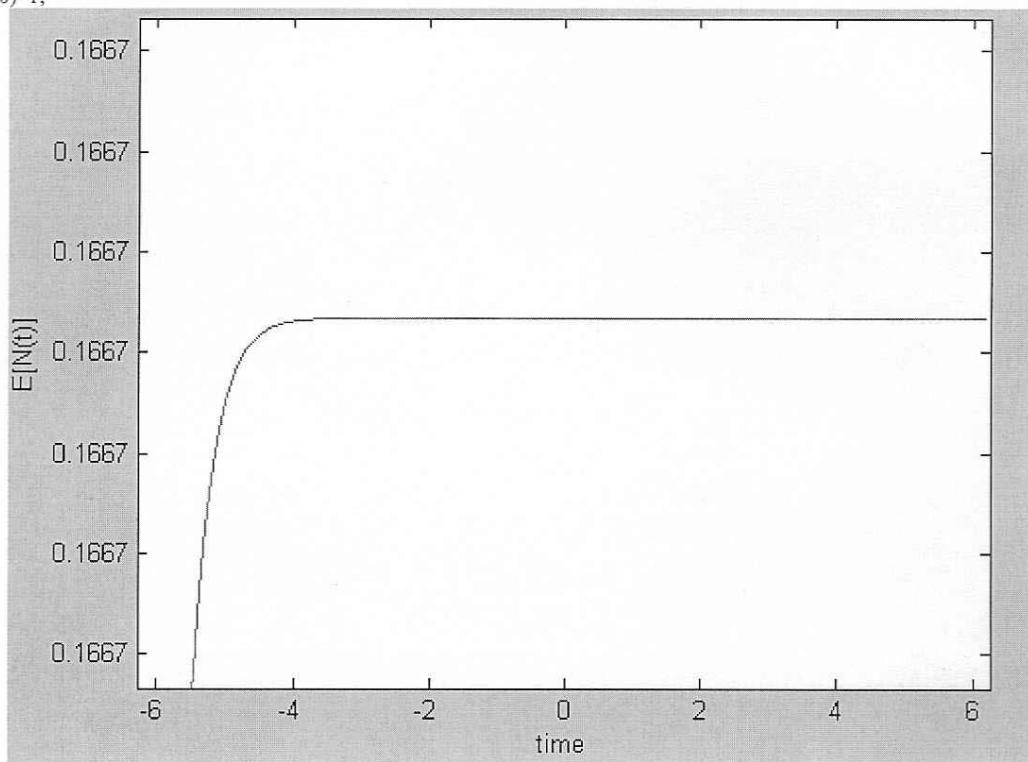
```
N0=6;
p0=zeros(1,6);
p0(N0)=1;
```



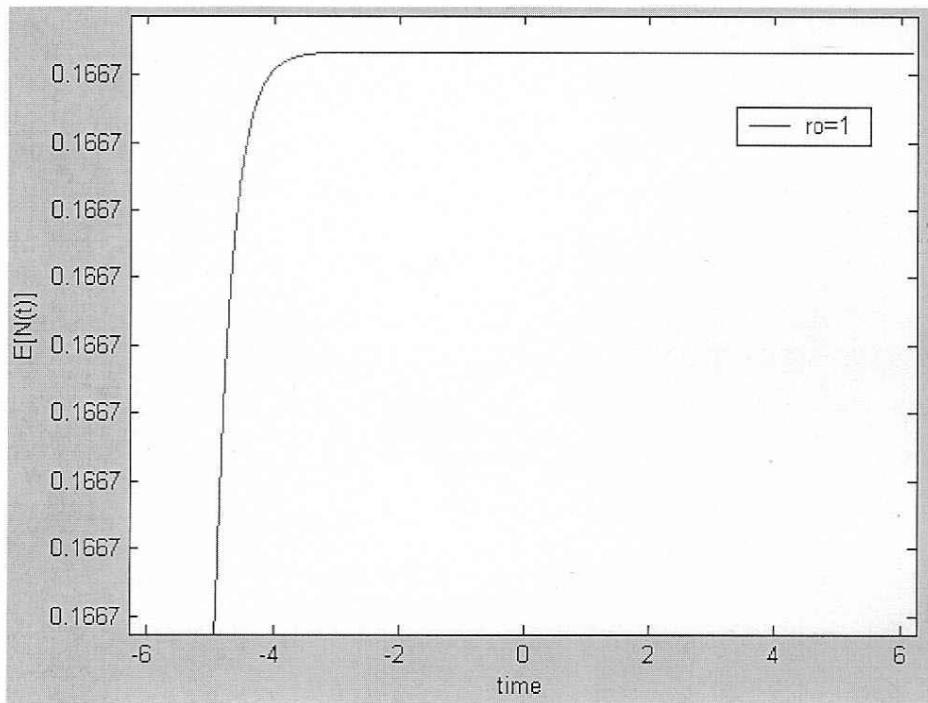
```
Part (ii)
Lambda=1;
N0=1;
p0=zeros(1,6);
p0(N0)=1;
```



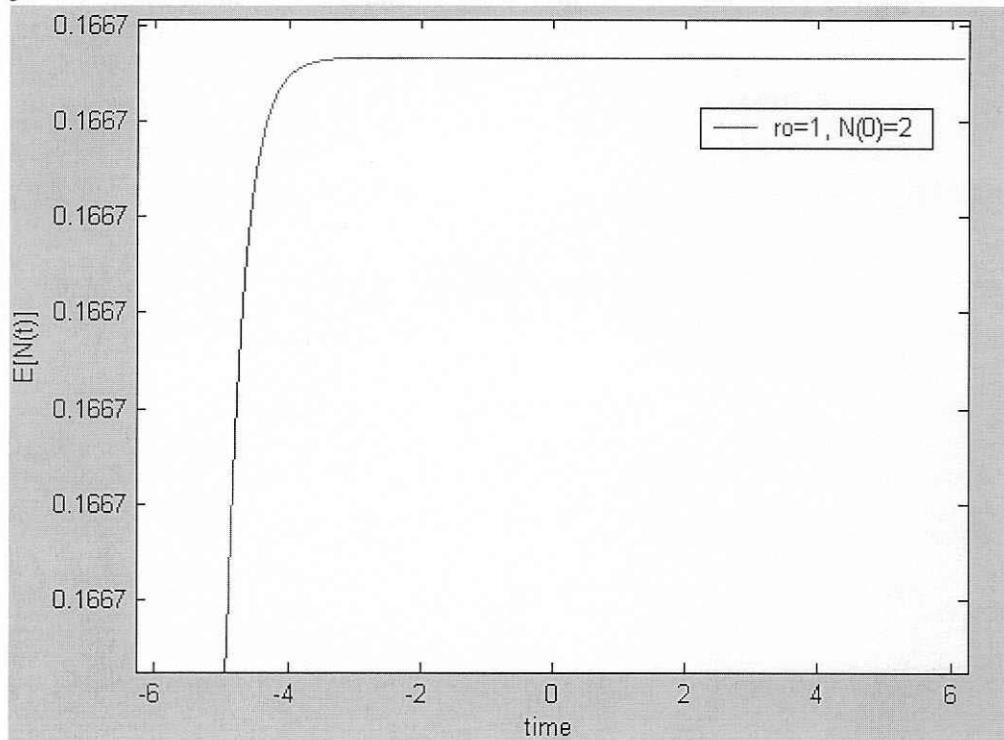
```
N0=6;
p0=zeros(1,6);
p0(N0)=1;
```



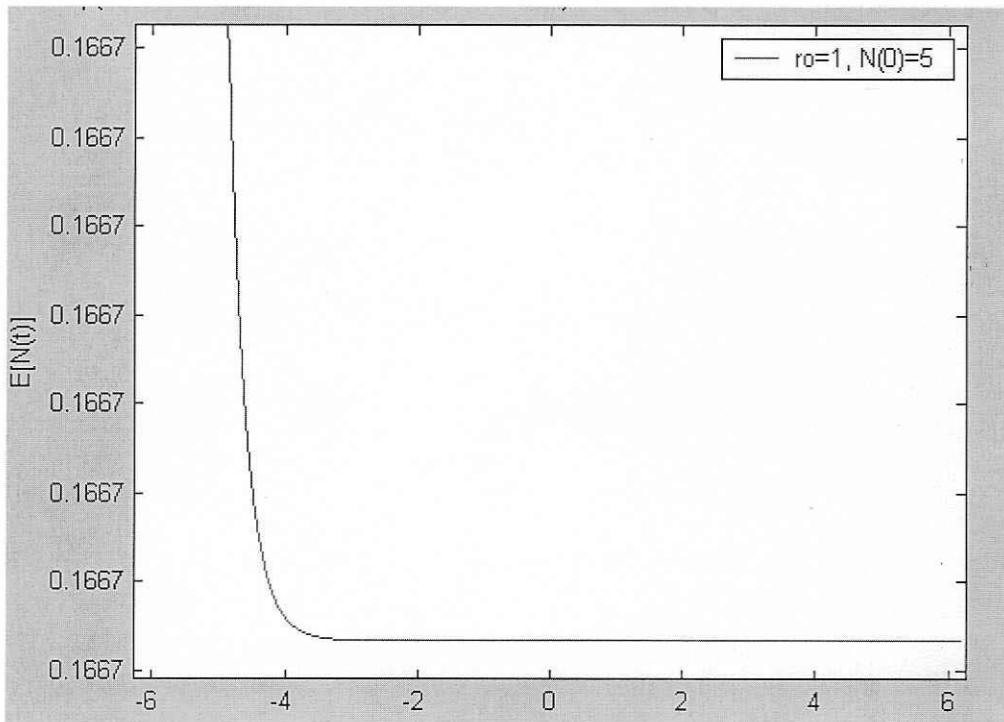
Part (ii)
 $\Lambda=1$;
 $N_0=1$;
 $p_0=zeros(1,6)$;
 $p_0(N_0)=1$;



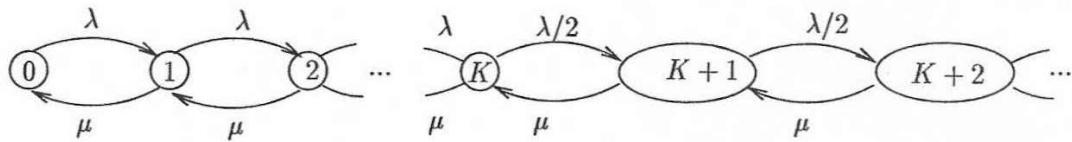
```
Lambda=1;  
N0=1;  
p0=zeros(1,6);  
p0(N0)=3
```



```
Lambda=1;  
N0=1;  
p0=zeros(1,6);  
p0(N0)=6
```



12.18

9.17 If $N < K$ arrival rate is λ If $N \geq K$ arrival rate is reduced to $\frac{\lambda}{2}$ For $0 \leq j \leq K$

$$P_j = \frac{\lambda}{\mu} P_{j-1} = \left(\frac{\lambda}{\mu}\right)^j P_0$$

For $K < j$

$$P_j = \frac{\lambda}{2\mu} P_{j-1} = \left(\frac{\lambda}{2\mu}\right)^{j-K} P_K = \left(\frac{\lambda}{2\mu}\right)^{j-K} \left(\frac{\lambda}{\mu}\right)^K P_0$$

$$\begin{aligned} 1 &= \sum_{j=0}^{\infty} P_j = P_0 \underbrace{\sum_{j=0}^{K-1} \left(\frac{\lambda}{\mu}\right)^j}_{\frac{1 - \left(\frac{\lambda}{\mu}\right)^K}{1 - \frac{\lambda}{\mu}}} + P_0 \underbrace{\sum_{j=K}^{\infty} \left(\frac{\lambda}{2\mu}\right)^{j-K} \left(\frac{\lambda}{\mu}\right)^K}_{\frac{\left(\frac{\lambda}{\mu}\right)^K}{1 - \frac{\lambda}{2\mu}}} \\ P_0 &= \left[\frac{1 - \left(\frac{\lambda}{\mu}\right)^K}{1 - \frac{\lambda}{\mu}} + \frac{\left(\frac{\lambda}{\mu}\right)^K}{1 - \frac{\lambda}{2\mu}} \right]^{-1} \end{aligned}$$

12.4 Multi-Server Systems: M/M/c, M/M/c/c, and M/M/∞

12.20

$$\begin{aligned} P[N \geq c+k] &= \sum_{j=c+k}^{\infty} \frac{\rho^{j-c}}{c!} a^c P_0 = \frac{a^c}{c!} P_0 \rho^k \sum_{j'=0}^{\infty} \rho^j \\ &= \frac{\frac{a^c}{c!} P_0 \rho^k}{1-\rho} = \frac{p_c \rho^k}{1-\rho} \end{aligned}$$

12.21

~~9.19~~ $\lambda = 12$ $\frac{1}{\mu} = \frac{5}{60}$ $c = 2$
 $\Rightarrow a = \frac{\lambda}{\mu} = 1$ $\rho = \frac{a}{2} = \frac{1}{2}$

a) $P[N \geq c] = \frac{\frac{a^c}{c!} P_0}{1-\rho} - C(c, a)$
 $P_0 = \left\{ \sum_{j=0}^1 \frac{a^j}{j!} + \frac{a^2}{2!} \sum_{j=0}^{\infty} \rho^j \right\}^{-1} = \left\{ 1 + 1 + \frac{1}{2} \frac{1}{1-\frac{1}{2}} \right\}^{-1} = \frac{1}{3}$
 $\Rightarrow P[N \geq 2] = \frac{\frac{1}{2!} \frac{1}{3}}{1-\frac{1}{2}} = \frac{1}{3} = C(2, 1)$

b) $\mathcal{E}[N] = \mathcal{E}[N_q] + a = \frac{\rho}{1-\rho} C(c, a) + a = \frac{\frac{1}{2}}{1-\frac{1}{2}} \frac{1}{3} + 1 = \frac{4}{3}$
 $\mathcal{E}[T] = \frac{1}{\lambda} \mathcal{E}[N] = \frac{1}{9}$

c) $P[N > 4] = P[N_q > 2] = \sum_{j=3}^{\infty} \rho^{j-2} P_2 = \frac{P_2 \rho}{1-\rho} = \frac{1}{6}$

12.22

$$\begin{aligned}
 \mathcal{E}[N_s] &= \sum_{j=0}^c j P_j + \sum_{j=c+1}^{\infty} c P_j \\
 &= \sum_{j=0}^c j \frac{a^j}{j!} P_0 + c \sum_{j=c+1}^{\infty} \rho^{j-c} \frac{a^c}{c!} P_0 \\
 &= \left[a \sum_{j=1}^c \frac{a^{j-1}}{(j-1)!} + \frac{ca^c}{c!} \frac{\rho}{1-\rho} \right] P_0 \quad \text{but } c\rho = a \\
 &= a \underbrace{\left[\sum_{j=0}^{c-1} \frac{a^j}{j!} + \frac{\frac{a^c}{c!}}{1-\rho} \right]}_{P_0^{-1}} P_0 = a = \lambda \mathcal{E}[\tau] \quad \checkmark
 \end{aligned}$$

12.23

$$9.21 \quad \lambda = 10 \quad \frac{1}{\mu} = \frac{1}{2} \quad a = \frac{\lambda}{\mu} = 5$$

$$\text{a) } \mathcal{E}[T] = \mathcal{E}[W] + \frac{1}{\mu} \leq 4 \Rightarrow \mathcal{E}[W] \leq 4 - \frac{1}{\mu} = 2$$

$$\mathcal{E}[W] = \frac{\frac{1}{\mu}}{c(1-\rho)} C(c, a) \leq 2 \Rightarrow C(c, a) \leq c(1-\rho)$$

$$C(c, a) \leq c - c\rho = c - a = c - 5$$

$$\Rightarrow c \geq 5 + C(c, a)$$

\Rightarrow try $c = 6$ and check waiting time requirement

$$C(6, 5) = \frac{p_0}{1-\rho} = 0.5875 \text{ where } p_0 = 0.004512$$

$$\therefore P[W \leq 8] = 1 - c(6, 5) e^{-6(\frac{1}{2})(1-\frac{5}{6})8} = 0.9892$$

$$\Rightarrow c = 6 \quad \text{OK}$$

$$p_0 = \left\{ \sum_{j=0}^5 \frac{5^j}{j!} + \frac{\frac{5^6}{6!}}{1-\frac{5}{6}} \right\}^{-1} = 0.004512$$

$$P[\text{all servers busy}] = P[N \geq c] = C(c, a) = 0.5875$$

(12.24)



$$\text{b) } P_j = \frac{\lambda}{c\mu} P_{j-1} \quad j \geq 1$$

$$= \left(\frac{\lambda}{c\mu} \right)^j P_0$$

$$\Rightarrow P_j = (1 - \rho)\rho^j \quad j \geq 0, \quad \rho = \frac{\lambda}{c\mu}$$

$$\text{c) } \mathcal{E}[N] = \frac{\rho}{1 - \rho} \quad \mathcal{E}[T] = \frac{1}{\lambda} \mathcal{E}[N] = \frac{\frac{1}{c\mu}}{1 - \rho}$$

$$\begin{aligned} \mathcal{E}[N_q] &= \sum_{k=c}^{\infty} (k - c) P_k = \sum_{k=c}^{\infty} (k - c)(1 - \rho)\rho^k \\ &= \rho^c (1 - \rho) \sum_{k'=0}^{\infty} k' \rho^{k'} = \frac{\rho^{c+1}}{1 - \rho} \end{aligned}$$

$$\mathcal{E}[W] = \frac{1}{\lambda} \mathcal{E}[N_q] = \frac{\rho^c}{1 - \rho} \frac{1}{c\mu}$$

d) M/M/1

$$\mathcal{E}[T] = \frac{\frac{1}{c\mu}}{1 - \rho} \quad \text{same as above system}$$

$$\mathcal{E}[W] = \frac{\rho}{1 - \rho} \frac{1}{c\mu}$$

M/M/2

$$C(2, a) = \frac{a^2/2}{1 - \rho} \left[1 + a + \frac{a^2/a}{1 - \rho} \right]^{-1} = \frac{2\rho^2}{1 + \rho} \quad \text{since } a = 2\rho$$

$$\mathcal{E}[W] = \frac{\frac{1}{\mu}}{c(1 - \rho)} C(c, a) = \frac{2\rho^2(1 + \rho)}{2\mu(1 - \rho)}$$

$$\mathcal{E}[T] = \mathcal{E}[W] + \frac{1}{\mu} = \frac{2/(1 + \rho)}{2\mu(1 - \rho)}$$

Comparison:

M/M/1 M/M/2 New System

$$\mathcal{E}[T] \quad \frac{1}{2\mu(1 - \rho)} \quad \frac{2/(1 + \rho)}{2\mu(1 - \rho)} \quad \frac{1}{2\mu(1 - \rho)}$$

$$\mathcal{E}[W] \quad \frac{\rho}{2\mu(1 - \rho)} \quad \frac{2\rho^2/(1 + \rho)}{2\mu(1 - \rho)} \quad \frac{1}{2\mu(1 - \rho)}$$

for $\mathcal{E}[T]$ New = M/M/1 < M/M/2for $\mathcal{E}[W]$ New < M/M/2 < M/M/1

12.25) for the queue of p. 12.24 we have:

$$P_i = \frac{\lambda}{c\mu} P_{i-1} \Rightarrow P_i = \left(\frac{\lambda}{c\mu}\right)^i P_0 = (1-p)^i$$

$$\sum_{j=0}^{\infty} P_j = 1 \Rightarrow P_0 + \sum_{j=1}^{\infty} P_j = 1 \Rightarrow \frac{1}{1-\frac{\lambda}{c\mu}} P_0 = 1 \Rightarrow P_0 = 1 - \frac{\lambda}{c\mu} = p$$

$$E(N) = \frac{p}{1-p}, \quad E(T) = \frac{E(W)}{\lambda} = \frac{\frac{1}{c\mu}}{1 - \frac{\lambda}{c\mu}} = \boxed{\frac{1}{c\mu - \lambda}}$$

$$E(N_q) = \sum_{i=c}^{\infty} (i-c) P_i = \sum_{i=c}^{\infty} (i-c)(1-p)^i = p^c(1-p) \sum_{j=0}^{\infty} j p^j$$

$$= p^c(1-p) \times \frac{p}{(1-p)^2} = \boxed{\frac{p^{c+1}}{1-p}}$$

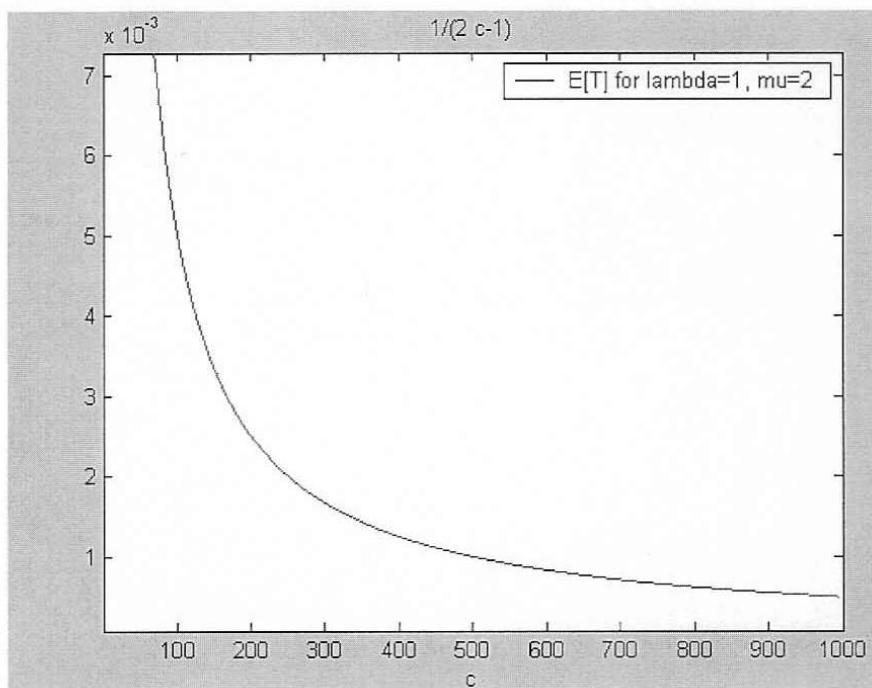
we write a simple code to plot $E[T] \leftarrow E(N_q)$

```
c=sym('c');
>> f=inline('1/(c*mu-lambda)');
>> f
```

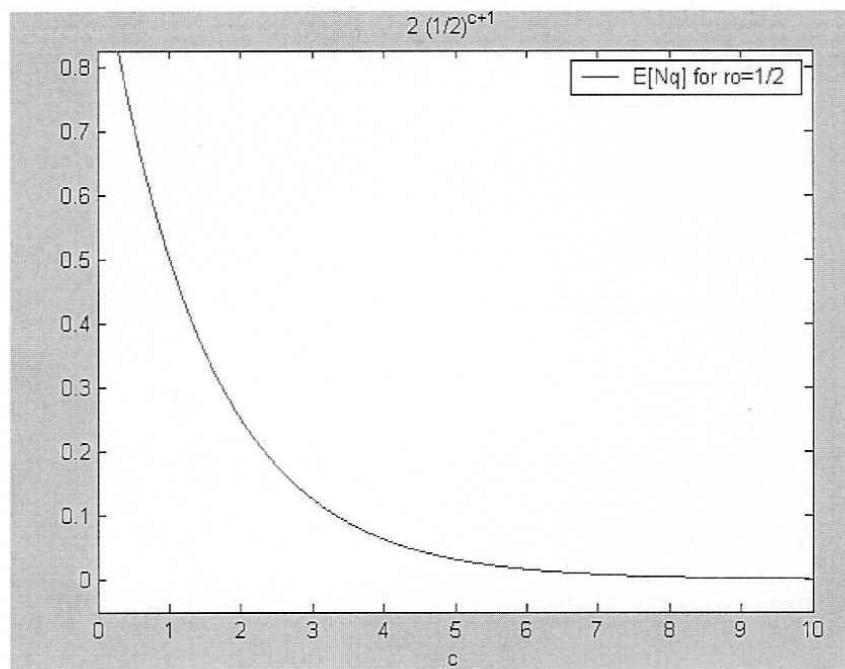
f =

Inline function:
 $f(c, \lambda, \mu) = 1/(c\mu - \lambda)$

```
>> ezplot(f(c,1,1))
```



```
f=inline('ro^(c+1)/(1-ro)');
ezplot(f(c,1/2))
```



12.26

$$\begin{aligned} B(c, a) &= \frac{\frac{a^c}{c!}}{\sum_{j=0}^c \frac{a^j}{j!}} = \frac{\frac{a^c}{c!}}{\frac{a^c}{c!} + \sum_{j=0}^{c-1} \frac{a^j}{j!}} = \frac{\frac{a^c}{c!}}{1 + \frac{a^c}{c!} / \sum_{j=0}^{c-1} \frac{a^j}{j!}} \\ &= \frac{\frac{a}{c} B(c-1, a)}{1 + \frac{a}{c} B(c-1, a)} = \frac{a B(c-1, a)}{c + a B(c-1, a)} \end{aligned}$$

12.27

9.24 $\lambda = 10$

$\frac{1}{\mu} = \frac{1}{2} \quad \frac{\lambda}{\mu} = 5 = a$

$B(0, 5) = 1$

$B(1, 5) = \frac{5 \cdot 1}{1 + 5 \cdot 1} = \frac{5}{6}$

$B(2, 5) = \frac{5 \left(\frac{5}{6}\right)}{2 + 5 \left(\frac{5}{6}\right)} = \frac{25}{37}$

$B(3, 5) = \frac{5 \left(\frac{25}{37}\right)}{3 + 5 \left(\frac{25}{37}\right)} = \frac{125}{236}$

$B(4, 5) = \frac{5 \left(\frac{125}{236}\right)}{4 + 5 \left(\frac{125}{236}\right)} = \frac{625}{1569}$

$\rightarrow B(5, 5) = \frac{5 \left(\frac{625}{1569}\right)}{5 + 5 \left(\frac{625}{1569}\right)} = \frac{625}{2194} \approx 28.5\%$

 \vdots

$B(8, 5) = 0.070 \quad \text{need 3 more servers}$

12.28

9.25 a) $\lambda = \frac{1}{2} \quad \frac{1}{\mu} = 2 \quad a = 1$

$B(4, 1) = \frac{1}{65} = 1.54\%$

b) $E[N] = a(1 - B(4, 1)) = \frac{64}{65} = 0.985$

c) $B(3, 1) = \frac{1}{16} = 6.25\% \text{ an increase of } 4.7\%$

12.29

9.26 For $a < c$

$$\begin{aligned}
 \text{a) } C(c, a) &= \frac{\frac{a^c}{c!} \frac{1}{1-\rho}}{\sum_{j=0}^{c-1} \frac{a^j}{j!} + \frac{a^c}{c!} \left(\frac{1}{1-\rho} \right)} = \frac{\frac{a^c}{c!}}{(1-\rho) \sum_{j=0}^{c-1} \frac{a^j}{j!} + \frac{a^c}{c!}} \quad \text{but } \rho = \frac{a}{c} \\
 &= \frac{\frac{a^c}{c!}}{\left(1 - \frac{a}{c}\right) \sum_{j=0}^{c-1} \frac{a^j}{j!} + \frac{a^c}{c!}} = \frac{\frac{a^c}{c!}}{(c-a) \sum_{j=0}^{c-1} \frac{a^j}{j!} + a \frac{a^{c-1}}{(c-1)!}} \\
 &= \frac{aB(c-1, a)}{(c-a) + 1B(c-1, a)} \quad \text{since } \frac{\frac{a^{c-1}}{(c-1)!}}{\sum_{j=0}^{c-1} \frac{a^j}{j!}} = B(c-1, a)
 \end{aligned}$$

$$\text{Problem 12.26} \Rightarrow aB(c-1, a) = \frac{cB(c, a)}{1 - B(c, a)}$$

$$\therefore C(c, a) = \frac{cB(c, a)}{(c-a)(1-B(c, a)) + cB(c, a)} = \frac{cB(c, a)}{c-a(1-B(c, a))}$$

$$\text{b) } C(c, a) = \frac{B(c, a)}{1 - \frac{a}{c}(1 - B(c, a))} > B(c, a) \text{ since}$$

$$\frac{1}{\underbrace{1 - \frac{a}{c}(1 - B(c, a))}_{<1}} > \frac{1}{1 - \underbrace{\frac{a}{c}}_{<1}} > 1$$

12.30

$$9.27 \lambda = 1 \quad \frac{1}{\mu} = 2 \quad a = \frac{\lambda}{\mu} = 2$$

$$\text{a) } P[\text{redirected}] = B(3, 2) = \frac{4}{19} = 21.1\%$$

$$\text{b) } P[\text{redirected}] = B(6, 4) = \frac{256}{2185} = 11.7\%$$

$$\lambda' = 2 \quad \frac{1}{\mu} = 2$$

12.31

M/M/10

$$2 \text{ cust/sec} \quad \frac{\lambda}{2} \text{ Type 1} \quad \bar{x}_1 = 1 \quad \bar{x} = \frac{1}{2} \bar{x}_1 + \frac{1}{2} \bar{x}_2 = \frac{1}{2}(1+3) = 2$$

$$\frac{\lambda}{2} \text{ Type 2} \quad \bar{x}_2 = 3$$

$$P_B = B(10, \alpha) \quad \alpha = \lambda \bar{x} = 2\lambda$$

alternative: 2 M/M/5 systems

$$P_{B_1} = B(5, \frac{\lambda}{2}) \quad \lambda=1 \quad 0\% \quad \lambda=2 \quad 0\% \quad \lambda=3 \quad 10\%$$

$$P_{B_2} = B(5, \frac{3\lambda}{2}) \quad 10\% \quad 11\% \quad 24\%$$

$$P_B = B(10, 2\lambda) \quad 0\% \quad 10\% \quad 40\%$$

Combined system does better for overall blocking;
 short service customers affected only at higher load

M/M/100

$$P_{B_1} = B(50, \frac{\lambda}{2}) \quad 0\% \quad 0\% \quad 0\%$$

$$P_{B_2} = B(50, \frac{3\lambda}{2}) \quad 31\% \quad 33\% \quad 35\%$$

$$P_B = B(100, 2\lambda) \quad 3\% \quad 5\% \quad 8\%$$

Combined system highly beneficial to customers with
 longer service time.

12.32

$$P[N = c] = \frac{a^c}{c!} e^{-a}$$

$$B(c, a) = \frac{\frac{a^c}{c!}}{\sum_{j=0}^c \frac{a^j}{j!}} < \frac{\frac{a^c}{c!}}{\sum_{j=0}^{\infty} \frac{a^j}{j!}} = \frac{a^c}{c!} e^{-a}$$

$\Rightarrow P[N = c]$ estimate is conservative

12.33

This system can be modeled with an M/M/ ∞

$$\lambda = 60 \quad \frac{1}{\mu} = 3600 \Rightarrow \mu = \frac{1}{3600}$$

$$\alpha = \frac{\lambda}{\mu} = 86,000$$

$$\Rightarrow P_j = \frac{\alpha^j}{j!} e^{-\alpha} \quad \text{a poisson dist.}$$

when α is large enough, the poisson dist. can be approximated with a Gaussian, which is the case in this problem

12.5 Finite-Source Queueing Systems

12.34

9.29 a)

$$\begin{aligned}\rho &= 1 - p_0 = 1 - \frac{1}{\sum_{k=0}^K \frac{K!}{(K-k)!} \left(\frac{\alpha}{\mu}\right)^k} \quad \text{let } j = K - k \\ &= 1 - \frac{\left(\frac{\mu}{\alpha}\right)^K / K!}{\sum_{j=0}^K \frac{(\mu\alpha)^j}{j!}} \\ &= 1 - B\left(K, \frac{\mu}{\alpha}\right) \quad \text{Erlang } B\end{aligned}$$

$$\begin{aligned}K &= 15 & \frac{1}{\mu} &= 2 & \frac{1}{\alpha} &= 30 & \frac{\mu}{\alpha} &= 15 \\ & & B(15, 15) &= 0.18\end{aligned}$$

$$\begin{aligned}\rho &= 1 - B\left(K, \frac{\mu}{\alpha}\right) = 11 - 0.18 = 0.82 \\ \lambda &= \mu\rho = \frac{1}{2}0.82 = 0.41 \\ \mathcal{E}[T] &= \frac{K}{\lambda} - \frac{1}{\alpha} = \frac{15}{0.41} - 30 = 6.6\end{aligned}$$

$$\text{b)} K^* = \frac{\frac{1}{\mu} + \frac{1}{\alpha}}{\frac{1}{\mu}} = \frac{32}{2} = 16$$

c) If we add 5 users we exceed K^* so

$$\begin{aligned}\mathcal{E}[T] &\approx \frac{K}{\mu} - \frac{1}{\alpha} = 20(2) - 30 = 10 \\ \lambda &\approx \mu = 2\end{aligned}$$

(12.35)

$$\text{a) } \frac{1}{\mu} = 0.01, \frac{1}{\lambda} = 5 \quad \frac{\lambda}{\mu} = \frac{1/5}{1/0.01} = 20$$

$$k^* \approx \frac{\frac{1}{\mu} + \frac{1}{\lambda}}{\frac{1}{\mu}} = \frac{0.01 + 5}{0.01} = 501$$

(b) The pmf of the number of request can be obtained using

$$\text{eq 12.84} \quad P[N_a=k] = \frac{[(K-1)! / (K-k-1)!] (\alpha/\mu)^k}{\sum_{j=0}^{K-1} [(K-1)! / (K-j-1)!] (\alpha/\mu)^j}$$

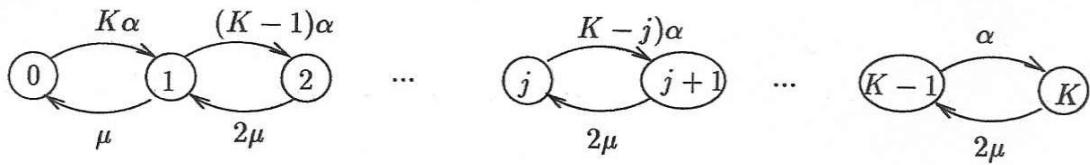
$$\Rightarrow P[N_a=k] = \frac{\frac{500!}{(500-k)!} (20)^k}{\sum_{j=0}^{500} \frac{500!}{(500-j)!} (20)^j}$$

% Problem 12.35

```
K=501;
alpha=1/5;
mu=100;
j=sym('j');
f=inline('factorial(500)/factorial(500-v)*(20)^v');
sum=0;
for j=0:(K-1)
    sum=sum+f(j);
end

sum
```

12.36



$$\begin{aligned}
 p_1 &= \frac{K\alpha}{\mu} p_0 \\
 p_{j+1} &= \frac{(K-j)\alpha}{2\mu} p_j \quad 1 < j \leq K-1 \\
 \Rightarrow p_j &= \frac{K(K-1)\dots(K-j+1)}{2^{j-1}} \left(\frac{\alpha}{\mu}\right)^j p_0 = 2 \frac{K!}{(K-j)!} \left(\frac{\alpha}{2\mu}\right)^j p_0 \\
 p_0 &= \left[1 + 2 \sum_{j=1}^K \frac{K!}{(K-j)!} \left(\frac{\alpha}{2\mu}\right)^j \right]^{-1}
 \end{aligned}$$

12.37

$$P[N_a = k] = \frac{\frac{(K-1)!(\alpha/\mu)^k}{(K-1-k)!}}{\sum_{k'=0}^{K-1} \frac{(K-1)!(\alpha/\mu)^{k'}}{(K-1-k')!}} = \frac{\frac{(\alpha/\mu)^k}{(K-1-k)!}}{\sum_{k'=0}^{K-1} \frac{(\alpha/\mu)^{k'}}{(K-1-k')!}}$$

$$\begin{aligned} \mathcal{E}[T] &= \frac{1}{\mu} \sum_{k=0}^{K-1} (k+1) P[N_a = k] \\ &= \frac{1}{\mu} \sum_{k=0}^{K-1} (k+1) \frac{\frac{(\alpha/\mu)^k}{(K-1-k)!}}{\sum_{k'=0}^{K-1} \frac{(\alpha/\mu)^{k'}}{(K-1-k')!}} \quad \text{Let } j = K-1-k, j' = K-1-k' \\ &= \frac{1}{\mu} \sum_{j=0}^{K-1} (K-j) \underbrace{\frac{\frac{(\mu/\alpha)^j}{j!}}{\sum_{j'=0}^{K-1} \frac{(\mu/\alpha)^{j'}}{j'!}}}_{\text{probs of M/M/K-1/K-1}} \\ &= \frac{1}{\mu} \left[K - \underbrace{\frac{\mu}{\alpha} \left(1 - B(K-1, \frac{\mu}{\alpha}) \right)}_{\text{mean # in M/M/K-1/K-1}} \right] \\ &= \frac{K}{\mu} - \frac{1}{\alpha} \left(1 - B(K-1, \frac{\mu}{\alpha}) \right) \end{aligned}$$

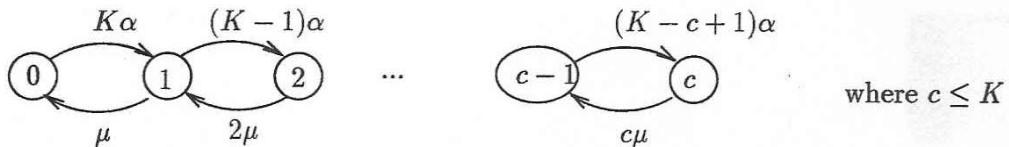
From Problem 12.26

$$\begin{aligned} B\left(K-1, \frac{\mu}{\alpha}\right) &= \frac{\frac{\alpha K}{\mu} B\left(K, \frac{\mu}{\alpha}\right)}{1 - B\left(K, \frac{\mu}{\alpha}\right)} \\ \mathcal{E}[T] &= \frac{K}{\mu} - \frac{1}{\alpha} + \frac{\frac{K}{\mu} B\left(K, \frac{\mu}{\alpha}\right)}{1 - B\left(K, \frac{\mu}{\alpha}\right)} \\ \mathcal{E}[T] &= \frac{K}{\mu} \left[1 + \frac{B\left(K, \frac{\mu}{\alpha}\right)}{1 - B\left(K, \frac{\mu}{\alpha}\right)} \right] - \frac{1}{\alpha} \\ &= \frac{K}{\mu} \frac{1}{1 - B\left(K, \frac{\mu}{\alpha}\right)} - \frac{1}{\alpha} \end{aligned}$$

But for Problem 12.34 solution

$$\begin{aligned} \rho &= \frac{\lambda}{\mu} = 1 - B\left(K, \frac{\mu}{\alpha}\right) \\ \Rightarrow \mathcal{E}[T] &= \frac{K}{\lambda} - \frac{1}{\alpha} \quad \text{as desired} \quad \checkmark \end{aligned}$$

12.38



$$K\alpha P_0 = \mu P_1 \Rightarrow P_1 = \frac{K\alpha}{\mu} P_0$$

$$(K-j+1)\alpha P_{j-1} + j\mu P_j \Rightarrow P_j = \frac{(K-j+1)\alpha}{j\mu} P_{j-1}$$

$$\Rightarrow P_j = \frac{K \dots (K-j+1)}{j!} \left(\frac{\alpha}{\mu}\right)^j P_0 = \underbrace{\frac{K!}{j!(K-j)!}}_{\binom{K}{j}} \left(\frac{\alpha}{\mu}\right)^j P_0$$

$$\Rightarrow P_0 = \left[\sum_{j=0}^c \binom{K}{j} \left(\frac{\alpha}{\mu}\right)^j \right]^{-1}$$

$$\therefore P_j = \frac{\binom{K}{j} \left(\frac{\alpha}{\mu}\right)^j}{\sum_{j'=0}^c \binom{K}{j'} \left(\frac{\alpha}{\mu}\right)^{j'}} = \frac{\binom{K}{j} \left(\frac{\alpha}{\alpha+\mu}\right)^j \left(\frac{\mu}{\alpha+\mu}\right)^{K-j}}{\sum_{j'=0}^c \binom{K}{j'} \left(\frac{\alpha}{\alpha+\mu}\right)^j \left(\frac{\mu}{\alpha+\mu}\right)^{K-j}}$$

$$= \frac{\binom{K}{j} p^j (1-p)^{K-j}}{\sum_{j'=0}^c \binom{K}{j'} p^{j'} (1-p)^{K-j'}}$$

$$\text{b) } P[\text{all servers busy}] = P_c = \frac{\binom{K}{c} p^c (1-p)^{K-c}}{\sum_{j'=0}^c \binom{K}{j'} p^{j'} (1-p)^{K-j'}}$$

c) Since an arriving customer sees the steady state of the system with one fewer server:

$P[\text{an arriving customer sees } c \text{ customers in system}]$

$$\begin{aligned} &= P_{K-1}[N = c] \\ &= \frac{\binom{K-1}{c} p^c (1-p)^{K-1-c}}{\sum_{j'=0}^c \binom{K-1}{j'} p^{j'} (1-p)^{K-1-j'}} \end{aligned}$$

12.39

The probability that all the servers are busy in Engset (P. 12.38) is:

$$P_{\text{busy}} = \frac{\sum_{i=0}^c \binom{K}{i} p^i (1-p)^{K-i}}{\sum_{i=0}^{\infty} \binom{K}{i} p^i (1-p)^{K-i}} \quad \text{where } p = \frac{\alpha}{\alpha + \mu}$$

$$\mu = \frac{1}{0.1} = 10 \Rightarrow p = \frac{1}{11} \quad c = 10$$

$$\alpha = 1$$

$$\text{we want } P_{\text{busy}} = 10\% = 0.1$$

$$\Rightarrow \frac{\sum_{i=0}^{10} \binom{K}{i} \left(\frac{1}{11}\right)^i \left(\frac{10}{11}\right)^{K-i}}{\sum_{i=0}^{\infty} \binom{K}{i} \left(\frac{1}{11}\right)^i \left(\frac{10}{11}\right)^{K-i}} = 0.1$$

12.6 M/G/1 Queueing Systems

12.40

9.33 A k -Erlang RV X with parameter k and λ has

$$\mathcal{E}[X] = \frac{k}{\lambda} \quad VAR[X] = \frac{k}{\lambda^2}$$

Since $\mathcal{E}[X] = \frac{1}{\mu}$ we have that $\lambda = k\mu$ and

$$\begin{aligned} VAR[X] &= \frac{k}{k^2\mu^2} = \frac{1}{k\mu^2} \\ \Rightarrow C_X^2 &= \frac{VAR[X]}{\mathcal{E}[X]^2} = \frac{1}{k} \\ \therefore \mathcal{E}[W]_{M/E_k/1} &= \frac{\rho(1 + C_X^2)}{2(1 - \rho)} \mathcal{E}[\tau] = \frac{\rho(1 + \frac{1}{k})}{2(1 - \rho)} \mathcal{E}[\tau] \end{aligned}$$

For M/M/1 we let $k = 1$ and obtain

$$\mathcal{E}[W]_{M/M/1} = \frac{2\rho}{2(1 - \rho)} \mathcal{E}[\tau]$$

For M/D/1 $C_X^0 = 0$ so

$$\mathcal{E}[W]_{M/D/1} = \frac{\rho}{2(1 - \rho)} \mathcal{E}[\tau]$$

$$\therefore \mathcal{E}[W]_{M/D/1} < \mathcal{E}[W]_{M/E_k/1} \leq \mathcal{E}[W]_{M/M/1}$$

Since $\mathcal{E}[T] = \mathcal{E}[W] + \mathcal{E}[\tau]$ the same ordering applies for total delay.

12.41

$$\begin{aligned} \mathcal{E}[\tau] &= \mathcal{E}[\tau|1]p + \mathcal{E}[\tau|2](1-p) = \frac{1}{\mu_1}p + \frac{1}{\mu_2}(1-p) \\ \mathcal{E}[\tau^2] &= \mathcal{E}[\tau^2|1]p + \mathcal{E}[\tau^2|2](1-p) = \frac{2}{\mu_1^2}p + \frac{2}{\mu_2^2}(1-p) \\ \mathcal{E}[W] &= \frac{\lambda\mathcal{E}[\tau^2]}{2(1-\rho)} = \frac{\lambda/2}{1-\rho} \left[\frac{2}{\mu_1^2}p + \frac{2}{\mu_2^2}(1-p) \right] \\ \mathcal{E}[T] &= \mathcal{E}[W] + \mathcal{E}[\tau] \end{aligned}$$

where $\rho = \lambda\mathcal{E}[\tau]$.

12.42

$$\begin{aligned}\mathcal{E}[\tau] &= \mathcal{E}[\tau|1]\alpha + \mathcal{E}[\tau|2](1-\alpha) = d\alpha + \frac{1}{\mu}(1-\alpha) \\ \mathcal{E}[\tau^2] &= \mathcal{E}[\tau^2|1]\alpha + \mathcal{E}[\tau^2|2](1-\alpha) = d^2\alpha + \frac{2}{\mu^2}(1-\alpha) \\ \mathcal{E}[W] &= \frac{\lambda\mathcal{E}[\tau^2]}{2(1-\rho)} = \frac{\lambda/2}{1-\rho} \left[\alpha d^2 + (1-\alpha) \frac{2}{\mu^2} \right] \\ \mathcal{E}[T] &= \mathcal{E}[W] + \mathcal{E}[\tau]\end{aligned}$$

where $\rho = \lambda\mathcal{E}[\tau]$.

12.43

$$\begin{aligned}\tau &= d + \tau_1 \\ \mathcal{E}[\tau] &= d + \mathcal{E}[\tau_1] = d + \frac{1}{\mu} \\ \mathcal{E}[\tau^2] &= d^2 + 2d\mathcal{E}[\tau_1] + \mathcal{E}[\tau_1^2] \\ &= d^2 + \frac{2d}{\mu} + \frac{2}{\mu^2} \\ \mathcal{E}[W] &= \frac{\lambda/2}{1-\rho} \mathcal{E}[\tau^2] = \frac{\lambda/2}{1-\rho} \left[d^2 + \frac{2d}{\mu} + \frac{2}{\mu^2} \right] \\ \mathcal{E}[T] &= \mathcal{E}[W] + \mathcal{E}[\tau]\end{aligned}$$

12.44

9.37 A message is transmitted until a successful acknowledgement is received:

a) $P[\tau = kd] = (1-p)p^{k-1} \quad k = 1, 2, \dots$

$$\mathcal{E}[\tau] = \frac{d}{1-p} \quad VAR[\tau] = \frac{d^2 p}{(1-p)^2}$$

b) $C_\tau^2 = \frac{VAR[\tau]}{\mathcal{E}[\tau]^2} = \frac{d^2 p}{(1-p)^2} \frac{(1-p)^2}{d^2} = p$

$$\begin{aligned}\mathcal{E}[T] &= \mathcal{E}[\tau] + \mathcal{E}[\tau] \frac{\rho}{2(1-\rho)} (1 + C_\tau^2) \quad \text{where } \rho = \frac{\lambda d}{1-p} \\ &= \frac{2 - \lambda d}{2(1-p - \lambda d)} d\end{aligned}$$

12.47

- 9.38 a)** Let τ = total job time, X = service time, $N(X) = \#$ breakdowns during X , R_i = repair times

$$\tau = X + \sum_{i=1}^{N(X)} R_i$$

where $N(X)$ is the total number of times the machine breaks down. To find $C[\tau]$ we use conditional expectation:

$$\begin{aligned}\mathcal{E}[\tau] &= \mathcal{E}[\mathcal{E}[\tau|X]] \\ \mathcal{E}[\tau|X=t] &= t + \mathcal{E}\left[\sum_{i=1}^{N(t)} R_i\right] = t + \alpha t \mathcal{E}[R] \quad \text{from Eq. 5.13} \\ \Rightarrow \mathcal{E}[\tau] &= \mathcal{E}[X + \alpha X \mathcal{E}[R]] = \underbrace{\mathcal{E}[X] + \alpha \mathcal{E}[X] \mathcal{E}[R]}_{\mathcal{E}[X](1+\alpha \mathcal{E}[R])} = \frac{1}{\mu} \left[1 + \frac{\alpha}{\beta}\right]\end{aligned}$$

We also use conditional expectation to find $E[\tau^2]$:

$$\begin{aligned}\mathcal{E}[\tau^2] &= \mathcal{E}[\mathcal{E}[\tau^2|X]] \\ \mathcal{E}[\tau^2|X=t] &= \mathcal{E}\left[\left(t + \sum_{i=1}^{N(t)} R_i\right)^2\right] \\ &= t^2 + 2t \mathcal{E}\left[\sum_{i=1}^{N(t)} R_i\right] + \mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2\right] \\ &= t^2 + 2t(\alpha t + \mathcal{E}[R]) + \mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2\right] \\ \mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2\right] &= \mathcal{E}\left[\mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2 | N(t)\right]\right] \\ \mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2 | N(t)=k\right] &= \mathcal{E}\left[\sum_{i=1}^k \sum_{j=1}^k R_i R_j\right] \\ &= k \mathcal{E}[R^2] + (k^2 - k) \mathcal{E}[R]^2 \\ \therefore \mathcal{E}\left[\left(\sum_{i=1}^{N(t)} R_i\right)^2\right] &= \mathcal{E}[N(t) \mathcal{E}[R^2] + [N^2(t) - N(t)] \mathcal{E}[R]^2] \\ &= \mathcal{E}[N(t)] \mathcal{E}[R^2] + (\mathcal{E}[N^2(t)] - \mathcal{E}[N(t)]) \mathcal{E}[R]^2\end{aligned}$$

$$\begin{aligned}
 &= \alpha t \mathcal{E}[R^2] + (\alpha t + (\alpha t)^2 - \alpha t) \mathcal{E}[R]^2 \\
 &= \alpha t \mathcal{E}[R^2] + \alpha^2 t^2 \mathcal{E}[R]^2 \\
 \therefore \mathcal{E}[\tau^2 | X = t] &= t^2 + 2\alpha t^2 \mathcal{E}[R] + \alpha t \mathcal{E}[R^2] + \alpha^2 t^2 \mathcal{E}[R]^2
 \end{aligned}$$

finally

$$\begin{aligned}
 \mathcal{E}[\tau^2] &= \mathcal{E}[X^2 + 2\alpha X^2 \mathcal{E}[R] + \alpha X \mathcal{E}[R^2] + \alpha^2 X^2 \mathcal{E}[R]^2] \\
 &= \mathcal{E}[X^2] \underbrace{[1 + 2\alpha \mathcal{E}[R] + \alpha^2 \mathcal{E}[R]^2]}_{(1+\alpha \mathcal{E}[R])^2} + \mathcal{E}[X] \alpha \mathcal{E}[R^2]
 \end{aligned}$$

$$\begin{aligned}
 \text{VAR}[\tau] &= \mathcal{E}[\tau^2] - \mathcal{E}[\tau]^2 \\
 &= \mathcal{E}[X^2] (1 + \alpha \mathcal{E}[R])^2 + \mathcal{E}[X] \alpha \mathcal{E}[R^2] - \mathcal{E}[X]^2 (1 + \alpha \mathcal{E}[R])^2 \\
 &= \text{VAR}[X] (1 + \alpha \mathcal{E}[R])^2 + \mathcal{E}[X] \alpha \mathcal{E}[R^2] \\
 &= \frac{1}{\mu^2} \left(1 + \frac{\alpha}{\beta}\right)^2 + \frac{\alpha}{\mu} \frac{2}{\beta^2}
 \end{aligned}$$

b) The coefficient of variation of τ is:

$$C_\tau^2 = \frac{\text{VAR}[\tau]}{\mathcal{E}[\tau]^2} = \frac{\frac{1}{\mu^2} \left(1 + \frac{\alpha}{\beta}\right)^2 + \frac{\alpha}{\mu} \frac{2}{\beta^2}}{\frac{1}{\mu^2} \left(1 + \frac{\alpha}{\beta}\right)^2} = 1 + \frac{2\alpha}{(\alpha + \beta)^2}$$

Thus the mean delay in the system is

$$\begin{aligned}
 \mathcal{E}[T] &= \mathcal{E}[\tau] + \mathcal{E}[\tau] \frac{\rho}{2(1-\rho)} (1 + C_\tau^2) \\
 &= \mathcal{E}[\tau] \left[1 + \frac{\rho}{(1-\rho)} \left(1 + \frac{\alpha}{(\alpha + \beta)^2} \right) \right]
 \end{aligned}$$

where

$$\rho = \lambda \mathcal{E}[\tau] = \frac{\lambda}{\mu} \left[1 + \frac{\alpha}{\beta} \right]$$

12.48

9.39 a) The proportion of time that the server works on low priority jobs is

$$\begin{aligned}\rho'_2 &= 1 - \rho_1 = \lambda'_2 \mathcal{E}[\tau_2] \\ \Rightarrow \lambda'_2 &= \frac{1 - \rho_1}{\mathcal{E}[\tau_2]} = \frac{1 - \lambda_1 \mathcal{E}[\tau_1]}{\mathcal{E}[\tau_2]}\end{aligned}$$

b) From Eq. (12.105)

$$\begin{aligned}\mathcal{E}[W_1] &= \frac{\lambda_1 \mathcal{E}[\tau_1^2] + \lambda'_2 \mathcal{E}[\tau_2^2]}{2(1 - \rho_1)} \\ &= \frac{\lambda_1 \mathcal{E}[\tau_1^2]}{2(1 - \rho_1)} + \frac{\lambda'_2 \mathcal{E}[\tau_2^2]}{2(1 - \rho_1)} \\ &= \frac{\frac{\lambda_1}{2} \mathcal{E}[\tau_1^2]}{1 - \lambda_1 \mathcal{E}[\tau_1]} + \frac{\mathcal{E}[\tau_2^2]}{2 \mathcal{E}[\tau_2]}\end{aligned}$$

since $\rho_1 = \lambda_1 \mathcal{E}[\tau_1]$, $1 - \rho_1 = \lambda'_2 \mathcal{E}[\tau_2]$

12.49

9.40 The server vacations can be viewed as the servicing of a low priority class of fictitious customers whose service times are a vacation time and whose arrival rate saturates the system. The result of Problem 9.39b then implies

$$\mathcal{E}[W] = \frac{\frac{1}{2} \lambda \mathcal{E}[\tau^2]}{1 - \lambda \mathcal{E}[\tau]} + \frac{\mathcal{E}[V^2]}{2 \mathcal{E}[V]}$$

where W and τ correspond to the real customers.

12.50

9.41 If we suppose that the server in Problem 12.49 takes vacations of fixed duration d , then we have the system described in the problem. Thus

$$\mathcal{E}[W] = \frac{\frac{1}{2} \lambda d^2}{1 - \lambda d} + \frac{d^2}{2d} = \frac{\frac{1}{2} \lambda d^2}{1 - \lambda d} + \frac{d}{2}$$

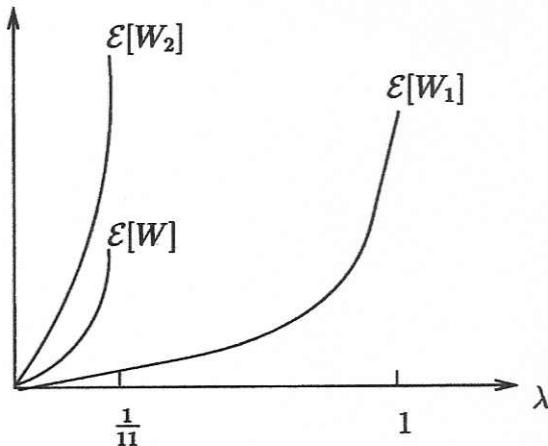
12.51

$$\begin{aligned}\mathcal{E}[\tau^2] &= \mathcal{E}[\tau^2|\text{type 1}]P[\text{type 1}] + \mathcal{E}[\tau^2|\text{type 2}]P[\text{type 2}] \\ &= \frac{2}{\mu_1^2} \frac{\lambda_1}{\lambda_1 + \lambda_2} + \frac{2}{\mu_2^2} \frac{\lambda_2}{\lambda_1 + \lambda_2} \\ &= 1 + 100 = 101\end{aligned}$$

where $\lambda = \lambda_1 = \lambda_2$ and $\mu_1 = 1$, $\mu_2 = \frac{1}{10}$.

$$\mathcal{E}[W_1] = \frac{\mathcal{E}[R'']}{1 - \rho_1} = \frac{\frac{2\lambda}{2} \mathcal{E}[\tau^2]}{1 - \rho_1} = \frac{101\lambda}{1 - \lambda} \quad \text{since } \rho_1 = \lambda/\mu_1 = \lambda$$

$$\begin{aligned}\mathcal{E}[W_2] &= \frac{101\lambda}{(1 - \lambda)(1 - \lambda - 10\lambda)} \quad \text{since } \rho_2 = \lambda/\mu_2 = 10\lambda \\ &= \frac{101\lambda}{(1 - \lambda)(1 - 11\lambda)} \\ \mathcal{E}[W] &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \mathcal{E}[W_1] + \frac{\lambda_2}{\lambda_1 + \lambda_2} \mathcal{E}[W_2] \\ &= \frac{1}{2} \frac{101\lambda}{1 - \lambda} \left[1 + \frac{1}{1 - 11\lambda} \right]\end{aligned}$$



12.52

9.43 a) The low priority customers are “invisible” to the high priority customers. Thus the mean waiting time and delay of high priority customers is that of a single-class M/G/1 system:

$$\begin{aligned}\mathcal{E}[W_1] &= \frac{\lambda_1 \mathcal{E}[\tau_1^2]}{2(1 - \rho_1)} \\ \mathcal{E}[T_1] &= \mathcal{E}[W_1] + \mathcal{E}[\tau]\end{aligned}$$

b) The time required to service all customers found by a low priority arrival is the time required to service all such customers in an ordinary M/G/1 system in which both classes are combined and neither receives priority. The reason for this is that the priority mechanism alters the order in which customers are served but not the rate at which the backlog is reduced. The mean waiting time in such a system is

$$\frac{\lambda \mathcal{E}[\tau^2]}{2(1 - \rho)} = \frac{\frac{1}{2} \sum_{j=1}^2 \lambda_j \mathcal{E}[\tau_j^2]}{(1 - \rho_1 - \rho_2)}$$

since $\lambda = \lambda_1 + \lambda_2$, $\rho = \rho_1 + \rho_2$

c) The mean time required to service all the high priority customers that arrive while a low priority customer is in the system is

$$\begin{aligned}\mathcal{E}\left[\sum_{i=1}^{N_1(T_2)} \tau_{i_1}\right] &= \mathcal{E}[N_1(T_2)] \mathcal{E}[\tau_1] = \lambda_1 \mathcal{E}[T_2] \mathcal{E}[\tau_1] \\ &= \rho_1 \mathcal{E}[T_2]\end{aligned}$$

d)

$$\begin{aligned}\mathcal{E}[T_2] &= \frac{R_2}{1 - \rho_1 - \rho_2} + \rho_1 \mathcal{E}[T_2] + \frac{1}{\mu_2} \\ \mathcal{E}[T_2] &= \frac{R_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{\frac{1}{\mu}}{1 - \rho_1} \\ &= \frac{\frac{1}{\mu}(1 - \rho_1 - \rho_2) + R_2}{(1 - \rho_1)(1 - \rho_1 - \rho_2)}\end{aligned}$$

12.53

9.44 $\lambda = \lambda_1 = \lambda_2$ $\mu_1 = 1$ $\mu_2 = \frac{1}{10}$

$$\begin{aligned}\mathcal{E}[W_1] &= \frac{\lambda_1 \mathcal{E}[\tau_1^2]}{2(1 - \rho_1)} = \frac{\lambda}{1 - \lambda} \\ \mathcal{E}[T_1] &= \frac{\lambda}{1 - \lambda} + 1 = \frac{1}{1 - \lambda} \\ R_2 &= \frac{1}{2} \lambda_1 \frac{2}{\mu_1^2} + \frac{1}{2} \lambda_2 \frac{2}{\mu_2^2} = \lambda + 100\lambda = 101\lambda \\ \mathcal{E}[T_2] &= \frac{101\lambda}{(1 - \lambda)(1 - 11\lambda)} + \frac{10}{1 - \lambda} = \frac{10 - 9\lambda}{(1 - \lambda)(1 - 11\lambda)}\end{aligned}$$

The mean waiting time and delay of class 1 is reduced greatly while those of class 2 are not significantly affected relative to the corresponding values for a non-preemptive priority system.

12.7 M/G/1 Analysis Using Embedded Markov Chains

$$12.54 \quad \rho = \frac{\lambda}{\mu} = \left(\frac{\mu}{2}\right) / \mu = \frac{1}{2}$$

a) For an M/G/1 system we have:

$$G_N(z) = \frac{(1-\rho)(z-1)\hat{r}(\lambda(1-z))}{z - \hat{r}(\lambda(1-z))}$$

where

$$\begin{aligned} \hat{r}(\lambda(1-z)) &= \frac{4\mu^2}{(s+2\mu)^2} \Big|_{s=\lambda(1-z)} = \frac{4\mu^2}{(\lambda - \lambda z + 2\mu)^2} \\ \Rightarrow G_N(z) &= \frac{\left(1 - \frac{1}{2}\right)(z-1)4\mu^2}{z(\lambda - \lambda z + 2\mu) - 4\mu^2} = \frac{8}{z^2 - 9z + 16} \\ &\text{where we used the fact that } \frac{\lambda}{\mu} = \frac{1}{2} \\ &= \frac{8}{(z - z_1)(z - z_2)} \quad z_1 = \frac{9 + \sqrt{17}}{2} \quad z_2 = \frac{9 - \sqrt{17}}{2} \\ &= \frac{8/z_1 z_2}{\left(1 - \frac{z}{z_1}\right)\left(1 - \frac{z}{z_2}\right)} = \frac{\frac{1}{2}}{\left(1 - \frac{1}{z_1}z\right)\left(1 - \frac{1}{z_2}z\right)} \\ &= \frac{A}{1 - \frac{1}{z_1}z} + \frac{B}{1 - \frac{1}{z_2}z} \Rightarrow \begin{array}{l} A = \frac{-z_2/2}{z_1 - z_2} \\ B = \frac{z_1/2}{z_1 - z_2} \end{array} \quad \text{partial fraction expansion} \\ &= \frac{z_1/2}{(z_1 - z_2)\left(1 - \frac{1}{z_2}z\right)} = \frac{z_2/2}{(z_1 - z_2)\left(1 - \frac{z}{z_1}\right)} \\ &= \frac{1}{2(z_1 - z_2)} \left[z_1 \sum_{j=0}^{\infty} \left(\frac{z}{z_2}\right)^j - z_2 \sum_{j=0}^{\infty} \left(\frac{z}{z_1}\right)^j \right] \\ \therefore P[N=j] &= \frac{z_1}{2(z_1 - z_2)} \left(\frac{1}{z_2}\right)^j - \frac{z_2}{2(z_1 - z_2)} \left(\frac{1}{z_1}\right)^j \quad \text{coefficient of } Z^j \\ P[N=j] &= \frac{9 + \sqrt{17}}{4\sqrt{17}} \left(\frac{2}{9 - \sqrt{17}}\right)^j - \frac{9 - \sqrt{17}}{4\sqrt{17}} \left(\frac{2}{9 + \sqrt{17}}\right)^j \\ &= \frac{8}{\sqrt{17}} \left(\frac{2}{9 - \sqrt{17}}\right)^{j+1} - \frac{8}{\sqrt{17}} \left(\frac{2}{9 + \sqrt{17}}\right)^j \quad j = 0, 1, \dots \end{aligned}$$

b) The Laplace Transform of the waiting time is:

$$\begin{aligned} \hat{W}(s) &= \frac{(1-\rho)s}{s - \lambda + \lambda \hat{r}(s)} = \frac{\frac{1}{2}s}{s - \lambda + \frac{\lambda 4\mu^2}{(s+2\mu)^2}} = \frac{1}{2} \left[\frac{s^2 + 8\lambda s + 16\lambda^2}{s^2 + 7\lambda s + 8\lambda^2} \right] \\ &= \frac{1}{2} \left[1 + \frac{\left(\frac{\sqrt{17}+9}{2\sqrt{17}}\right)\lambda}{s + \left(\frac{7-\sqrt{17}}{2}\right)\lambda} + \frac{\left(\frac{\sqrt{17}-9}{2\sqrt{17}}\right)\lambda}{s + \left(\frac{7+\sqrt{17}}{2}\right)\lambda} \right] \end{aligned}$$

$$\begin{aligned}
 f_W(t) &= \mathcal{L}^{-1}[\hat{W}(s)] \\
 &= \frac{1}{2}\delta(t) + \frac{1}{2} \left(\frac{\sqrt{17} + 9}{2\sqrt{17}} \right) \lambda e^{-\left(\frac{7-\sqrt{17}}{2}\right)\lambda t} u(t) \\
 &\quad + \frac{1}{2} \left(\frac{\sqrt{17} - 9}{2\sqrt{17}} \right) \lambda e^{-\left(\frac{7+\sqrt{17}}{2}\right)\lambda t} u(t)
 \end{aligned}$$

The total delay transform is:

$$\begin{aligned}
 \hat{T}(s) &= \frac{(1-\rho)s\hat{\tau}(s)}{s-\lambda+\lambda\hat{\tau}(s)} = \frac{\frac{1}{2}s\frac{4\mu^2}{(s+2\mu)^2}}{s-\lambda+\lambda\frac{4\mu^2}{(s+2\mu)^2}} \\
 &= \frac{8\lambda^2}{s^2+7\lambda s+8\lambda^2} \\
 \hat{T}(s) &= \frac{8\lambda}{\sqrt{17}} \left[\frac{1}{s+\left(\frac{7-\sqrt{17}}{2}\right)\lambda} - \frac{1}{s+\left(\frac{7+\sqrt{17}}{2}\right)\lambda} \right] \\
 f_T(t) &= \mathcal{L}^{-1}[\hat{T}(s)] = \frac{8\lambda}{\sqrt{17}} \left[e^{-\left(\frac{7-\sqrt{17}}{2}\right)\lambda t} - e^{-\left(\frac{7+\sqrt{17}}{2}\right)\lambda t} \right] u(t)
 \end{aligned}$$

12.55

~~12.46~~ a) $\tau = X + \sum_{i=1}^{N(X)} R_i$ (see solution to 12.47)

$$\begin{aligned}
 \hat{\tau}(s) &= \mathcal{E}[e^{-s\tau}] = \mathcal{E}[\mathcal{E}[e^{-s\tau}|X]] \\
 \mathcal{E} \left[e^{-s \left(X + \sum_{i=1}^{N(x)} R_i \right)} | X = t \right] &= e^{-st} \mathcal{E} \left[e^{-s \sum_{i=1}^{N(t)} R_i} \right] \\
 \mathcal{E} \left[e^{-s \sum_{i=1}^{N(t)} R_i} \right] &= \mathcal{E} \left[\mathcal{E} \left[e^{-s \sum_{i=1}^N R_i} | N \right] \right] \\
 &= \mathcal{E} \underbrace{[\mathcal{E}[e^{-sR}]^N]}_{= G_N[\hat{R}(s)]} \\
 &= G_N[\hat{R}(s)] \quad \text{but} \quad G_N(z) = e^{\alpha t(z-1)} \\
 &= e^{\alpha t(\hat{R}(s)-1)}
 \end{aligned}$$

$$\begin{aligned}\hat{\tau}(s) &= \mathcal{E}[e^{-sX} e^{\alpha X(\hat{R}(s)-1)}] = \mathcal{E}[e^{-X(s-\alpha(\hat{R}(s)-1))}] \\ &= \hat{X}(s - \alpha(\hat{R}(s) - 1))\end{aligned}$$

But $\hat{R} = \frac{\beta}{s+\beta}$ and $\hat{X}(s) = \frac{\mu}{s+\mu}$

$$\begin{aligned}\therefore \hat{\tau}(s) &= \frac{\mu}{s - \alpha(\hat{X}(s) - 1) + \mu} = \frac{\mu}{s - \alpha \frac{-s}{s+\beta} + \mu} \\ &= \frac{\mu(s+\beta)}{(s+\mu)(s+\beta) + \alpha s} \quad \text{as required.}\end{aligned}$$

b)

$$\begin{aligned}\hat{W}(s) &= \frac{(1-\rho)s}{s - \lambda + \lambda \hat{\tau}(s)} = \frac{(1-\rho)s[(s+\mu)(s+\beta) + \alpha s]}{(s-\lambda)[(s+\mu)(s+\beta) + \alpha s] + \lambda \mu(s+\beta)} \\ &= (1-\rho)s[s^2 + (\alpha + \beta + \mu)s + \mu\beta] \\ &= \frac{(1-\rho)(s^2 + (\alpha + \beta + \mu)s + \mu\beta)}{s^2 + (\alpha + \beta + \mu - \lambda)s - (\alpha + \beta + \mu - \mu\beta - \lambda\mu)} \\ &= (1-\rho) \left[1 + \frac{\lambda s + (\alpha + \beta + \mu - \lambda\mu)}{s^2 + (\alpha + \beta + \mu - \lambda)s - (\alpha + \beta + \mu - \mu\beta - \lambda\mu)} \right] \\ &\quad \text{where } \lambda_1 \text{ and } \lambda_2 \text{ are roots of denominator} \\ &= (1-\rho) \left[1 + \frac{A}{s + \lambda_1} + \frac{B}{s + \lambda_2} \right] \\ f_W(t) &= (1-\rho)\delta(t) + (Ae^{-\lambda_1 t} + Be^{-\lambda_2 t})\mu(t)\end{aligned}$$

where A, B are obtained from a partial fraction expansion

$$\begin{aligned}\hat{T}(s) &= \frac{(1-\rho)s\hat{\tau}(s)}{s - \lambda + \lambda \hat{\tau}(s)} = \frac{(1-\rho)s\mu(s+\beta)}{(s-\lambda)[(s+\mu)(s+\beta) + \alpha s] + \lambda \mu(s+\beta)} \\ &= \frac{(1-\rho)\mu(s+\beta)}{s^2 + (\alpha + \beta + \mu - \lambda)s - (\alpha + \beta + \mu - \mu\beta - \lambda\mu)} \\ &= \frac{A'}{s + \lambda_1} + \frac{B'}{s + \lambda_2} \\ f_T(t) &= (A'e^{-\lambda_1 t} + B'e^{-\lambda_2 t})\mu(t)\end{aligned}$$

12.56

$$\text{9.47 a) } N_j = N_{j-1} - U(N_{j-1}) + M_j = \begin{cases} N_{j-1} - 1 + M_j & N_{j-1} \geq 1 \\ M_j & N_{j-1} = 0 \end{cases} \quad (9.110\text{a})$$

✓

b)

$$\begin{aligned} \mathcal{E}[N_j] &= \mathcal{E}[N_j] - \mathcal{E}[U(N_{j-1})] + \mathcal{E}[M_j] \\ \Rightarrow \mathcal{E}[M_j] &= \mathcal{E}[U(N_{j-1})] = P[N_{j-1} > 0] \\ \Rightarrow P[N > 0] &= \mathcal{E}[M] = \lambda \mathcal{E}[\tau] \end{aligned}$$

c)

$$\begin{aligned} N_j^2 &= N_{j-1}^2 - 2N_{j-1}U(N_{j-1}) + U(N_{j-1})^2 + M_j^2 \\ &\quad + 2(N_{j-1} - U(N_{j-1}))M_j \end{aligned}$$

$$N_{j-1}U(N_{j-1}) = N_{j-1} \text{ and } U(N_{j-1})^2 = U(N_{j-1})$$

$$\begin{aligned} 0 &= -2\mathcal{E}[N_{j-1}] + \mathcal{E}[U(N_{j-1})] + \mathcal{E}[M_j^2] \\ &\quad + 2\mathcal{E}[N_{j-1}]\mathcal{E}[M_j] - \mathcal{E}[U(N_{j-1})]\mathcal{E}[M_j] \end{aligned}$$

$$\begin{aligned} \mathcal{E}[N_{j-1}] &= \frac{\mathcal{E}[U(N_{j-1})](1 - \mathcal{E}[M_j]) + \mathcal{E}[M_j^2]}{2(1 - \mathcal{E}[M_j])} \\ &= \mathcal{E}[U(N_{j-1})] + \frac{\mathcal{E}[M_j^2]}{2(1 - \mathcal{E}[M_j])} \end{aligned}$$

$$\text{From part b) } \mathcal{E}(U[N_j]) = \mathcal{E}[M] = \lambda \mathcal{E}[\tau]$$

$$\begin{aligned} \mathcal{E}[M^2] &= \mathcal{E}[\mathcal{E}[M^2|\tau]] = \mathcal{E}[X\tau + \lambda^2\tau^2] \\ &= \lambda \mathcal{E}[\tau] + \lambda^2 \mathcal{E}[\tau^2] \end{aligned}$$

Finally

$$\begin{aligned} \mathcal{E}[N] &= \lambda \mathcal{E}[\tau] + \frac{\lambda \mathcal{E}[\tau] + \lambda^2 \mathcal{E}[\tau^2]}{2(1 - \lambda \mathcal{E}[\tau])} \quad C_\tau^2 = \frac{\lambda^2 \mathcal{E}[\tau^2]}{\lambda^2 \mathcal{E}[\tau]^2} \\ &= \lambda \mathcal{E}[\tau] + \lambda \mathcal{E}[\tau] \frac{(1 + C_\tau^2)}{2(1 - \lambda \mathcal{E}[\tau])} \\ &= \lambda \mathcal{E}[T] \leftarrow \text{as given by 9.94} \quad \checkmark \end{aligned}$$

12.57

9.48 a)

$$\begin{aligned} G_N(z) &= \frac{(1-\rho)(z-1)\hat{\tau}(\lambda(1-z))}{z-\hat{\tau}(\lambda(1-z))} \quad \hat{\tau}(s) = e^{-sd} \\ &= \frac{(1-\rho)(z-1)e^{-\lambda d(t-z)}}{z-e^{-\lambda d(1-z)}} = \frac{(1-\rho)(1-z)}{1-ze^{\rho(1-z)}} \quad \text{where } \rho = \lambda d \end{aligned}$$

b)

$$\begin{aligned} \frac{1}{1-e^{\rho(1-z)}z} &= \sum_{k=0}^{\infty} e^{\rho(1-z)k} z^k = \sum_{k=0}^{\infty} e^{-\rho z k} e^{\rho k} z^k \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-\rho k z)^l}{l!} e^{\rho k} z^k \\ \therefore G_N(z) &= (1-\rho)(1-z) \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-\rho k)^l}{l!} e^{\rho k} z^{l+k} \\ &= (1-\rho) \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-\rho k)^l}{l!} e^{\rho k} (z^{l+k} - z^{l+k+1}) \\ &= \sum_{k'=0}^{\infty} P[N=k'] z^{k'} \end{aligned}$$

where

$$\begin{aligned} P[N=k'] &= (1-\rho) \left\{ \sum_{l,k:l+k=k'} \frac{(-\rho k)^l e^{\rho k}}{l!} - \sum_{l,k:l+k+1=k'} \frac{(-\rho k)^l e^{\rho k}}{l!} \right\} \\ P[N=k'] &= (1-\rho) \sum_{j=0}^{k'} \left[\frac{(-j\rho)^{k'-j} e^{j\rho}}{(k'-j)!} - \frac{(-j\rho)^{k'-j-1} e^{j\rho}}{(k'-j-1)!} \right] \\ &= (1-\rho) \sum_{j=0}^{k'} \frac{(-j\rho)^{k'-j-1} (-j\rho - k' + j)}{(k'-j)!} e^{j\rho} \quad \checkmark \end{aligned}$$

12.58

S. 45

$$\begin{aligned}\hat{W}(s) &= \frac{(1-\rho)s}{s - \lambda + \lambda\hat{\tau}(s)} = \frac{1-\rho}{1 - \lambda\frac{1-\hat{\tau}(s)}{s}} = \frac{1-\rho}{1 - \rho\frac{1-\hat{\tau}(s)}{s\mathcal{E}[\tau]}} \\ &= \frac{1-\rho}{1 - \rho\hat{R}(s)} \quad \text{where } \hat{R}(s) = \frac{1-\hat{\tau}(s)}{s\mathcal{E}[\tau]}\end{aligned}$$

But

$$f_R(t) = \mathcal{L}^{-1} \left[\frac{1-\hat{\tau}(s)}{s\mathcal{E}[\tau]} \right] = \frac{1}{\mathcal{E}[\tau]} [1 - F_\tau(x)]$$

which is pdf of residual service time as given by Eqn.12.87

$$\begin{aligned}\therefore \hat{W}(s) &= (1-\rho) \sum_{k=0}^{\infty} (\rho\hat{R}(s))^k \\ &= \sum_{k=0}^{\infty} (1-\rho)\rho^k \hat{R}^k(s)\end{aligned}$$

and

$$f_W(t) = \sum_{k=0}^{\infty} (1-\rho)\rho^k f^{(k)}(x)$$

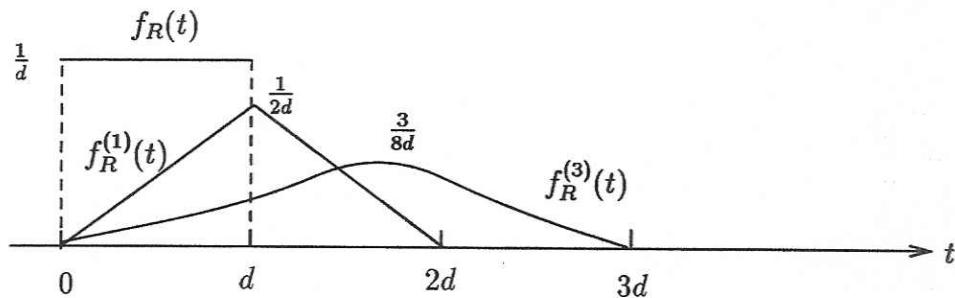
where

$$f^{(k)}(x) = \mathcal{L}^{-1}[\hat{R}^k(s)]$$

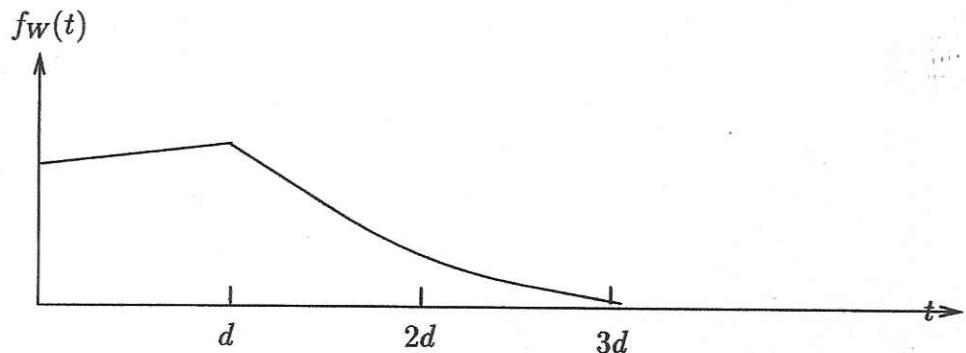
12.59

9.50 For M/D/1 $\hat{r}(s) = e^{-sd}$ and

$$\begin{aligned}\hat{R}(s) &= \frac{1 - e^{-sd}}{sd} \\ \Rightarrow f_R(t) &= \frac{1 - \mu(t - d)}{d} = \begin{cases} 1 & 0 \leq t \leq d \\ 0 & \text{else} \end{cases}\end{aligned}$$



$$\begin{aligned}\therefore f_W(t) &\approx (1 - \rho)f_R(t) + (1 - \rho)\rho f_R^{(1)}(t) + (1 - \rho)\rho^2 f_R^{(2)}(t) \\ &= \frac{1}{2}f_R(t) + \frac{1}{4}f_R^{(1)}(t) + \frac{1}{8}f_R^{(2)}(t)\end{aligned}$$



12.8 Burke's Theorem: Departures From M/M/c Systems

12.60

9.51 a) If a departure leaves the system nonempty, then another customer commences service immediately. Thus the time until the next departure is an exponential random variable with mean $1/\mu$.

b) If a departure leaves the system empty, then the time until the next departure is equal to the sum of an exponential interarrival time (of mean $1/\lambda$) followed by an exponential service time (of mean $1/\mu$).

c) The Laplace transform of the interdeparture time is

$$\frac{\mu}{s + \mu} \quad \text{when a departure leaves system nonempty}$$

$$\frac{\lambda}{s + \lambda} \frac{\mu}{s + \mu} \quad \text{when a departure leaves system empty}$$

$$\begin{aligned} \therefore \mathcal{E}[e^{-sT_d}] &= \frac{\mu}{s + \mu} \underbrace{\rho}_{\substack{\text{prob. system} \\ \text{left nonempty}}} + \frac{\lambda}{s + \lambda} \frac{\mu}{s + \mu} \underbrace{(1 - \rho)}_{\substack{\text{prob. system} \\ \text{left empty}}} \\ &= \frac{\lambda}{s + \mu} + \frac{\lambda(\mu - \lambda)}{(s + \lambda)(s + \mu)} = \frac{\lambda(s + \lambda) + \lambda\mu - \lambda^2}{(s + \lambda)(s + \mu)} \\ &= \frac{\lambda}{s + \lambda} \Rightarrow T_d \text{ exponential with mean } 1/\lambda \end{aligned}$$

12.61

9.52 Claim:

$$P[N_1 = n, N_2 = m] = (1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^m \quad n, m \geq 0$$

$$\rho_i = \lambda/\mu_i$$

Eq. 9.135a

$$\begin{aligned} \lambda P[N_1 = 0, N_2 = 0] &= \lambda(1 - \rho_1)(1 - \rho_2) \\ &= \mu_2(1 - \rho_1)(1 - \rho_2)\rho_2 = \mu_2 P[N_1 = 0, N_2 = 1] \quad \checkmark \end{aligned}$$

Eq. 9.135b

$$\begin{aligned} (\lambda + \mu_1)P[N_1 = 0, N_2 = 0] &= (\lambda + \mu_1)(1 - \rho_1)\rho_1^n(1 - \rho_2) \\ &= \lambda(1 - \rho_1)\rho_1^n(1 - \rho_2) + \mu_1(1 - \rho_1)\rho_1^n(1 - \rho_2) \\ &= \mu_2\rho_2(1 - \rho_1)\rho_1^n(1 - \rho_2) + \lambda(1 - \rho_1)\rho_1^{n-1}(1 - \rho_2) \\ &= \mu_2 P[N_1 = n, N_2 = 1] + \lambda P[N_1 = n - 1, N_2 = 0] \quad \checkmark \end{aligned}$$

Eqn. 9.135c

$$\begin{aligned} (\lambda + \mu_2)P[N_1 = 0, N_2 = m] &= (\lambda + \mu_2)(1 - \rho_1)(1 - \rho_2)\rho_2^m \\ &= \mu_2(1 - \rho_1)(1 - \rho_2)\rho_2^{m+1} + \mu_1(1 - \rho_1)\rho_1(1 - \rho_2)\rho_2^{m-1} \\ &= \mu_2 P[N_1 = 0, N_2 = m + 1] + \mu_1 P[N_1 = 1, N_2 = m - 1] \quad \checkmark \end{aligned}$$

Eqn. 9.135d

$$\begin{aligned} (\lambda + \mu_1 + \mu_2)P[N_1 = n, N_2 = m] &= \lambda(1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^m + \mu_1(1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^m \\ &\quad + \mu_2(1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^m \\ &= \mu_2(1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^{m+1} + \lambda(1 - \rho_1)\rho_1^{n-1}(1 - \rho_2)\rho_2^m \\ &\quad + \mu_1(1 - \rho_1)\rho_1^{n+1}(1 - \rho_2)\rho_2^{m-1} \\ &= \mu_2 P[N_1 = n, N_2 = m + 1] + \lambda P[N_1 = n - 1, N_2 = m] \\ &\quad + \mu_1 P[N_1 = n + 1, N_2 = m - 1] \quad \checkmark \end{aligned}$$

12.62

9.53 The arrival process at queue #3 is the merge of two independent Poisson processes with combined rate

$$\lambda_1 + \frac{1}{2}\lambda_2$$

The state of queue #3 at time t is independent of those at queues #1 and #2 at time t :

$$P[N_1(t) = i, N_2(t) = j, N_3(t) = k] = (1 - \rho_1)\rho_1^i(1 - \rho_2)\rho_2^j(1 - \rho_3)\rho_3^k$$

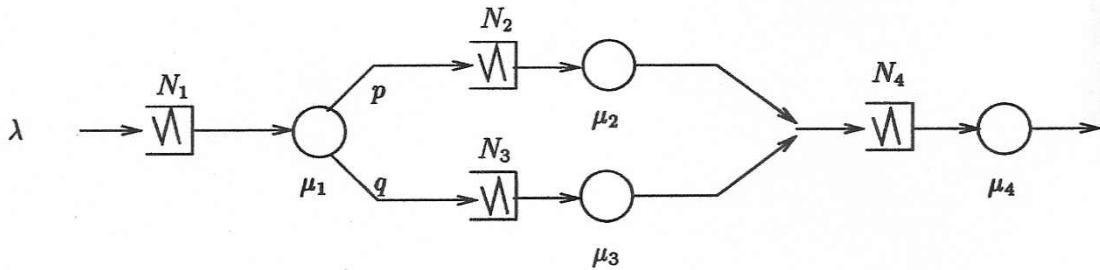
where

$$\rho_1 = \frac{\lambda_1}{\mu_1} < 1$$

$$\rho_2 = \frac{\lambda_2}{\mu_2} < 1$$

$$\text{and } \rho_3 = \frac{(\lambda_1 + \frac{1}{2}\lambda_2)}{\mu_3} < 1$$

12.63



Claim:

$$\begin{aligned} P[N_1 = k, N_2 = l, N_3 = m, N_4 = n] &= (1 - \rho)\rho_1^k(1 - \rho_2)\rho_2^l(1 - \rho_3)\rho_3^m(1 - \rho_4)\rho_4^n \\ &\triangleq A\rho_1^k\rho_2^l\rho_3^m\rho_4^n \end{aligned}$$

where

$$\rho_1 = \frac{\lambda}{\mu_1} \quad \rho_2 = \frac{p\lambda}{\mu_2} \quad \rho_3 = \frac{q\lambda}{\mu_3} \quad \rho_4 = \frac{\lambda}{\mu_4}$$

Let \underline{e}_i be the i th unit vector.Let $\underline{s} = (klmn)$ where $k, l, m, n > 0$. The balance equation for this state is

$$\begin{aligned} (\lambda + \mu_1 + \mu_2 + \mu_3 + \mu_4)P[\underline{s}] &= \lambda P[\underline{s} - \underline{e}_1] + p\mu_1 P[\underline{s} + \underline{e}_1 - \underline{e}_2] \\ &\quad + q\mu_1 P[\underline{s} + \underline{e}_1 - \underline{e}_2] + \mu_2 P[\underline{s} + \underline{e}_2 - \underline{e}_4] \\ &\quad + \mu_3 P[\underline{s} + \underline{e}_3 - \underline{e}_4] + \mu_4 P[\underline{s} + \underline{e}_4] \\ &= \lambda P[\underline{s}]\rho_1^{-1} + p\mu_1 P[\underline{s}]\rho_1\rho_2^{-1} + q\mu_1 P[\underline{s}]\rho_1\rho_3^{-1} \\ &\quad + \mu_2 P[\underline{s}]\rho_2\rho_4^{-1} + \mu_3 P[\underline{s}]\rho_3\rho_4^{-1} + \mu_4 P[\underline{s}]\rho_4 \\ &= \mu_1 P[\underline{s}] + \mu_2 P[\underline{s}] + \mu_3 P[\underline{s}] + \mu_4 P[\underline{s}] \\ &\quad + \lambda P[\underline{s}] \quad \checkmark \end{aligned}$$

 $\therefore P[\underline{s}] = A\rho_1^k\rho_2^l\rho_3^m\rho_4^n$ satisfies this balance equation.There are 15 other special cases of boundary balance equations. These are shown to be satisfied by $P[\underline{s}]$ in similar fashion.

12.9 Networks of Queues: Jackson's Theorem

12.64

$$9.55 \quad \lambda_1 = \lambda \quad \lambda_2 = \frac{1}{2}\lambda + \frac{1}{2}\lambda_3 \quad \lambda_3 = \lambda_2 + \frac{1}{2}\lambda$$

$$\Rightarrow \lambda_1 = \lambda \quad \lambda_2 = \frac{3}{2}\lambda \quad \lambda_3 = 2\lambda$$

$$\Rightarrow \rho_1 = \frac{\lambda}{\mu_1} \quad \rho_2 = \frac{3\lambda}{2\mu_2} \quad \rho_3 = \frac{2\lambda}{\mu_3}$$

Assuming $\rho_i < 1$, $i = 1, 2, 3$, then

$$P[N_1 = k, N_2 = l, N_3 = m] = (1 - \rho_1)\rho_1^k(1 - \rho_2)\rho_2^l(1 - \rho_3)\rho_3^m \quad k, l, m \geq 0$$

12.65

~~9.56~~ $I = 3$

$$\left. \begin{array}{rcl} \pi_0 & = & p\pi_0 + \pi_1 + \pi_2 \\ \pi_1 & = & \frac{1}{2}(1-p)\pi_0 \\ \pi_2 & = & \frac{1}{2}(1-p)\pi_0 \end{array} \right\} \quad \begin{array}{l} \pi_0 = \frac{1}{2-p} \\ \pi_1 = \pi_2 = \frac{1-p}{2(2-p)} \end{array}$$

a) Then

$$\begin{aligned} \lambda_0 &= \lambda(3)\pi_0 = \frac{\lambda(3)}{2-p} & \rho_0 &= \frac{\lambda_0}{\mu} \\ \lambda_1 &= \lambda(3)\pi_1 = \frac{\lambda(3)(1-p)}{2(2-p)} & \rho_1 &= \frac{\lambda_1}{\mu_1} \\ \lambda_2 &= \lambda_1 & \rho_2 &= \frac{\lambda_1}{\mu_2} \end{aligned}$$

$$\begin{aligned} S(3) &= (1 - \rho_0)(1 - \rho_1)(1 - \rho_2)[\rho_0^3 + \rho_1^3 + \rho_2^3 + \rho_0\rho_1^2 + \rho_0\rho_2^2 \\ &\quad + \rho_1\rho_2^2 + \rho_1\rho_0^2 + \rho_2\rho_0^2 + \rho_2\rho_1^2 + \rho_0\rho_1\rho_2] \\ &= (1 - \rho_0)(1 - \rho_1)(1 - \rho_2)[(\rho_0^2 + \rho_1^2 + \rho_2^2)(\rho_0 + \rho_1 + \rho_2) + \rho_0\rho_1\rho_2] \end{aligned}$$

$$\therefore P[N_0 = i, N_1 = j, N_2 = 3 - i - j] = \frac{\rho_0^i \rho_1^j \rho_2^{3-i-j}}{(\rho_0^2 + \rho_1^2 + \rho_2^2)(\rho_0 + \rho_1 + \rho_2) + \rho_0\rho_1\rho_2} \\ 0 \leq i, j \quad \text{and} \quad i + j \leq 3$$

b) The program completion rate is

$$p\mu[1 - P[N_0 = 0]] = p\mu \frac{\rho_0^3 + \rho_0^2\rho_1 + \rho_0^2\rho_2 + \rho_0\rho_2^2 + \rho_0\rho_1^2 + \rho_0\rho_1\rho_2}{(\rho_0^2 + \rho_1^2 + \rho_2^2)(\rho_0 + \rho_1 + \rho_2) + \rho_0\rho_1\rho_2}$$

12.10 Simulation and Data Analysis of Queueing Systems

12.66

~~9.5~~ From Eq. (12.56) we have:

$$\pi_0 = \frac{1}{2-p} \quad \pi_1 = \pi_2 = \frac{1-p}{2(2-p)}$$

We need to find $p\lambda_0(3) = p\pi_0\lambda(3)$

$I = 1$

$$\mathcal{E}[T_0(1)] = \frac{1}{\mu} \quad \mathcal{E}[T_1(1)] = \frac{1}{\mu_1} \quad \mathcal{E}[T_2(1)] = \frac{1}{\mu_2}$$

$$\begin{aligned}\lambda(1) &= 1 \left[\frac{\frac{1}{\mu}}{2-p} + \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) \frac{1-p}{2(2-p)} \right]^{-1} \\ \mathcal{E}[N_0(1)] &= \frac{\frac{1}{\mu}}{\frac{1}{\mu} + \frac{1-p}{2} \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)} \triangleq \frac{a}{a+b+c} \\ \mathcal{E}[N_1(1)] &= \frac{\frac{1-p}{2} \frac{1}{\mu_1}}{\frac{1}{\mu} + \frac{1-p}{2} \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)} \triangleq \frac{b}{a+b+c} \\ \mathcal{E}[N_2(1)] &= \frac{\frac{1-p}{2} \frac{1}{\mu_2}}{\frac{1}{\mu} + \frac{1-p}{2} \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)} \triangleq \frac{c}{a+b+c}\end{aligned}$$

where

$$a \triangleq \frac{1}{\mu} \quad b = \frac{1-p}{2} \frac{1}{\mu_1} \quad c = \frac{1-p}{2} \frac{1}{\mu_2}$$

$I = 2$

$$\begin{aligned}\mathcal{E}[T_0(2)] &= \frac{1}{\mu} \left[\frac{2a+b+c}{a+b+c} \right] \\ \mathcal{E}[T_1(2)] &= \frac{1}{\mu_1} \left[\frac{a+2b+c}{a+b+c} \right] \\ \mathcal{E}[T_2(2)] &= \frac{1}{\mu_2} \left[\frac{a+b+2c}{a+b+c} \right]\end{aligned}$$

$$\lambda(2) = 2 \left[\frac{1}{\frac{1}{2-p} \mathcal{E}[T_0(2)] + \frac{1-p}{2} \frac{1}{2-p} \mathcal{E}[T_1(2)] + \frac{1-p}{2} \frac{1}{2-p} \mathcal{E}[T_2(2)]} \right]$$

$$\begin{aligned}
 &= 2 \left[\underbrace{\frac{1}{\mu} (2a + b + c)}_a + \underbrace{\frac{1-p}{2} \frac{1}{\mu_1} (a + 2b + c)}_b + \underbrace{\frac{1-p}{2} \frac{1}{\mu_2} (a + b + 2c)}_c \right] \\
 &= 2 \frac{(2-p)(a+b+c)}{2a^2 + 2b^2 + 2c^2 + 2ab + 2ac + 2bc} \\
 &= \frac{(2-p)(a+b+c)}{a^2 + b^2 + c^2 + ab + ac + bc}
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{E}[N_0(2)] &= \lambda(2)\pi_0 \mathcal{E}[T_0(2)] = \lambda(2) \frac{1}{2-p} \frac{1}{\mu} \left[\frac{2a+b+c}{a+b+c} \right] \\
 &= \frac{1}{\mu} \frac{2a+b+c}{a^2 + b^2 + c^2 + ab + ac + bc} = \frac{2a^2 + ab + ac}{a^2 + b^2 + c^2 + ab + ac + bc} \\
 \mathcal{E}[N_1(2)] &= \frac{\overbrace{\frac{1}{\mu_1} \frac{1-p}{2}}^b (a+2b+c)}{a^2 + b^2 + c^2 + ab + ac + bc} = \frac{ab + 2b^2 + bc}{a^2 + b^2 + c^2 + ab + ac + bc} \\
 \mathcal{E}[N_2(2)] &= \frac{c(a+b+2c)}{a^2 + b^2 + c^2 + ab + ac + bc} = \frac{ac + bc + 2c^2}{a^2 + b^2 + c^2 + ab + ac + bc}
 \end{aligned}$$

$I = 3$

$$\begin{aligned}
 \mathcal{E}[T_0(3)] &= \frac{1}{\mu} [1 + \mathcal{E}[N_0(2)]] = \frac{3a^2 + b^2 + c^2 + 2ab + 2ac + bc}{a^2 + b^2 + c^2 + ab + ac + bc} \\
 \mathcal{E}[T_1(3)] &= \frac{1}{\mu_1} \left[\frac{2ab + 3b^2 + 2bc + a^2 + c^2 + ac}{a^2 + b^2 + c^2 + ab + ac + bc} \right] \\
 \mathcal{E}[T_2(3)] &= \frac{1}{\mu_2} \left[\frac{2ac + 2bc + 3c^2 + a^2 + b^2 + c^2 + ab}{a^2 + b^2 + c^2 + ab + ac + bc} \right] \\
 \lambda(3) &= 3 \frac{1}{\frac{1}{2-p} \mathcal{E}[T_0(3)] + \frac{1-p}{2} \frac{1}{2-p} \mathcal{E}[T_1(3)] + \frac{1-p}{2} \frac{1}{2-p} \mathcal{E}[T_2(3)]}
 \end{aligned}$$

Program completion rate is

$$\begin{aligned}
 p\lambda_0 &= px_0\lambda(3) \\
 &= \frac{p}{2-p}\lambda(3)
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{3p}{\mathcal{E}[T_0(3)] + \frac{1-p}{2}\mathcal{E}[T_1(3)] + \frac{1-p}{2}\mathcal{E}[T_2(3)]} \\
 &= [3p(a^2 + b^2 + c^2 + ab + ac + bc)]/[a(3a^2 + b^2 + c^2 + 2ab + 2ac + bc) \\
 &\quad + b(2ab + 3b^2 + 2bc + a^2 + c^2 + ac) + c(2ac + 2bc + 3c^2 + a^2 + b^2 + ab)] \\
 &= \frac{p\mu \overbrace{\left(\frac{1}{\mu}\right)}^a (a^2 + b^2 + c^2 + ab + ac + bc)}{a^3 + b^3 + c^3 + ab^2 + ac^2 + a^2b + a^2c + b^2c + bc^2 + abc}
 \end{aligned}$$

We will multiply the numerator and denominator above by $\left(\frac{\lambda(3)}{2-p}\right)^3$ but first note that

$$\begin{aligned}
 \frac{\lambda(3)a}{2-p} &= \frac{\lambda(3)}{\mu(2-p)} = \lambda_0(3)\frac{1}{\mu} = \rho_0 \\
 \frac{\lambda(3)b}{2-p} &= \frac{\lambda(3)(1-p)}{2(2-p)\mu_1} = \rho_1 \quad \frac{\lambda(3)c}{2-p} = \rho_2
 \end{aligned}$$

\therefore Completion Rate

$$= p\mu \frac{\rho_0^3 + \rho_0^2\rho_1 + \rho_0^2\rho_2 + \rho_0\rho_1^2 + \rho_0\rho_1\rho_2}{(\rho_0^2 + \rho_1^2 + \rho_2^2)(\rho_0 + \rho_1 + \rho_2) + \rho_1\rho_2\rho_3} \quad \checkmark$$

12.67

```

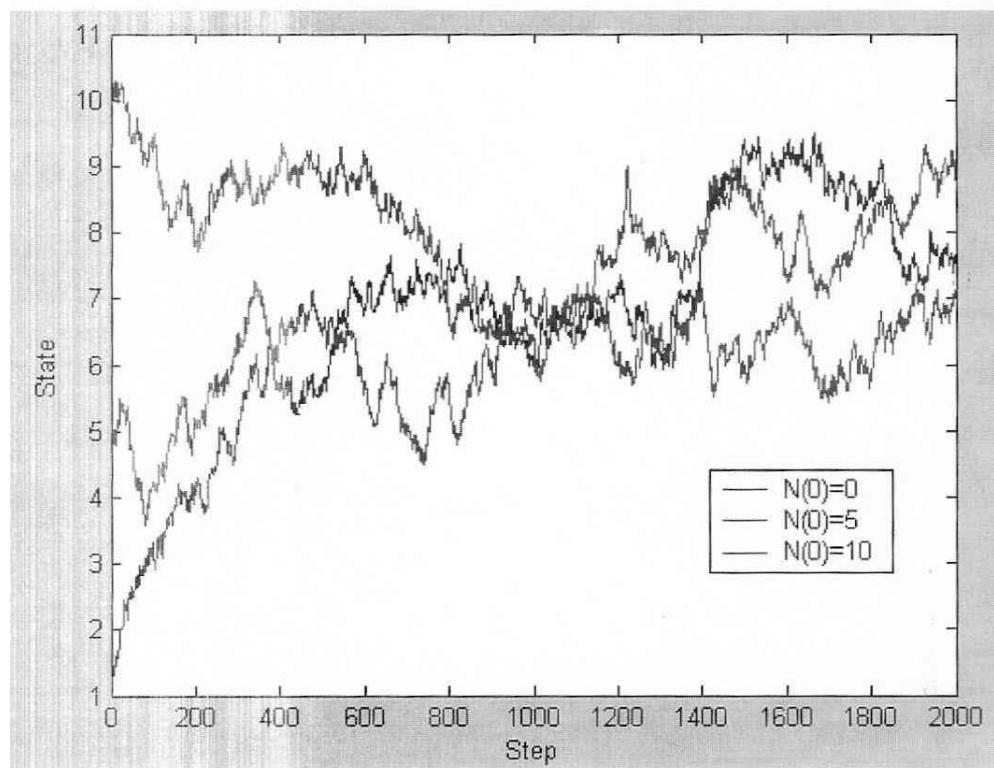
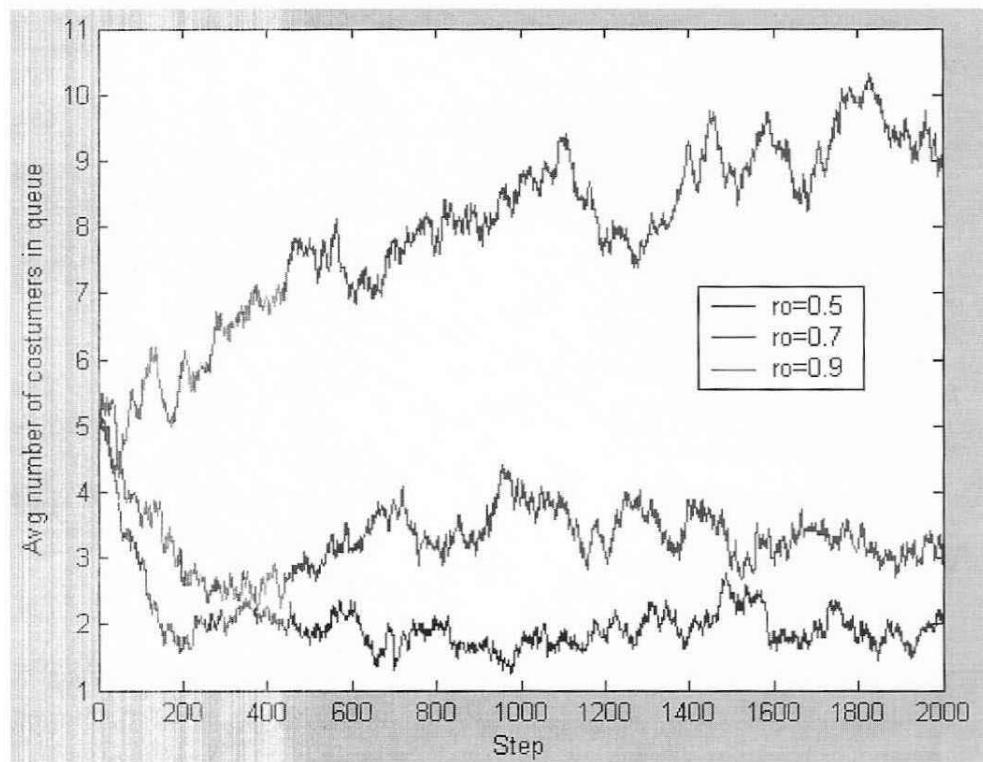
Nmax=50;
P=zeros(Nmax+1,3);
mu=1;
lambda=.9;
delta=.1;
a=delta*lambda;
b=delta*mu;
P(1,:)=[0,1-a,a];
r=[(1-a)*b,a*b+(1-a)*(1-b),(1-b)*a];
for n=2:Nmax;
    P(n,:)=r;
end
P(Nmax+1,:)=[(1-a)*b,1-(1-a)*b,0];
IC=zeros(Nmax+1,1);
IC(1,1)=1;
L=2000;
avg_seq=zeros(L,1);
avg_cor=zeros(L,1);
for j=1:25
    seq=queueState(Nmax,P,IC,L);
    cor_seq=autocorr(seq,L);
    for l=1:L
        avg_seq(l)=(avg_seq(l)*(j-1)+seq(l))/j;
        avg_cor(l)=(avg_cor(l)*(j-1)+cor_seq(l))/j;
    end
end
plot(avg_seq);

```

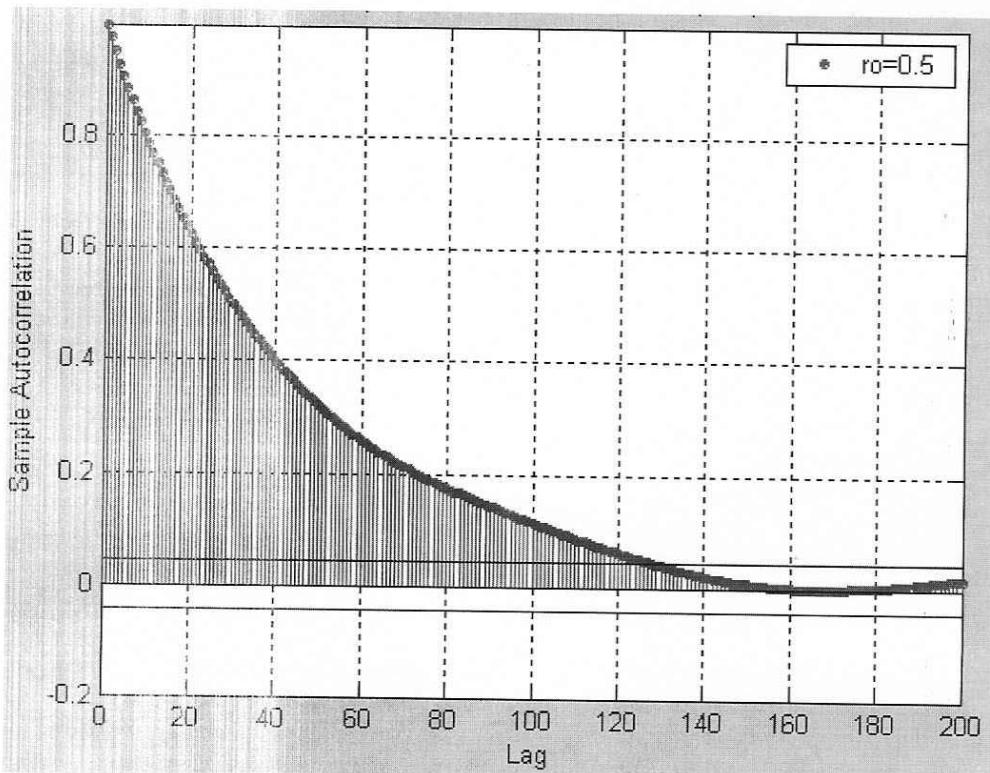
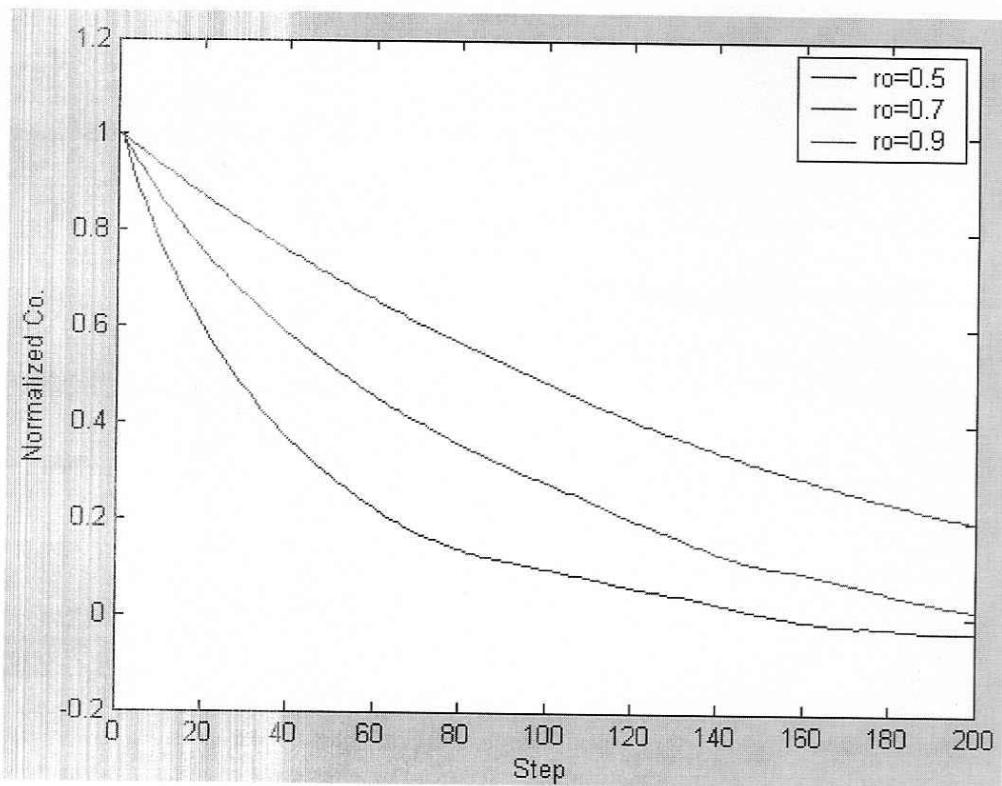
```

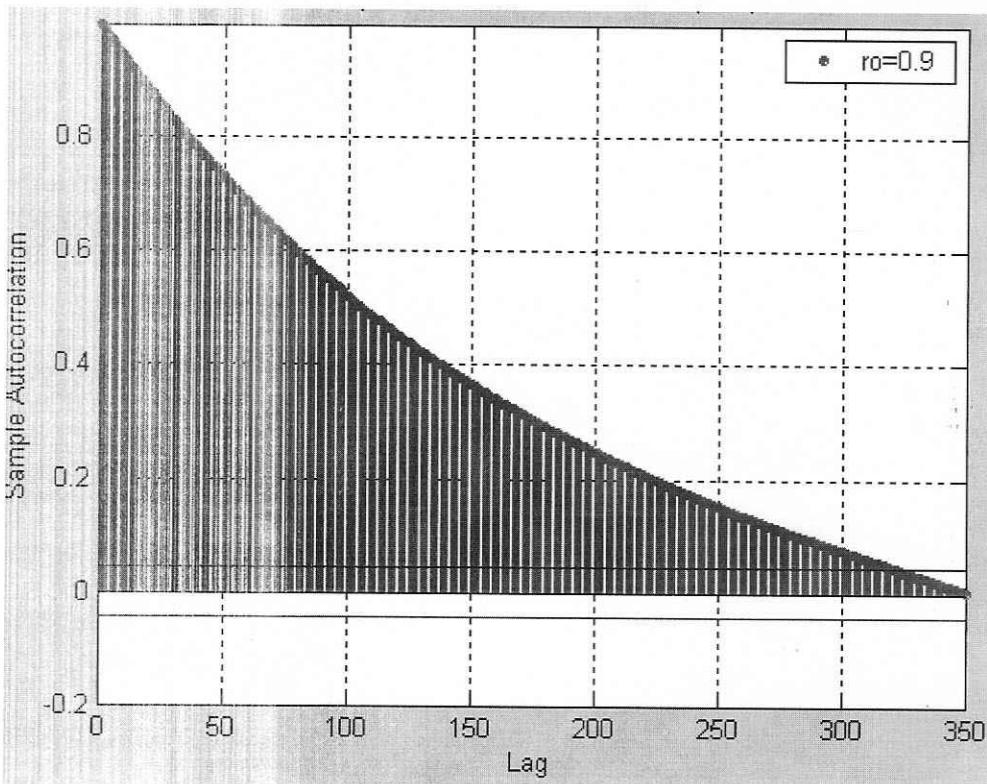
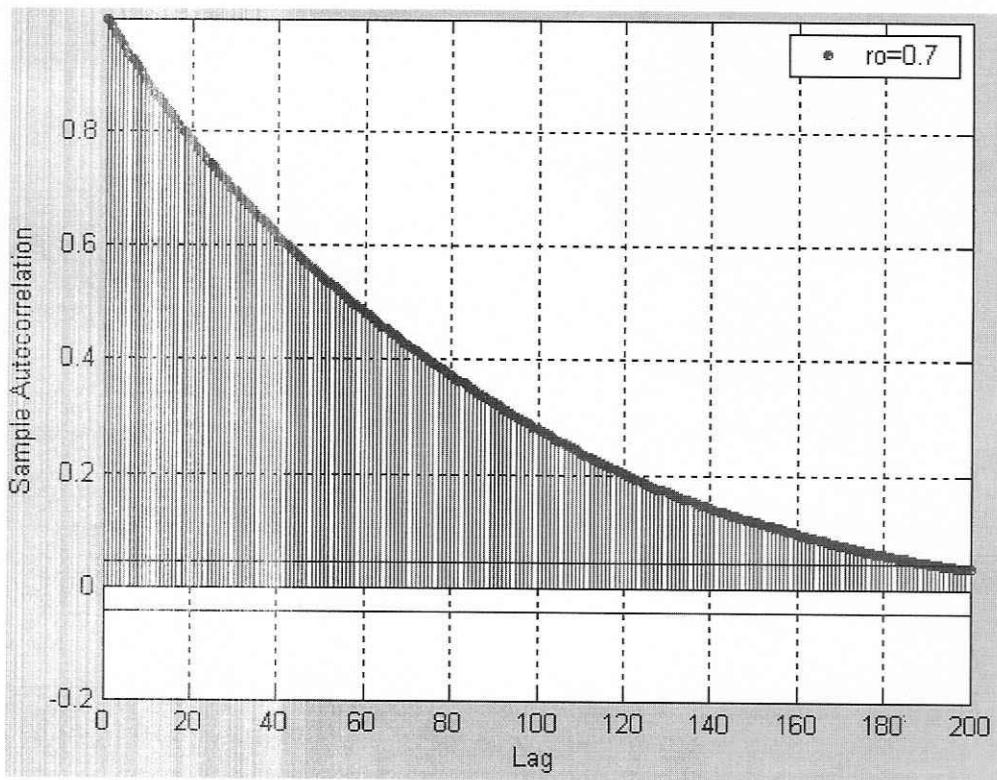
function stseq = queue_state(Nmax,P,IC,L)
stseq=zeros(1,L);
s=[1:Nmax+1];
step=[-1,0,1];
%Initst= floor(1000*rand);
Initst=ceil(10*rand);
stseq(1)=Initst;
for n=2:L+1
    k=rand;
    if(k<P(stseq(n-1),1))
        nxt=step(1);
    elseif (k<(P(stseq(n-1)+1,1)+P(stseq(n-1)+1,2)))
        nxt=step(2);
    else
        nxt=step(3);
    end
    nextst=stseq(n-1)+nxt;
    stseq(n)=nextst;
end

```



12.68





12.70

```

Nmax=50;
P=zeros(Nmax+1,3);
mu=1;
lambda=.5;
delta=.1;
a=delta*lambda;
b=delta*mu;
P(1,:)=[0,1-a,a];
r=[(1-a)*b,a*b+(1-a)*(1-b),(1-b)*a];
for n=2:Nmax;
    P(n,:)=r;
end
P(Nmax+1,:)=[(1-a)*b,1-(1-a)*b,0];
IC=zeros(Nmax+1,1);
IC(1,1)=1;
L=5000;
n=50;
avg_seq=zeros(L,1);
avg_cor=zeros(L,1);
avg=zeros(n,6);
conf=zeros(n,2);
avgseq=zeros(n,1);
prcnt=zeros(n,1);
z=1.68; %90% confidence interval

for j=1:n
    seq=queueState(Nmax,P,IC,L);
    avgseq(j)=sum(seq(1:L))/L;
    conf(j,1)=avgseq(j)-z*sqrt(var(seq))/sqrt(n);
    conf(j,2)=avgseq(j)+z*sqrt(var(seq))/sqrt(n);
    for l=1:L
        avg_seq(l)=(avg_seq(l)*(j-1)+seq(l))/j;
    end
    prent(j)=accuracy(avg_seq,conf(j,1), conf(j,2),L);
    for i=1:6
        ii=200*i+1;
        avg(j,i)=sum(seq(ii:ii+800))/800;
    end
end

avg1=zeros(6,1);
for i=1:6
    ii=200*i+1;
    avg1(i)=sum(avg_seq(ii:ii+800))/800;
end
plot(avg_seq);

function acc = accuracy(seq,a,b,L)
cnt=0;
for i=1:L
    a1=seq(i)-a;
    a2=seq(i)-b;
    if a1>=0 && a2<=0
        cnt=cnt+1;
    end
end
acc=cnt/L;

```

Each column shows mean for one batch is 50 simulations ($\rho=0.5$):

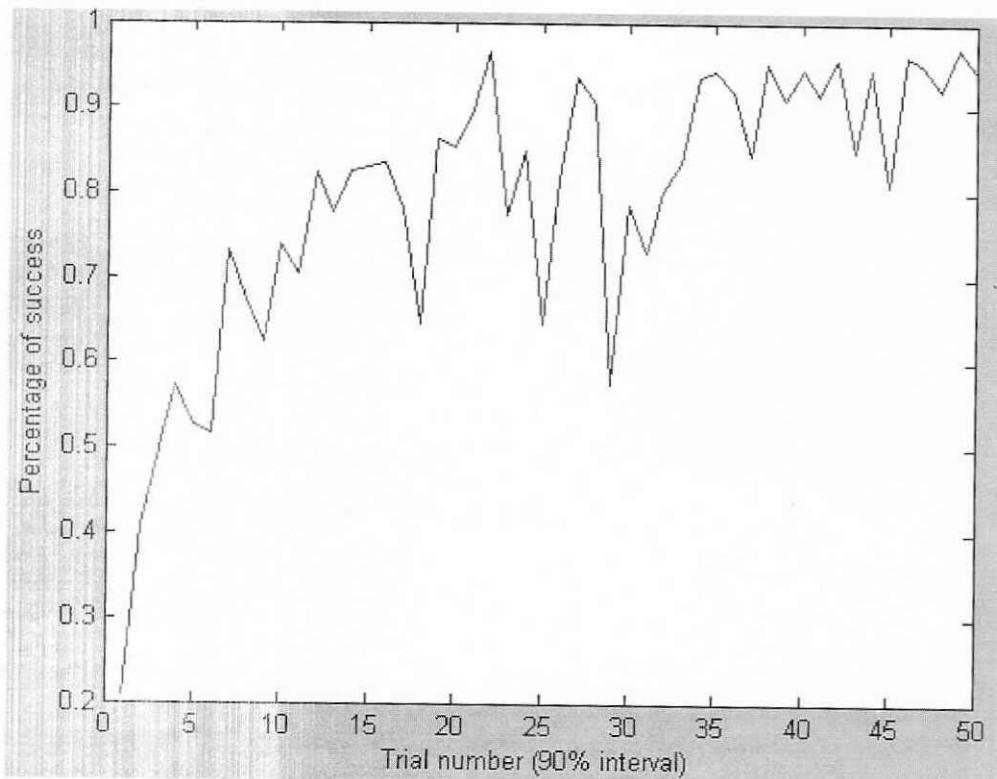
avg =

1.8275	1.7862	2.0550	2.1800	2.0513	1.8300
1.9150	1.8663	1.7900	1.2825	1.2788	2.1550
1.7038	1.6450	2.2388	2.3138	2.4337	2.2788
1.7938	1.8900	1.9163	1.6563	1.5562	1.7650
2.2212	2.3112	2.1363	1.5375	1.8975	2.0825
1.7763	1.8975	1.9175	1.8313	1.4763	1.4950
1.9288	1.9238	1.6275	1.5137	1.6863	1.7250
1.5900	1.7000	1.5337	1.5225	1.6000	1.5738
2.2425	2.3075	2.3438	1.5213	1.5362	1.5225
2.0013	1.9500	1.9788	1.5862	1.6450	1.7900
1.7637	1.5700	1.7200	1.9212	1.9225	1.8925
1.4688	1.6863	2.4663	2.3500	2.3863	2.2037
2.3537	2.3100	2.3737	2.3950	2.2212	2.0850
2.1050	2.2837	2.0313	2.2763	2.1250	1.9288
1.9412	1.8713	2.2588	2.1175	2.1913	1.8825
2.7375	2.8725	2.8363	3.2287	2.3912	3.0638
1.8162	1.5662	1.6438	1.6600	1.5738	2.1425
3.3963	3.3363	2.7062	1.7825	1.6888	1.4475
2.2500	2.1975	2.2763	1.6912	1.4800	1.4400
1.7388	1.9087	2.0275	1.9863	1.8713	2.2925
1.5900	1.6087	2.4137	2.3262	2.4550	2.3912
2.1500	1.7475	1.4238	1.4150	1.5313	1.6100
1.6612	1.6838	1.4775	1.4288	1.4663	1.4325
1.5488	1.6200	1.5800	1.6650	1.6950	1.8438
1.6587	1.6688	1.7175	1.7563	1.6187	1.5425
1.5425	1.5337	1.5900	1.9538	1.8375	2.0838
1.5350	1.7900	1.8587	2.8700	3.3163	3.2437
1.6175	1.5438	1.5413	1.6087	1.5388	2.0675
1.9925	2.0125	1.9813	2.0412	1.7962	1.7313
2.2825	1.8875	1.5700	1.4837	1.5788	1.5538
2.0187	1.8925	1.5950	1.3175	1.3987	1.3875
1.6538	1.6775	2.0675	1.8963	1.9688	1.9800
1.7588	1.8275	1.8150	1.6463	1.8663	1.7913
1.8125	1.7950	1.9737	1.8400	1.7825	2.0100
2.3287	2.4613	2.3712	1.7413	1.9125	1.6550
1.7038	1.5538	1.7950	1.8575	1.7962	1.9163
2.5675	2.4562	1.8725	1.9763	1.7237	1.7575
1.9038	1.7363	1.8075	1.7537	1.6638	1.4188
1.7163	1.4400	1.4000	1.4950	1.4775	1.6825
2.1450	2.2175	2.1338	2.2525	1.9913	2.0713
1.8538	1.9825	1.9475	1.8162	1.9525	1.8825
2.4000	1.8975	1.9512	1.7988	1.5950	1.5575
1.9188	2.1400	2.8988	2.4925	2.5750	2.2687
1.6663	1.8438	1.7962	1.7563	1.5037	1.4988
3.2550	1.8038	1.5300	2.0300	1.9675	2.4712
2.4362	2.6200	2.6675	1.8900	1.7875	2.2050
2.2050	1.9087	1.8550	1.6475	1.5263	1.6538
2.0888	2.3625	2.2363	2.1200	1.6575	1.5100
1.6400	1.8963	2.0950	2.0137	1.9150	1.5300
1.8188	1.8750	2.0425	2.4188	2.3963	2.4562

conf =

1.6179	2.1865
1.4681	2.0367
1.5845	2.1531
1.6285	2.1971
1.3779	1.9465
1.3155	1.8841
2.0663	2.6349
1.8899	2.4585
1.6341	2.2027
1.5071	2.0757
1.5911	2.1597
1.6489	2.2175
1.5311	2.0997
1.5765	2.1451
1.5279	2.0965
1.5345	2.1031
1.8395	2.4081
1.8585	2.4271
1.5859	2.1545
1.7831	2.3517
1.6665	2.2351
1.6427	2.2113
1.6305	2.1991
1.8207	2.3893
1.7083	2.2769
1.5471	2.1157
1.4711	2.0397
1.6845	2.2531
1.5713	2.1399
1.4545	2.0231
1.5913	2.1599
1.8241	2.3927
1.6721	2.2407
1.6101	2.1787
1.4299	1.9985
1.6529	2.2215
1.4947	2.0633
1.6019	2.1705
1.5661	2.1347
1.6317	2.2003
1.9669	2.5355
1.4301	1.9987
1.7863	2.3549
2.0861	2.6547
1.9275	2.4961
1.3875	1.9561
1.5473	2.1159
1.5055	2.0741
1.6113	2.1799
1.3331	1.9017

The percentage of times that the real value of the state is in the confidence interval is shown below:



12.71

```

Nmax=50;
P=zeros(Nmax+1,3);
mu=1;
lambda=2;
delta=.1;
c=3;
a=delta*lambda;
b=delta*mu*c;
b2=delta*mu^2;
P(1,:)=[0,1-a,a];
P(2,:)=[(1-a)*b2,a*b2+(1-a)*(1-b2),(1-b2)*a];
r=[(1-a)*b,a*b+(1-a)*(1-b),(1-b)*a];
for n=3:Nmax;
    P(n,:)=r;
end
P(Nmax+1,:)=[(1-a)*b,1-(1-a)*b,0];
IC=zeros(Nmax+1,1);
IC(1,1)=1;
L=5000;
n=50;
avg_seq=zeros(L,1);
avg_cor=zeros(L,1);
avg=zeros(n,6);
conf=zeros(n,2);
avgseq=zeros(n,1);
prcnt=zeros(n,1);
prq=zeros(n,1);
z=1.68; %90% confidence interval

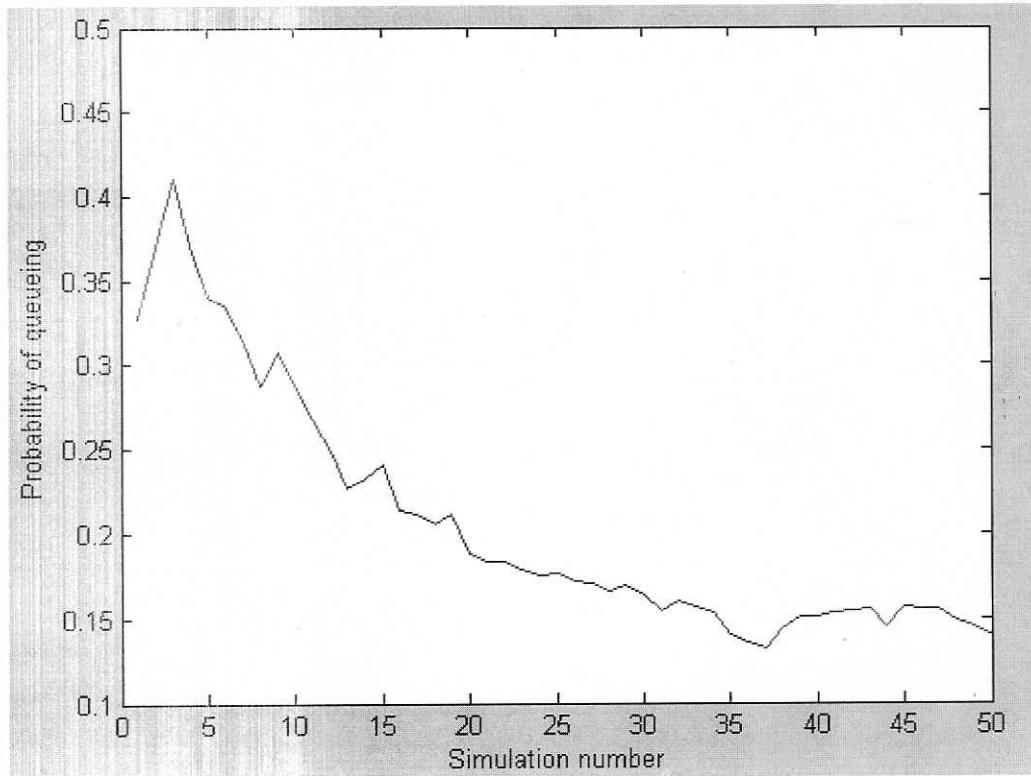
for j=1:n
    seq=queueState(Nmax,P,IC,L);
    avgseq(j)=sum(seq(1:L))/L;
    conf(j,1)=avgseq(j)-z*sqrt(var(seq))/sqrt(n);
    conf(j,2)=avgseq(j)+z*sqrt(var(seq))/sqrt(n);
    for l=1:L
        avg_seq(l)=(avg_seq(l)*(j-1)+seq(l))/j;
    end
    prcnt(j)=accuracy(avg_seq,conf(j,1), conf(j,2),L);
    prq(j)=prQue(avg_seq,c,L);
    for i=1:6
        ii=200*i+1;
        avg(j,i)=sum(seq(ii:ii+800))/800;
    end
end

```

```
avg1=zeros(6,1);
for i=1:6
    ii=200*i+1;
    avg1(i)=sum(avg_seq(ii:ii+800))/800;
end
plot(seq);
```

```
function acc = prQue(seq,c,L)
cnt=0;
for i=1:L
    if seq(i)>c
        cnt=cnt+1;
    end
end
acc=cnt/L;
```

Probability of being queued before getting the service:



12.72

```

Nmax=50;
mu=1;
lambda=.9;
G=zeros(Nmax,2);
cnt=zeros(Nmax,1);
for i=1:Nmax
    G(i,:)=[mu,lambda];
end
G(1,1)=0;
%G(Nmax,2)=0;
IC=zeros(Nmax+1,1);
IC(1,1)=1;
T=100;
[stseq,OccTime,n]= contTm(Nmax,G,IC,T);
%pmf
l=length(stseq);
for i=1:l
    idx=stseq(i);
    cnt(idx)=cnt(idx)+1;
end

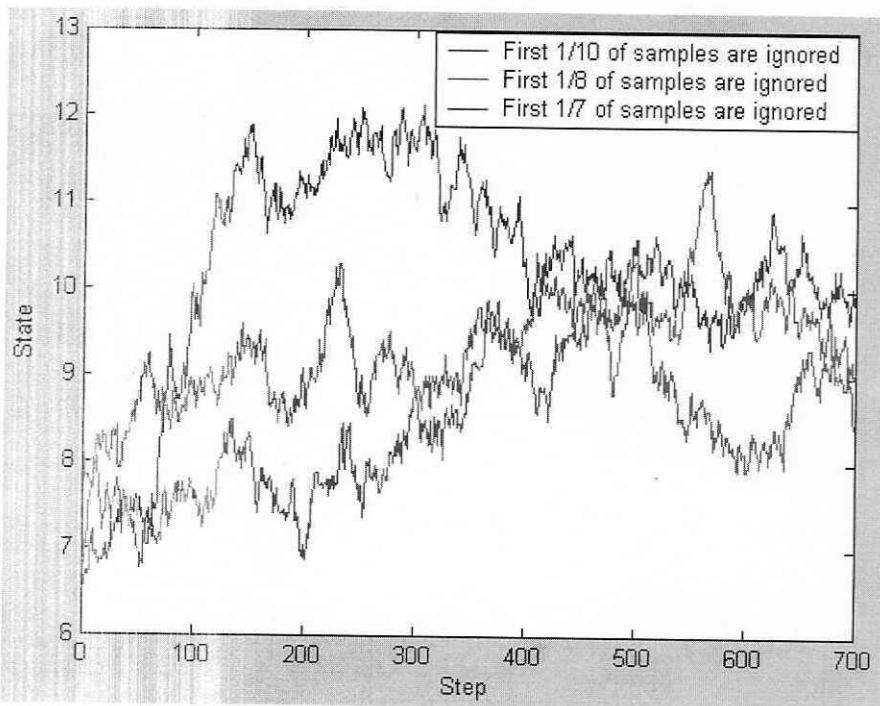
for i=1:Nmax
    pmf(i)=cnt(i)/l;
end

function [stseq,OccTime,n]= contTime(Nmax,G,IC,T)

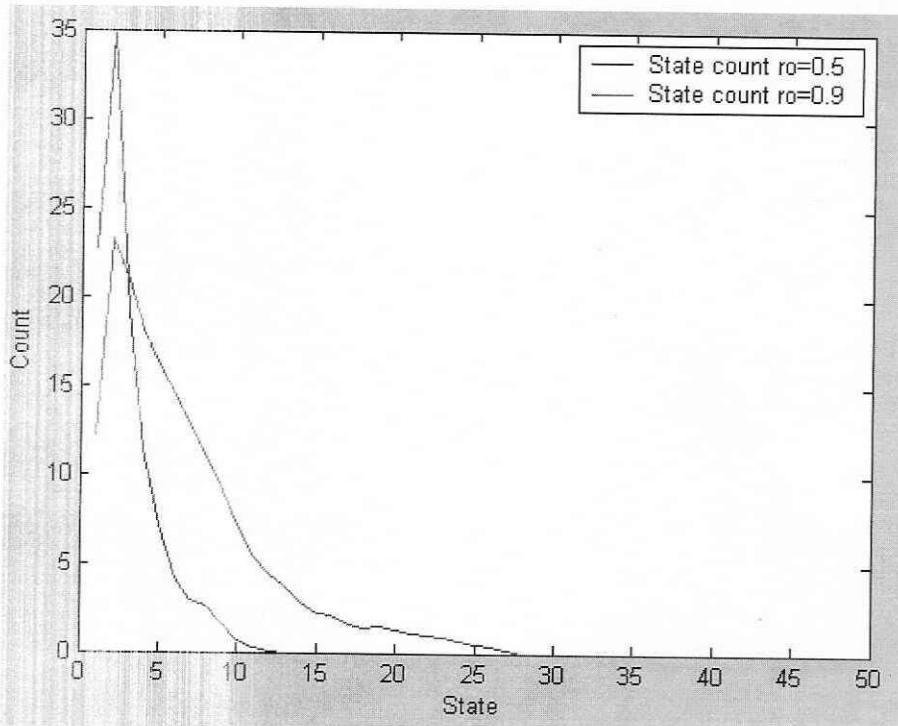
Taggr=-1;
L=round(T*(G(Nmax-1,1)+G(Nmax-1,2)));
%stseq=zeros(1,L);
%OccTime=zeros(1,L+1);
%Q=zeros(Nmax,2);
Q=G/(G(Nmax-1,1)+G(Nmax-1,2));
s=[1:Nmax+1];
step=[-1,1];
Initst= ceil(10*rand);
stseq(1)=Initst;
%Initst=1;
n=1;
OccTime(n)=randraw('exp',G(stseq(n),1)+G(stseq(n),2));
Taggr=OccTime(n);
while(Taggr<T)
    n=n+1;
    Q(stseq(n-1),:)= [G(stseq(n-1),1),G(stseq(n-1),2)]/(G(stseq(n-1),1)+G(stseq(n-1),2));
    nxt=dscRnd(1,Q(stseq(n-1),:),step);
    nextst=stseq(n-1)+nxt;
    stseq(n)=nextst;
    OccTime(n)=randraw('exp',G(stseq(n),1)+G(stseq(n),2));
    Taggr=Taggr+OccTime(n);
end

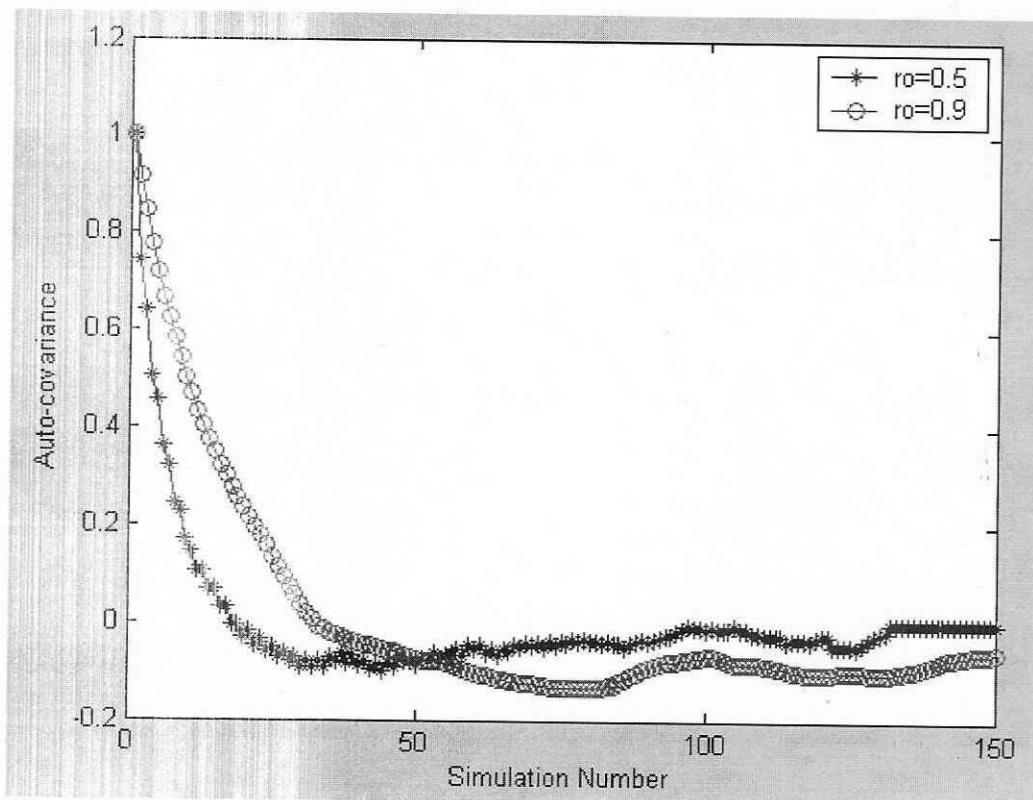
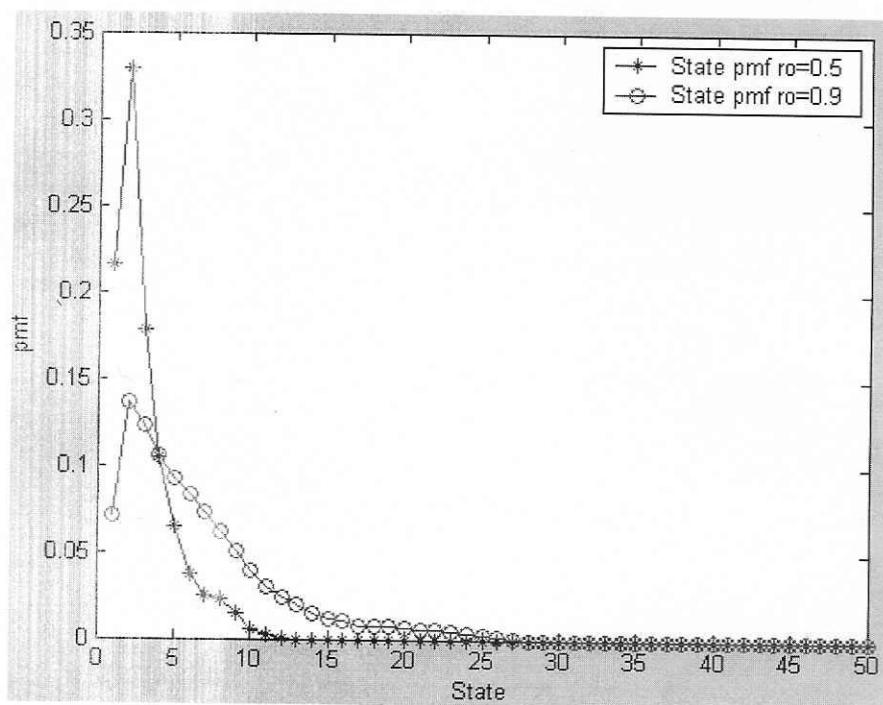
```

This diagram compares three different warm-up periods:



The diagram of stay in each one of the states and the pmf's:





12.73

```
% Prepare Transition Probability Matrix
ro=.7;
s=1000;
cnt=zeros(s+1,1);
P=zeros(s,s);
for j=1:s
    P(1,j)=exp(-ro)*ro^j/factorial(j);
    P(2,j)=exp(-ro)*ro^j/factorial(j);
end

for i=3:s
    for j=i-1:s
        P(i,j)=exp(-ro)*ro^(j-i+2)/factorial(j-i+2);
    end
end

L=s;
stseq=zeros(1,L+1);
step=1:s;
Initst= ceil(10*rand);
%Initst=1;
stseq(1)=Initst;
for n=2:L+1
    stseq(n)=dscRnd(1, P(stseq(n)+1,:), step);;
end

l=length(stseq);
for i=1:l
    idx=stseq(i);
    cnt(idx+1)=cnt(idx+1)+1;
end

for i=1:s+1
    pmf(i)=cnt(i)/s;
end

plot(stseq);

function [sample] = dscRnd(np, pdf, v)
    %if (sum(pdf) ~= 1)
    % error('Probabilities does not sum up to 1');
    %end

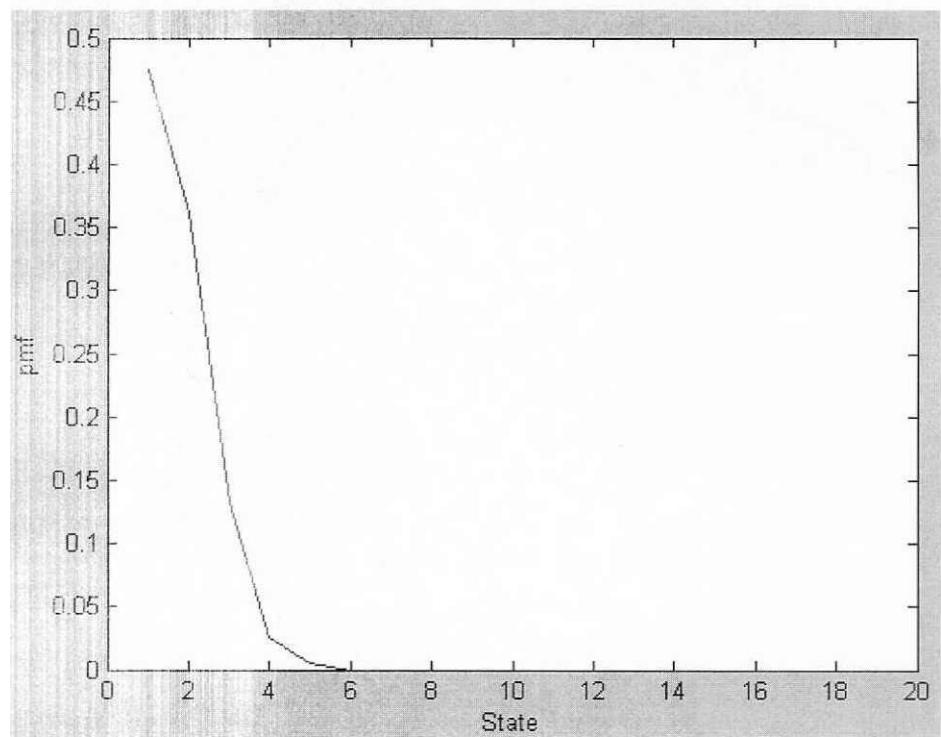
    n = length(pdf);

    if (nargin==2)
        v = [1:n];
    end

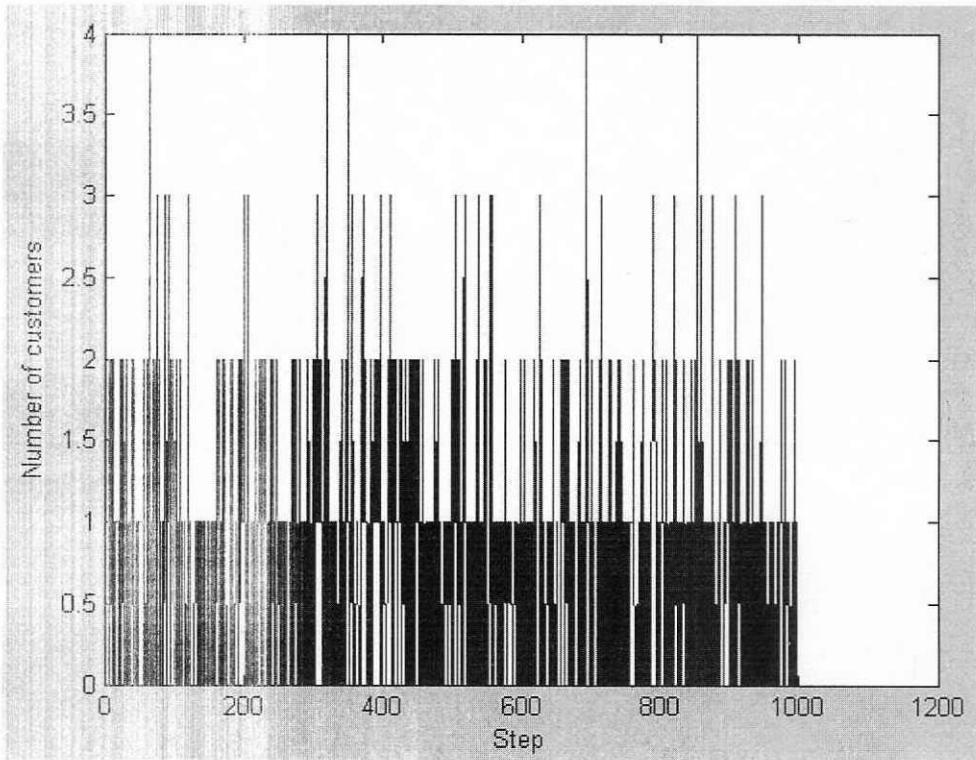
    cumprob = [0 cumsum(pdf)];

    runi = rand(1, np); % random uniform sample

    sample = zeros(1, np);
    for j=1:n
        ind = find((runi>cumprob(j)) & (runi<=cumprob(j+1)));
        sample(ind) = v(j);
    end
```

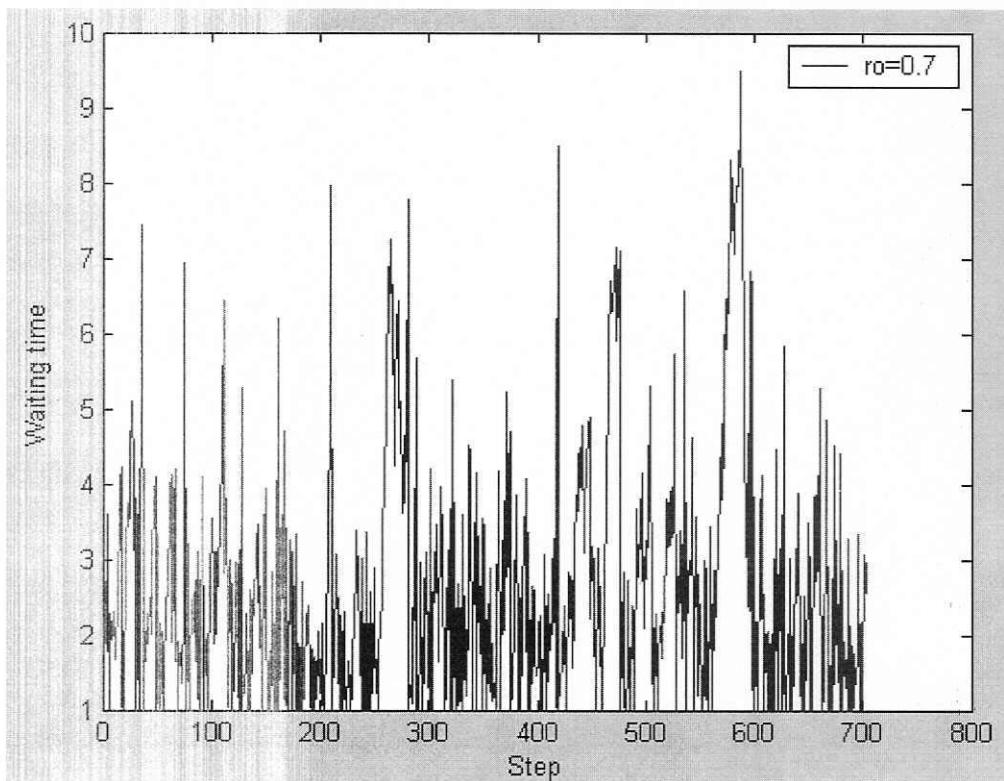


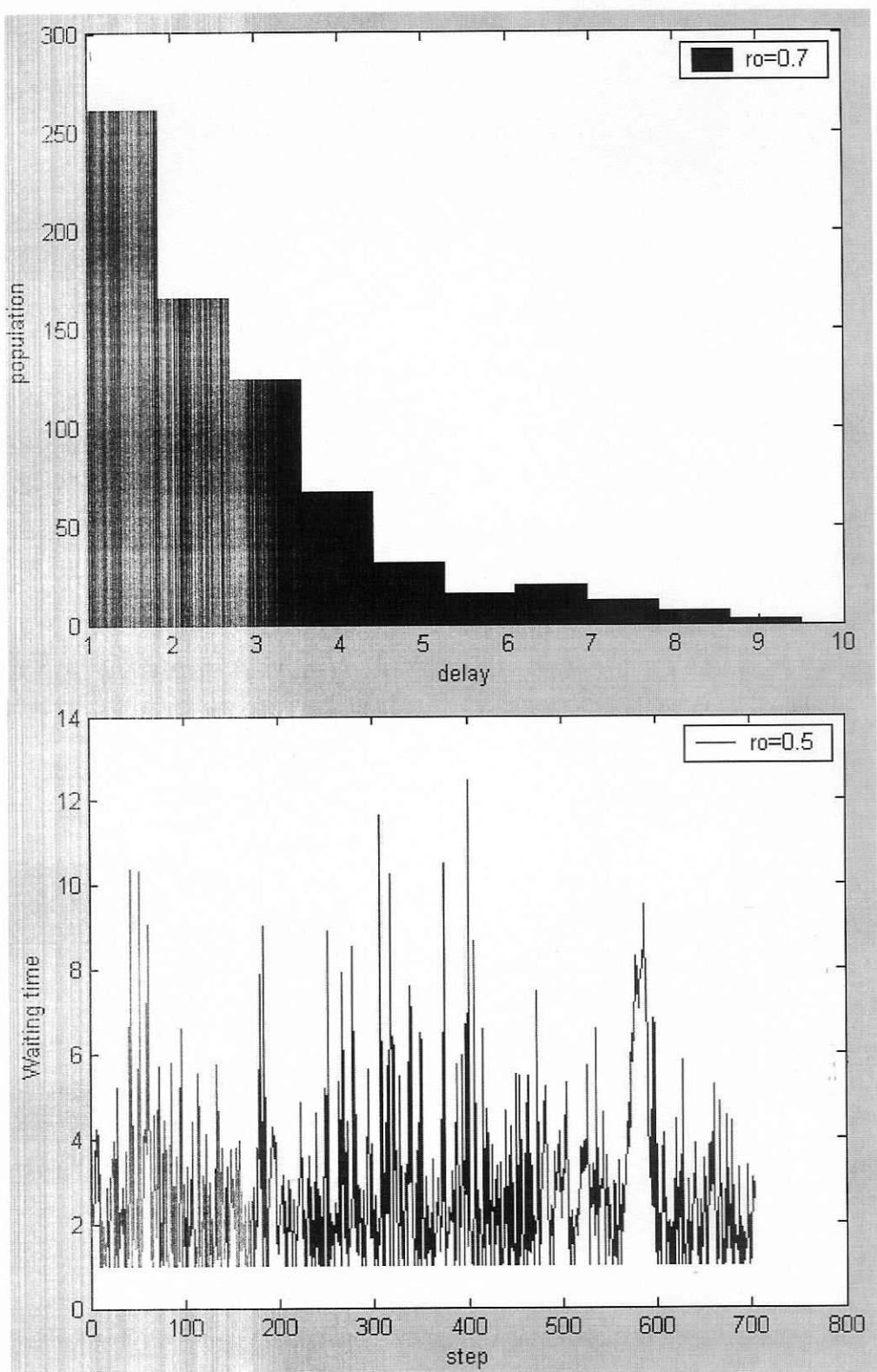
Number of customers in each step:

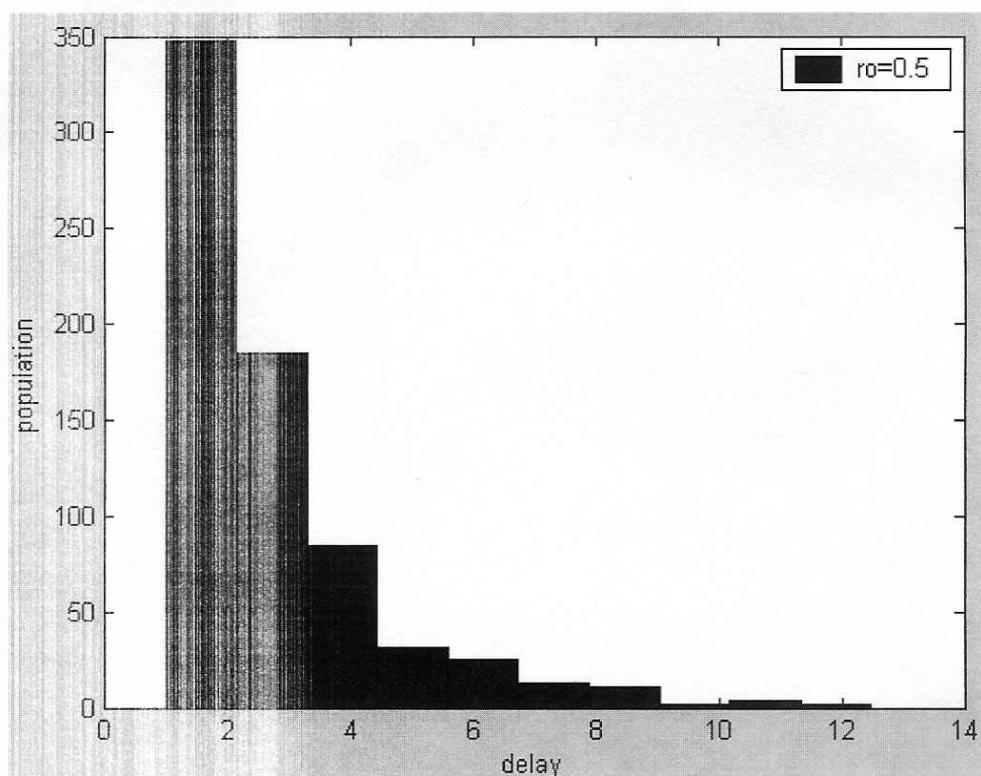


12.74

```
T=1000;
lambda=0.5;
arrtime=-log(rand)/lambda; % Poisson arrivals
i=1;
while (min(arrtime(i,:))<=T)
    arrtime = [arrtime; arrtime(i, :) - log(rand)/lambda];
    i=i+1;
end
n=length(arrtime);      % arrival times t_1,...t_n
w=ones(1,n+1);
for j=2:n-1
    w(j+1)=max(0, w(j)+1-(arrtime(j+1)-arrtime(j)));
    TT(j)=w(j)+arrtime(j)-arrtime(j-1);
end
plot(TT);
```





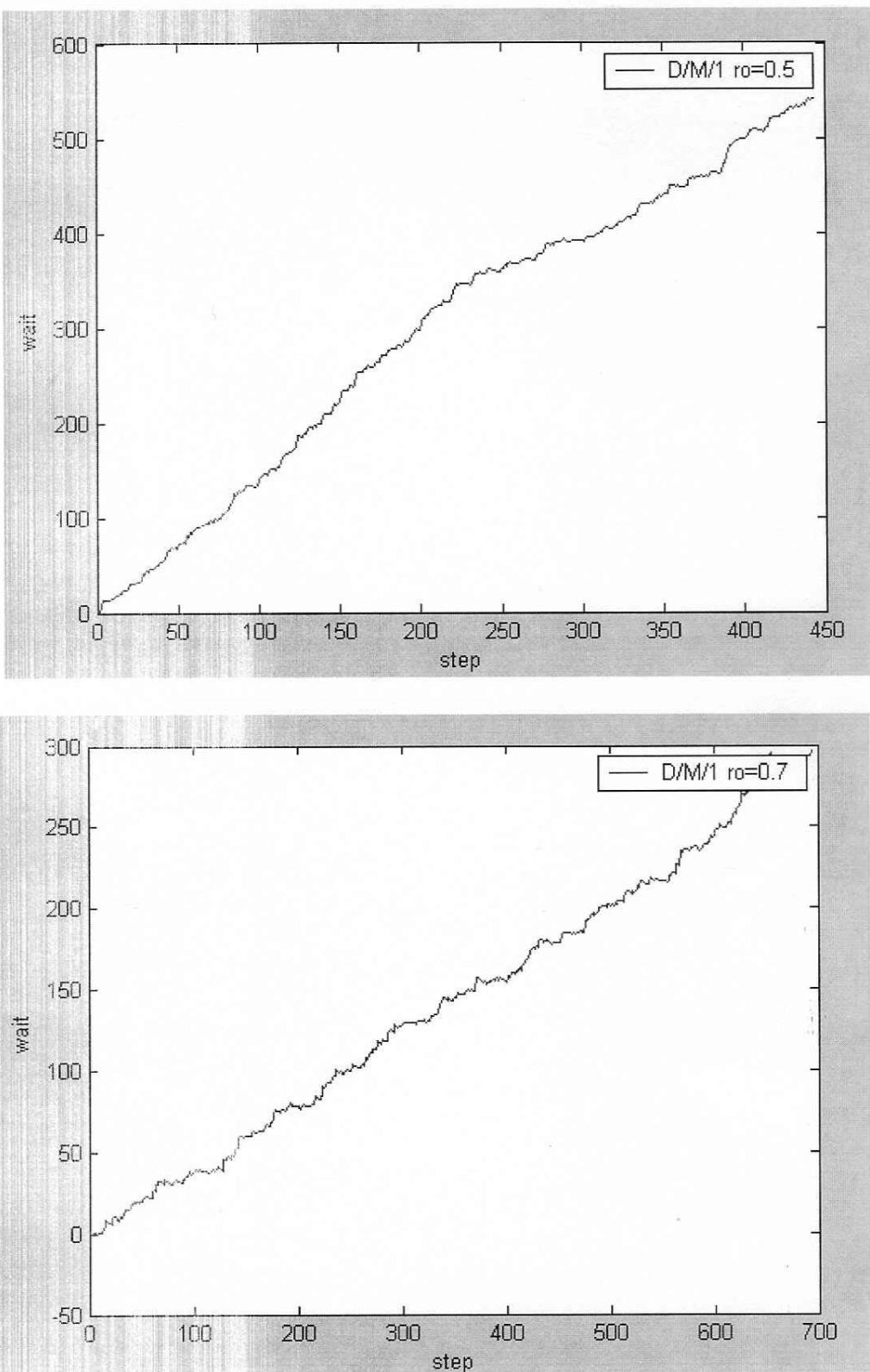


12.75

```

T=1000;
lambda=0.5;
srvtime=-log(rand)/lambda; % Poisson arrivals
i=1;
while (min(srvtime(i,:))<=T)
    srvtime = [srvtime; srvtime(i,:)-log(rand)/lambda];
    i=i+1;
end
n=length(srvtime);      % arrival times t_1,...t_n
w=ones(1,n+1);
for j=2:n-1
    w(j+1)=max(0, w(j)+(srvtime(j+1)-srvtime(j))-1);
    TT(j)=w(j)-1;
end
plot(TT(1:n-1));

```

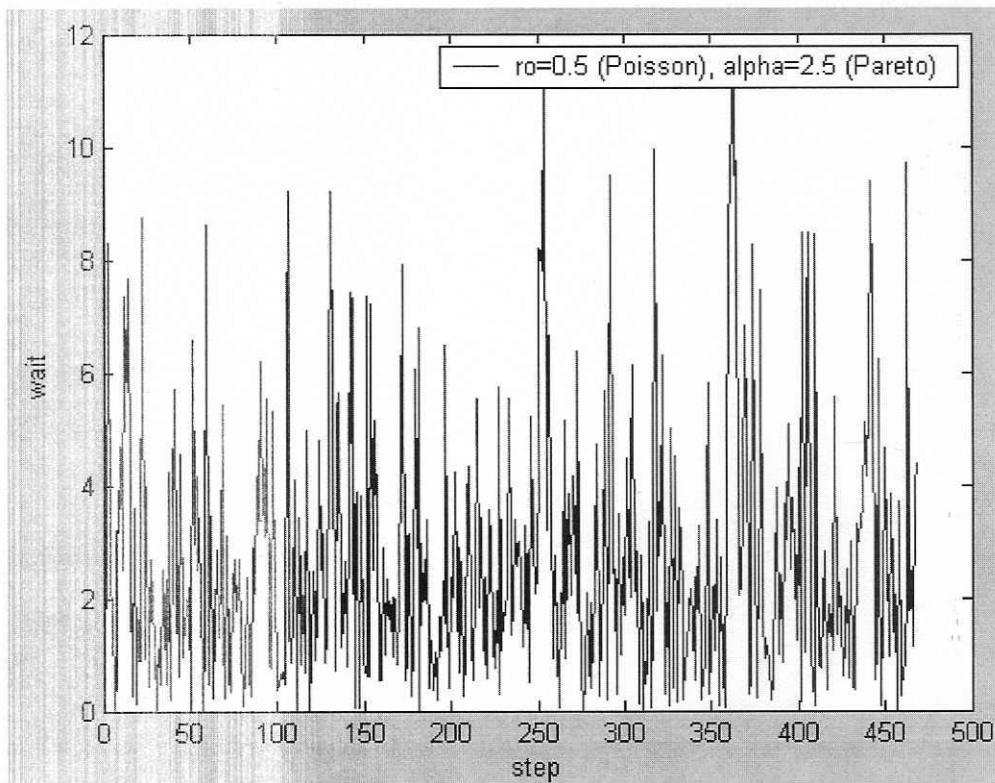


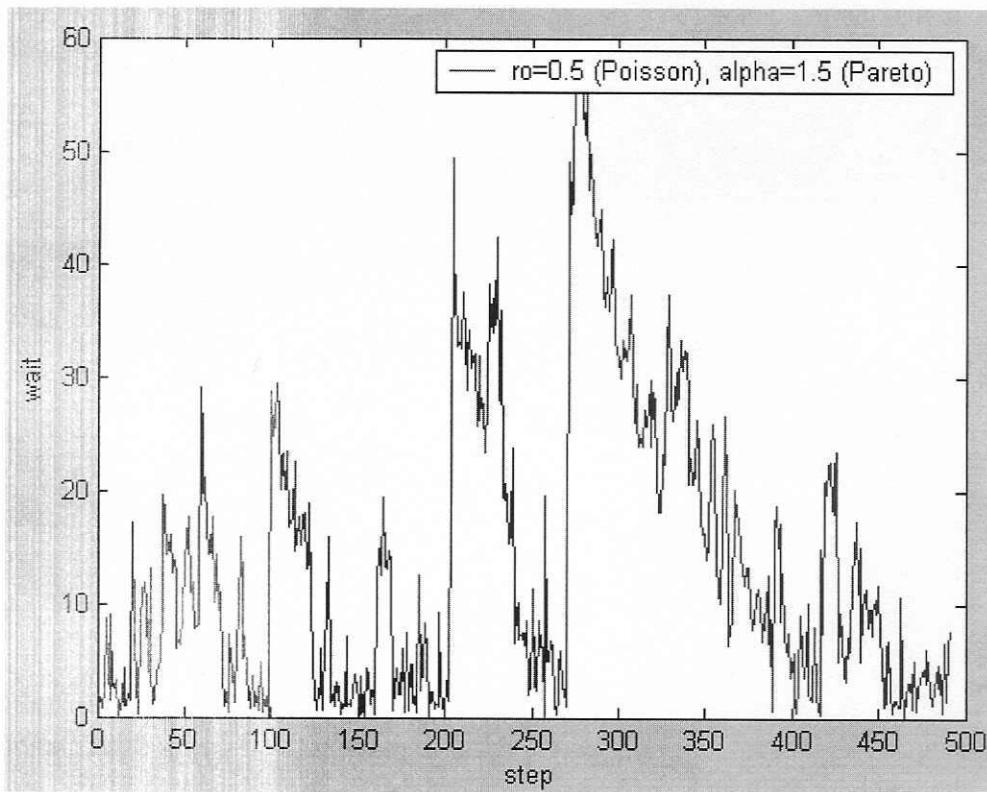
12.76

```

T=1000;
lambda=0.5;
alpha=1.5;
arrrtime=-log(rand)/lambda; % Poisson arrivals
srvtime=(1-rand).^(1/alpha)-1; % Pareto
i=1;
while (min(arrrtime(i,:))<=T)
    arrtime = [arrrtime; arrrtime(i, :)-log(rand)/lambda];
    srvtime = [srvtime; srvtime(i, :)+(1-rand).^(1/alpha)-1];
    i=i+1;
end
n=length(arrrtime);      % arrival times t_1,...t_n
w=ones(1,n+1);
for j=2:n-1
    w(j+1)=max(0, w(j)+(srvtime(j)-srvtime(j-1))-(arrrtime(j+1)-arrrtime(j)));
    TT(j)=w(j)+(arrrtime(j+1)-arrrtime(j));
end
plot(TT(1:n-1));

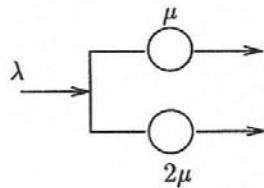
```





Problems Requiring Cumulative Knowledge

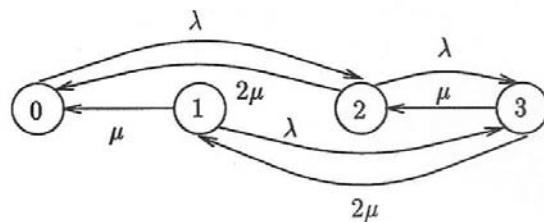
12.80



a) There are four possible cases: both servers are idle, the slower server is idle, the faster server is idle or both servers are busy. If we define each of these cases as a state, then $X(t)$ will be a four-state continuous-time Markov chain.

b) In this case, the transition rate diagram will be as follows:

- State 1: Slower server is busy & faster server is idle.
State 2: Faster server is busy & slower server is idle.



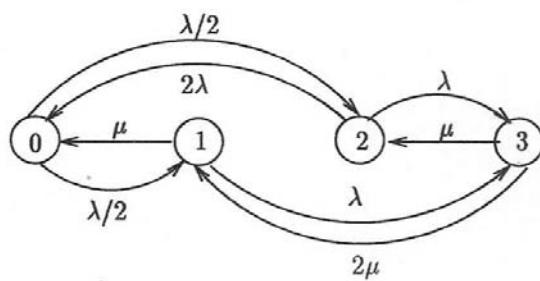
Writing global balance equations, we will have:

$$\begin{aligned} P_0\lambda &= \mu P_1 + 2\mu P_2 \\ 2\mu P_3 &= (\mu + \lambda)P_1 \\ (\lambda + 2\mu)P_2 &= \mu P_3 + \lambda P_0 \\ 2\mu P_3 &= \lambda P_2 + \lambda P_1 \end{aligned}$$

Incorporating the fact that $\sum_i P_i = 1$ and having $\rho = \frac{\lambda}{\mu}$, the following is obtained:

$$\begin{aligned} P_0 &= \frac{2\rho + 3}{\rho^3 + 2\rho^2 + 3.5\rho + 3} \\ P_1 &= \frac{\rho^2}{\rho^3 + 2\rho^2 + 3.5\rho + 3} \\ P_2 &= \frac{0.5\rho^2 + 1.5\rho}{\rho^3 + 2\rho^2 + 3.5\rho + 3} \\ P_3 &= \frac{\rho^3 + 0.5\rho^2}{\rho^3 + 2\rho^2 + 3.5\rho + 3} \end{aligned}$$

c) The Markov chain is as follows:

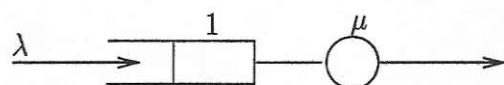


$$\begin{aligned}\rho P_0 &= \mu P_1 + 2\mu P_2 \\ 2\mu P_3 + \frac{\lambda}{2} P_0 &= (\lambda + \mu) P_1 \\ (\lambda + 2\mu) P_2 &= \mu P_3 + \frac{\lambda}{2} P_0 \\ 3\mu P_3 &= \lambda P_2 + \lambda P_1\end{aligned}$$

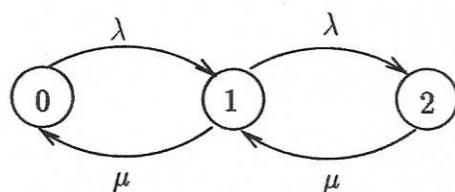
$$\begin{aligned}\Rightarrow P_0 &= \frac{4}{\rho^2 + 3\rho + 4} \\ P_1 &= \frac{2\rho}{\rho^2 + 3\rho + 4} \\ P_2 &= \frac{\rho}{\rho^2 + 3\rho + 4} \\ P_3 &= \frac{\rho^2}{\rho^2 + 3\rho + 4}\end{aligned}$$

It is expected that the average waiting time in the case of part (b) will be less than that of part (c).

12.81

9.59 M/M/1/2

This is a three-state continuous-time Markov chain:



The associated Chapman-Kolmogorov: $P'_i = \sum_j \Pi_{ji} P_j$

$$\Rightarrow \begin{aligned} P'_0 &= -\lambda P_0 + \mu P_1 \\ P'_1 &= \lambda P_0 + \mu P_2 - (\mu + \lambda) P_1 \\ P'_2 &= \lambda P_1 - \mu P_2 \end{aligned}$$

This set of differential equations can be solved much easier using Laplace transform:

$$\left\{ \begin{array}{l} sP_0(s) - P_0(0) = -\lambda P_0(s) + \mu P_1(s) \\ sP_1(s) - P_1(0) = \lambda P_0(s) + \mu P_2(s) - (\mu + \lambda) P_1(s) \\ sP_2(s) - P_2(0) = \lambda P_1(s) - \mu P_2(s) \end{array} \right.$$

$P_0(s)$, $P_1(s)$, and $P_2(s)$ are obtained by solving this set of linear equations from which the transient pmfs, $P_i(t) = P[N(t) = i]$, are obtained.

a) Suppose that the system is empty at $t = 0 \Rightarrow P_0(0) = 1, P_1(0) = P_2(0) = 0$

$$\Rightarrow \begin{aligned} P_1(s) &= \frac{\lambda(s + \mu)}{s(s^2 + 2(\mu + \lambda)s + \mu\lambda + \lambda^2 + \mu^2)} = \frac{\lambda(s + \mu)}{F(s)} \\ P_0(0) &= \frac{\mu\lambda(s + \mu)}{s(s + \mu)(s^2 + 2(\mu + \lambda)s + \mu\lambda + \lambda^2 + \mu^2)} + \frac{1}{s + \lambda} \\ P_2(s) &= \frac{\lambda^2}{s(s^2 + 2(\mu + \lambda)s + \mu\lambda + \lambda^2 + \mu^2)} \end{aligned}$$

\Rightarrow By partial fraction expansion, we will have:

$$P_i(s) = \frac{A_i}{s} + \frac{B_i}{s + \lambda} + \frac{C_i}{s + a_1} + \frac{D_i}{s + a_2}$$

where $a_1, a_2 = -(\mu + \lambda) \pm \sqrt{\mu\lambda}$.

Applying inverse Laplace transform, we obtain:

$$P_i(t) = A_i + B_i e^{-\lambda t} + C_i e^{-a_1 t} + D_i e^{-a_2 t}$$

Finding the factors $A_i - D_i$ is very easy because $P_i(s)$ have only simple 1st order roots.

b) If the system is full at $t = 0 \Rightarrow P_2(0) = 1, P_0(0) = P_1(0) = 0$

$$\begin{aligned} \Rightarrow P_1(s) &= \frac{\mu(s + \lambda)}{F(s)} \\ P_0(s) &= \frac{\mu^2}{F(s)} \\ P_2(s) &= \frac{\mu\lambda(s + \lambda)}{(s + \mu)F(s)} + \frac{1}{s + \mu} \end{aligned}$$

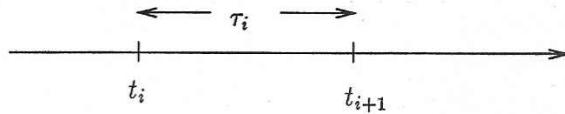
The expansion of these fractions will be as follows:

$$P_i(s) = \frac{A_i}{s} + \frac{B_i}{s + \mu} + \frac{C_i}{s + a_1} + \frac{P_i}{s + a_2}$$

$$\Rightarrow P_i(t) = A_i + B_i e^{-\mu t} + C_i e^{-a_1 t} + D_i e^{-a_2 t}$$

12.82

- 9.60 a)** Let's divide the time axis into time intervals or cycles τ_i which are the times when customers arrive to an empty system.



We have $t_j = \tau_0 + \tau_1 + \tau_2 + \dots + \tau_i$.

If the τ_i 's are identically distributed independent random variables, then they form a renewal process. In an M/G/1, each time a customer arrives to an empty system restarts a process which is independent of all past events. In addition, all cycles have the same statistics, i.e. they constitute a repeated trial.

- b) Let $N(t)$ be the number of customers in the system.

$$P[N(t) = j] = \lim_{t \rightarrow \infty} \frac{\text{Total time that there are 'j' customers in the system}}{t}$$

To find this, let's associate a cost, C_i , to each renewal interval τ_i . C_i is equal to amount of time that there are ' j ' customers in the system in time τ_i . Also, $C(t) = \sum_i C_i$.

$$P[N(t) = j] = \lim_{t \rightarrow \infty} \frac{C(t)}{t}$$

According to the renewal theorem

$$\lim_{t \rightarrow \infty} \frac{C(t)}{t} = \frac{E(C)}{E[\tau]}$$

Therefore,

$$P[N(t) = j] = \frac{E(C)}{E(\tau)} \approx \frac{\frac{1}{n} \sum_{j=1}^n C_j}{\frac{1}{n} \sum_{j=1}^n \tau_j}$$

- c) We can find confidence intervals for the numerator and denominator of the above expression using the methods from section 8.4.

12.83

$$\begin{aligned} \text{9.61 a)} \quad P[N_1(t) = i, N_2(t) = j] &= P[N_1(t) = i, N(t) = i + j] \\ &= P[N_1(t) = i / N(t) = i + j] P[N(t) = i + j] \end{aligned}$$

If we know that $N(t) = i + j$, then each of the $i + j$ arrivals picks its arrival time according to a uniform distribution in the interval $(0, t)$ independently of the other arrivals

Therefore the probability that an arrival for $N(t)$ is also an arrival for $N_1(t)$ is

$$\int_0^t Pr[\text{arrival for } N_1(t)/\text{arrival time is } t] \frac{1}{t} dt = \frac{1}{t} \int_0^t p(t) dt \triangleq p$$

Since the $i + j$ arrivals are independent, we have

$$P[N_1(t) = i | N(t) = i + j] = \binom{i+j}{i} p^i (1-p)^j$$

Finally, we have

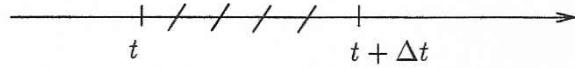
$$\begin{aligned}
 P[N_1(t) = i, N_2(t) = j] &= \binom{i+j}{i} p^i (1-p)^j \frac{(\lambda t)^{i+j}}{(i+j)!} e^{-\lambda t} \\
 &= \frac{p^i (1-p)^j}{i! j!} (\lambda t)^{i+j} e^{-\lambda t} \\
 &= \frac{(\lambda p)^i}{i!} \frac{(\lambda(1-p))^j}{j!} e^{-\lambda t} e^{-\lambda(1-p)t} \\
 &= P[N_1(t) = i] P[N_2(t) = j].
 \end{aligned}$$

b) In order for $N_1(t)$ and $N_2(t)$ to be independent random processes we require that

$$P[N_1(t_1) = i, N_2(t_2) = j] = P[N_1(t_1) = i] P[N_2(t_2) = j]$$

for any i, j and any choices of times t_1 and t_2 .

Consider a very small time interval at time t of duration Δt :



Consider the evolution of the joint process $(N_1(t), N_2(t))$

$$\begin{aligned}
 P[N_1(t + \Delta t) = i, N_2(t + \Delta t) = j] &= P[N_1(t) = i - 1, \text{ a type 1 arrival in } \Delta t] \\
 &\quad + P[N_2(t) = j - 1, \text{ a type 2 arrival in } \Delta t] \\
 &\quad + P[N_1(t) = i, N_2(t) = j, \text{ no arrival in } \Delta t] \\
 &\quad + O(\Delta t) \\
 &= P[N_1(t) = i - 1] \lambda p(t) \Delta t \\
 &\quad + P[N_2(t) = j - 1] \lambda(1 - p(t)) \Delta t \\
 &\quad + P[N_1(t) = i, N_2(t) = j] (1 - \lambda \Delta t) \\
 &\quad + O(\Delta t)
 \end{aligned}$$

Now consider the individual evolution of the two processes:

$$\begin{aligned}
 P[N_1(t + \Delta t) = i] &= P[N_1(t) = i - 1] \lambda p(t) \Delta t \\
 &\quad + P[N_1(t) = i] (1 - \lambda p(t) \Delta t) + O(\Delta t) \\
 P[N_2(t + \Delta t) = j] &= P[N_2(t) = j - 1] \lambda(1 - p(t)) \Delta t \\
 &\quad + P[N_2(t) = j] (1 - \lambda(1 - p(t)) \Delta t) + O(\Delta t)
 \end{aligned}$$

If we multiply these two equations we obtain:

$$\begin{aligned}
 P[N_1(t + \Delta t) = i] P[N_2(t + \Delta t) = j] &= P[N_1(t) = i - 1] P[N_2(t) = j] \lambda p(t) \Delta t \\
 &\quad + P[N_1(t) = i] P[N_2(t) = j - 1] \lambda(1 - p(t)) \Delta t \\
 &\quad + P[N_1(t) = i] P[N_2(t) = j] (1 - \lambda \Delta t) + O(\Delta t)
 \end{aligned}$$

By comparing this equation to that of the joint process, we see that we will end up with independent random processes.

12.84

- 9.62 a) Suppose a customer arrived at time $t_1 < t$, then the customer has completed its service time by time t with probability

$$P[X < t - t_1] = F_X(t - t_1) \triangleq p(t_1)$$

Therefore customers that arrived in the interval $(0, t)$ have probability of completing service:

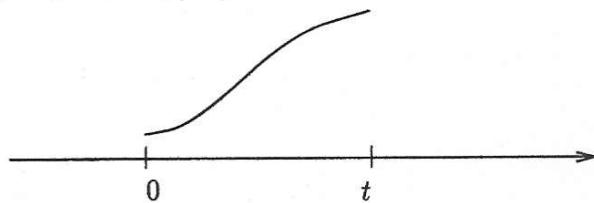
$$p = \frac{1}{t} \int_0^t F_X(t - t_1) dt_1$$

Thus

$$P[N_1(t) = i, N_2(t) = j] = \frac{(\lambda t p)^i}{i!} e^{-\lambda t p} \frac{(\lambda t(1-p))^j}{j!} e^{-\lambda t(1-p)}$$

Note that

$$\begin{aligned} \lambda t p &= \lambda \int_0^t F_X(t - t_1) dt_1 \\ \lambda t(1-p) &= \lambda t \left[\frac{1}{t} \int_0^t (1 - F_X(t - t_1)) dt_1 \right] \\ &= \lambda \int_0^t (1 - F_X(t - t_1)) dt_1 \end{aligned}$$

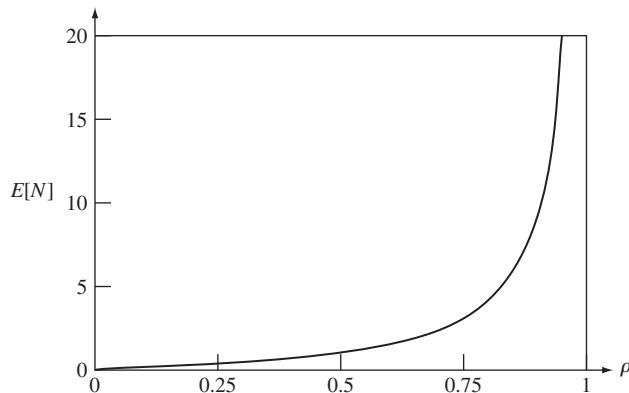


- b) As $t \rightarrow \infty$

$$\begin{aligned} \lambda t(1-p) &\rightarrow \lambda \int_0^\infty (1 - F_X(t)) dt = \lambda E[X] \\ \therefore P[N_2(t) = j] &= \frac{(\lambda E[X])^j}{j!} e^{-\lambda E[X]} \quad \text{Poisson RV!} \end{aligned}$$

- c) Little's formula

$$E[N] = \lambda E[X] \quad \checkmark$$

**FIGURE 12.5**

Mean number of customers in the system versus utilization for M/M/1 queue.

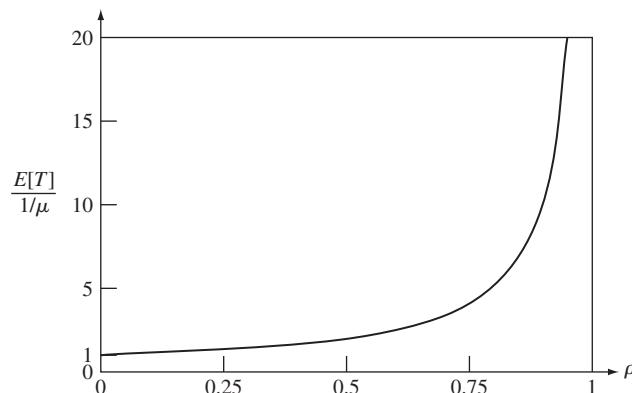
The server utilization (defined in Example 12.2) is given by

$$1 - p_0 = 1 - (1 - \rho) = \rho = \frac{\lambda}{\mu}. \quad (12.29)$$

Figures 12.5 and 12.6 show $E[N]$ and $E[T]$ versus ρ . It can be seen that as ρ approaches one, the mean number in the system and the system delay become arbitrarily large.

Example 12.3

A router receives packets from a group of users and transmits them over a single transmission line. Suppose that packets arrive according to a Poisson process at a rate of one packet every 4 ms, and suppose that packet transmission times are exponentially distributed with mean 3 ms.

**FIGURE 12.6**

Mean total customer delay versus utilization for M/M/1 system. The delay is expressed in multiples of mean service times.

Find the mean number of packets in the system and the mean total delay in the system. What percentage increase in arrival rate results in a doubling of the above mean total delay?

The arrival rate is 1/4 packets/ms and the mean service time is 3 ms. The utilization is therefore

$$\rho = \frac{1}{4}(3) = \frac{3}{4}.$$

The mean number of packets in the system is then

$$E[N] = \frac{\rho}{1 - \rho} = 3.$$

The mean time in the system is

$$E[T] = \frac{E[N]}{\lambda} = \frac{3}{1/4} = 12 \text{ ms.}$$

The mean time in the system will be doubled to 24 ms when

$$24 = \frac{E[\tau]}{1 - \rho'} = \frac{3}{1 - \rho'}.$$

The resulting utilization is $\rho' = 7/8$ and the corresponding arrival rate is $\lambda' = \rho'\mu = 7/24$. The original arrival rate was 6/24. Thus an increase in arrival rate of $1/6 = 17\%$ leads to a 100% increase in mean system delay.

The point of this example is that *the onset of congestion is swift*. The mean delay increases rapidly once the utilization increases beyond a certain point.

Example 12.4 Concentration and Effect of Scale

A large processor handles transactions at a rate of $K\mu$ transactions per second. Suppose transactions arrive according to a Poisson process of rate $K\lambda$ transactions/second, and that transactions require an exponentially distributed amount of processing time. Suppose that a proposal is made to eliminate the large processor and to replace it with K processors, each with a processing rate of μ transactions per second and an arrival rate of λ . Compare the mean delay performance of the existing and the proposed systems.

The large processor system is an M/M/1 queue with arrival rate $K\lambda$, service rate $K\mu$, and utilization $\rho = K\lambda/K\mu = \lambda/\mu$. The mean delay is given by Eq. (12.26):

$$E[T] = \frac{E[\tau]}{1 - \rho} = \frac{1/K\mu}{1 - \rho}.$$

Each of the small processors is an M/M/1 system with arrival rate λ , service rate μ , and utilization $\rho = \lambda/\mu$. The mean delay is

$$E[T'] = \frac{E[\tau']}{1 - \rho} = \frac{1/\mu}{1 - \rho} = KE[T].$$

Thus, the system with the single large processor with processing rate $K\mu$ has a smaller mean delay than the system with K small processors each of rate μ . In other words, the concentration of customer demand into a single system results in significant delay performance improvement.

12.3.2 Delay Distribution in M/M/1 System and Arriving Customer's Distribution

Let N_a denote the number of customers found in the system by a customer arrival. We call $P[N_a = k]$ the **arriving customer's distribution**. We now show that if arrivals are Poisson and independent of the system state and customer service times, then the arriving customer's distribution is equal to the steady state distribution for the number in the system. A customer that arrives at time $t + \delta$ finds k in the system if $N(t) = k$, thus

$$\begin{aligned} P[N_a(t) = k] &= \lim_{\delta \rightarrow 0} P[N(t) = k | A(t + \delta) - A(t) = 1] \\ &= \lim_{\delta \rightarrow 0} \frac{P[N(t) = k, A(t + \delta) - A(t) = 1]}{P[A(t + \delta) - A(t) = 1]} \\ &= \lim_{\delta \rightarrow 0} \frac{P[A(t + \delta) - A(t) = 1 | N(t) = k] P[N(t) = k]}{P[A(t + \delta) - A(t) = 1]}, \end{aligned}$$

where we have used the definition of conditional probability. The probability of an arrival in the interval $(t, t + \delta]$ is independent of $N(t)$, thus

$$\begin{aligned} P[N_a(t) = k] &= \lim_{\delta \rightarrow 0} \frac{P[A(t + \delta) - A(t) = 1] P[N(t) = k]}{P[A(t + \delta) - A(t) = 1]} \\ &= P[N(t) = k]. \end{aligned}$$

Thus the probability that $N_a = k$ is simply the proportion of time during which the system has k customers in the system. For the M/M/1 queueing system under consideration we have

$$P[N_a = k] = P[N(t) = k] = (1 - \rho)\rho^k. \quad (12.30)$$

We are now ready to compute the distribution for the total time T that a customer spends in an M/M/1 system. Suppose that an arriving customer finds k in the system, that is, $N_a = k$. If the service discipline is “first come, first served,” then T is the residual service time of the customer found in service, the service times of the $k - 1$ customers found in queue, and the service time of the arriving customer. The memoryless property of the exponential service time implies that the residual service time of the customer found in service has the same distribution as a full service time. Thus T is the sum of $k + 1$ iid exponential random variables. In Example 7.5 we saw that this sum has the gamma pdf

$$f_T(x | N_a = k) = \frac{(\mu x)^k}{k!} \mu e^{-\mu x} \quad x > 0. \quad (12.31)$$

The pdf of T is found by averaging over the probability of an arriving customer finding k messages in the system, $P[N_a = k]$. Thus the pdf of T is

$$\begin{aligned} f_T(x) &= \sum_{k=0}^{\infty} \frac{(\mu x)^k}{k!} \mu e^{-\mu x} P[N(t) = k] \\ &= \sum_{k=0}^{\infty} \frac{(\mu x)^k}{k!} \mu e^{-\mu x} (1 - \rho)\rho^k \end{aligned}$$

The probability that an arriving customer finds all servers busy and has to wait in queue is an important parameter of the M/M/c system:

$$P[W > 0] = P[N \geq c] = \sum_{j=c}^{\infty} \rho^{j-c} p_c = \frac{p_c}{1 - \rho}. \quad (12.50)$$

This probability is called the **Erlang C formula** and is denoted by $C(c, a)$:

$$C(c, a) = \frac{p_c}{1 - \rho} = P[W > 0]. \quad (12.51)$$

The mean number of customers in queue is given by

$$\begin{aligned} E[N_q] &= \sum_{j=c}^{\infty} (j - c) \rho^{j-c} p_c = p_c \sum_{j'=0}^{\infty} j' \rho^{j'} \\ &= \frac{\rho}{(1 - \rho)^2} p_c \\ &= \frac{\rho}{1 - \rho} C(c, a). \end{aligned} \quad (12.52)$$

The mean waiting time is found from Little's formula:

$$\begin{aligned} E[W] &= \frac{E[N_q]}{\lambda} \\ &= \frac{1/\mu}{c(1 - \rho)} C(c, a). \end{aligned} \quad (12.53)$$

The mean total time in the system is

$$E[T] = E[W] + E[\tau] = E[W] + \frac{1}{\mu}. \quad (12.54)$$

Finally, the mean number in the system is found from Little's formula:

$$E[N] = \lambda E[T] = E[N_q] + a, \quad (12.55)$$

where we have used Equation (12.54).

Example 12.8

A company has two 1 Megabit/second lines connecting two of its sites. Suppose that packets for these lines arrive according to a Poisson process at a rate of 150 packets per second, and that packets are exponentially distributed with mean 10 kbits. When both lines are busy, the system queues the packets and transmits them on the first available line. Find the probability that a packet has to wait in queue.

First we need to compute p_0 . The system parameters are $c = 2$, $\lambda = 150$ packets/sec, $1/\mu = 10$ kbit/1 Mbit/s = 10 ms, $a = \lambda/\mu = 1.5$ and $\rho = \lambda/c\mu = 3/4$. Therefore:

$$p_0 = \left\{ 1 + 1.5 + \frac{(1.5)^2}{2!} \frac{1}{1 - 3/4} \right\}^{-1} = \frac{1}{7}.$$

The probability of having to wait is then

$$C(2, 1.5) = \frac{(1.5)^2}{2!} p_0 \frac{1}{1 - \rho} = \frac{9}{14}.$$

Example 12.9 M/M/1 Versus M/M/c

Compare the mean delay and mean waiting time performance of the two systems shown in Fig. 12.11. Note that both systems have the same processing rate.

For the M/M/1 system, $\rho = \lambda/\mu = (1/2)/1 = 1/2$, so the mean waiting time is

$$E[W] = \frac{\rho/\mu}{1 - \rho} = 1 \text{ s},$$

and the mean total delay is

$$E[T] = \frac{1/\mu}{1 - \rho} = 2 \text{ s}.$$

For the M/M/2 system, $a = \lambda/\mu' = 1$, and $\rho = \lambda/2\mu' = 1/2$. The probability of an empty system is

$$p_0 = \left\{ 1 + a + \frac{a^2/2}{1 - 1/2} \right\}^{-1} = \frac{1}{3}.$$

The Erlang C formula is

$$C(2, 1) = \frac{a^2/2}{1 - \rho} p_0 = \frac{1}{3}.$$

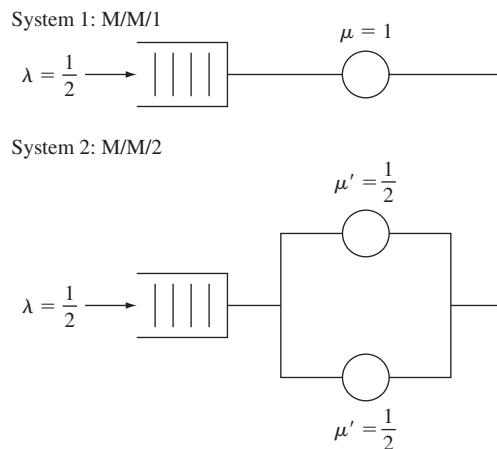


FIGURE 12.11
M/M/1 and M/M/2 systems with the same arrival rate
and the same maximum processing rate.

The mean waiting time is then

$$E[W'] = \frac{1/\mu'}{2(1 - \rho)} C(2, 1) = \frac{2}{3},$$

and the mean delay is

$$E[T'] = \frac{2}{3} + \frac{1}{\mu'} = \frac{8}{3}.$$

Thus the M/M/1 system has a smaller total delay but a larger waiting time than the M/M/2. In general, increasing the number of servers decreases the waiting time but increases the total delay.

12.4.2 Waiting Time Distribution for M/M/c

Before we compute the pdf of the waiting time, consider the conditional probability that there are $j - c > 0$ customers in queue given that all servers are busy (i.e., $N(t) \geq c$):

$$\begin{aligned} P[N(t) = j | N(t) \geq c] &= \frac{P[N(t) = j, N(t) \geq c]}{P[N(t) \geq c]} = \frac{P[N(t) = j]}{P[N(t) \geq c]} \quad j \geq c \\ &= \frac{\rho^{j-c} p_c}{p_c / (1 - \rho)} = (1 - \rho) \rho^{j-c} \quad j \geq c. \end{aligned} \quad (12.56)$$

This geometric pmf suggests that when all the servers are busy, the M/M/c system behaves like an M/M/1 system. We use this fact to compute the cdf of W .

Suppose that a customer arrives when there are k customers in queue. There must be $k + 1$ service completions before our customer enters service. From Eq. (12.43), each service completion is exponentially distributed with rate $c\mu$. Thus the waiting time for our customer is the sum of $k + 1$ iid exponential random variables with parameter $c\mu$, which we know is a gamma random variable with parameter $c\mu$:

$$f_W(x | N = c + k) = \frac{(c\mu x)^k}{k!} c\mu e^{-c\mu x}. \quad (12.57)$$

The cdf for W given that $W > 0$, or equivalently $N \geq c$, is obtained by combining Eqs. (12.56) and (12.57):

$$\begin{aligned} F_W(x | W > 0) &= \sum_{k=0}^{\infty} F_W(x | N = c + k) P[N = c + k | N \geq c] \\ &= \sum_{k=0}^{\infty} \int_0^x \frac{(c\mu y)^k}{k!} c\mu e^{-c\mu y} dy (1 - \rho) \rho^k \\ &= (1 - \rho) \int_0^x \sum_{k=0}^{\infty} \frac{(c\mu y)^k}{k!} \rho^k c\mu e^{-c\mu y} dy \\ &= (1 - \rho) c\mu \int_0^x e^{-c\mu(1-\rho)y} dy \\ &= 1 - e^{-c\mu(1-\rho)x}. \end{aligned}$$

Examples 12.18 and 12.19 demonstrate that the Pollaczek–Khinchin transform equations can be used to obtain closed-form expressions for the pmf of $N(t)$ and the pdf's of W and T when the Laplace transform of the service time pdf is a rational function of s , that is, a ratio of polynomials in s . This result is particularly important because it can be shown that the Laplace transform of any service time pdf can be approximated arbitrarily closely by a rational function of s . Thus in principle we can obtain exact expressions for the pmf of $N(t)$ and pdf's of W and T .

In addition it should be noted that *the Pollaczek–Khinchin transform expressions can always be inverted numerically* using fast Fourier transform methods such as those discussed in Section 7.6. This numerical approach does not require that the Laplace transform of the pdf be a rational function of s .

12.8 BURKE'S THEOREM: DEPARTURES FROM M/M/c SYSTEMS

In many problems, a customer requires service from several service stations before a task is completed. These problems require that we consider a *network* of queueing systems. In such networks, the departures from some queues become the arrivals to other queues. This is the reason why we are interested in the statistical properties of the departure process from a queue.

Consider two queues in tandem as shown in Fig. 12.21, where the departures from the first queue become the arrivals at the second queue. Assume that the arrivals to the first queue are Poisson with rate λ and that the service time at queue 1 is exponentially distributed with rate $\mu_1 > \lambda$. Assume that the service time in queue 2 is also exponentially distributed with rate $\mu_2 > \lambda$.

The state of this system is specified by the number of customers in the two queues, $(N_1(t), N_2(t))$. This state vector forms a Markov process with the transition rate diagram shown in Fig. 12.22, and global balance equations:

$$\lambda P[N_1 = 0, N_2 = 0] = \mu_2 P[N_1 = 0, N_2 = 1] \quad (12.135a)$$

$$(\lambda + \mu_1)P[N_1 = n, N_2 = 0] = \mu_2 P[N_1 = n, N_2 = 1] + \lambda P[N_1 = n - 1, N_2 = 0] \quad n > 0 \quad (12.135b)$$

$$(\lambda + \mu_2)P[N_1 = 0, N_2 = m] = \mu_2 P[N_1 = 0, N_2 = m + 1] + \mu_1 P[N_1 = 1, N_2 = m - 1] \quad m > 0 \quad (12.135c)$$

$$(\lambda + \mu_1 + \mu_2)P[N_1 = n, N_2 = m] = \mu_2 P[N_1 = n, N_2 = m + 1] + \mu_1 P[N_1 = n + 1, N_2 = m - 1] + \lambda P[N_1 = n - 1, N_2 = m] \quad n > 0, m > 0. \quad (12.135d)$$

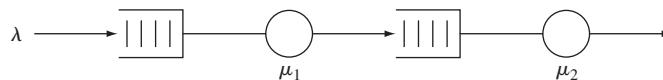
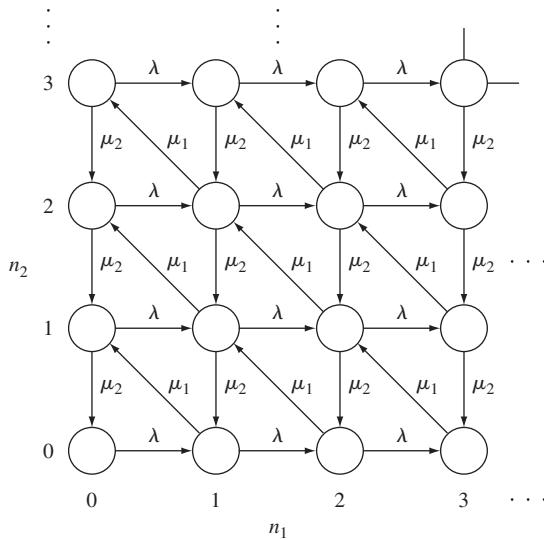


FIGURE 12.21
Two tandem exponential queues with Poisson input.

**FIGURE 12.22**

Transition rate diagram for two tandem exponential queues with Poisson input.

It is easy to verify that the following joint state pmf satisfies Eqs. (12.135a) through (12.135d):

$$P[N_1 = n, N_2 = m] = (1 - \rho_1)\rho_1^n(1 - \rho_2)\rho_2^m \quad n \geq 0, m \geq 0, \quad (12.136)$$

where $\rho_i = \lambda/\mu_i$. We know that the first queue is an M/M/1 system, so

$$P[N_1 = n] = (1 - \rho_1)\rho_1^n \quad n = 0, 1, \dots \quad (12.137)$$

By summing Eq. (12.136) over all n , we obtain the marginal state pmf of the second queue:

$$P[N_2 = m] = (1 - \rho_2)\rho_2^m \quad m \geq 0. \quad (12.138)$$

Equations (12.136) through (12.138) imply that

$$P[N_1 = n, N_2 = m] = P[N_1 = n]P[N_2 = m] \quad \text{for all } n, m. \quad (12.139)$$

In words, *the number of customers at queue 1 and the number at queue 2 at the same time instant are independent random variables*. Furthermore, *the steady state pmf at the second queue is that of an M/M/1 system with Poisson arrival rate λ and exponential service time μ_2* .

We say that a network of queues has a **product-form solution** when the joint pmf of the vector of numbers of customers at the various queues is equal to the product of the marginal pmf's of the number in the individual queues. We now discuss Burke's theorem, which states the fundamental result underlying the product-form solution in Eq. (12.139).

Burke's Theorem

Consider an M/M/1, M/M/c, or M/M/ ∞ queueing system at steady state with arrival rate λ , then

1. The departure process is Poisson with rate λ .
 2. At each time t , the number of customers in the system $N(t)$ is independent of the sequence of departure times prior to t .
-

The product-form solution for the two tandem queues follows from Burke's theorem. Queue 1 is an M/M/1 queue, so from part 1 of the theorem the departures from queue 1 form a Poisson process. Thus the arrivals to queue 2 are a Poisson process, so the second queue is also an M/M/1 system with steady state pmf given by Eq. (12.138). It remains to show that the numbers of customers in the two queues at the same time instant are independent random variables.

The arrivals to queue 2 prior to time t are the departures from queue 1 prior to time t . By part 2 of Burke's theorem the departures from queue 1, and hence the arrivals to queue 2, prior to time t are independent of $N_1(t)$. Since $N_2(t)$ is determined by the sequence of arrivals from queue 1 prior to time t and the independent sequence of service times, it then follows that $N_1(t)$ and $N_2(t)$ are independent. Equation (12.139) then follows. Note that Burke's theorem does not state that $N_1(t)$ and $N_2(t)$ are independent random processes. This would require that $N_1(t_1)$ and $N_2(t_2)$ be independent random variables for all t_1 and t_2 . This is clearly not the case.

Burke's theorem implies that the generalization of Eq. (12.139) holds for the tandem combination of any number of M/M/1, M/M/c, or M/M/ ∞ queues. Indeed, the result holds for any “feedforward” network of queues in which a customer cannot visit any queue more than once.

Example 12.21

Find the joint state pmf for the network of queues shown in Fig. 12.23, where queue 1 is driven by a Poisson process of rate λ_1 , where the departures from queue 1 are randomly routed to queues 2 and 3, and where queue 3 also has an additional independent Poisson arrival stream of rate λ_2 .

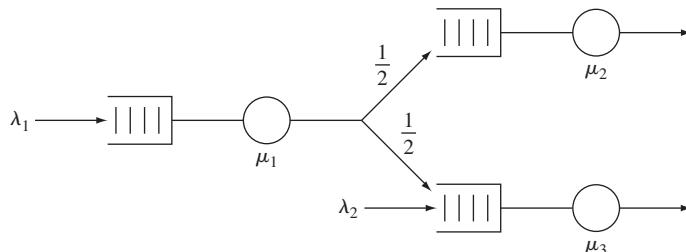


FIGURE 12.23
A feedforward network of queues.

From Burke's theorem $N_1(t)$ and $N_2(t)$ are independent, as are $N_1(t)$ and $N_3(t)$. Since the random split of a Poisson process yields independent Poisson processes, we have that the inputs to queues 2 and 3 are independent. The input to queue 2 is Poisson with rate $\lambda_1/2$. The input to queue 3 is Poisson of rate $\lambda_1/2 + \lambda_2$ since the merge of two independent Poisson processes is also Poisson. Thus

$$\begin{aligned} P[N_1(t) = k, N_2(t) = m, N_3(t) = n] \\ = (1 - \rho_1)\rho_1^k(1 - \rho_2)\rho_2^m(1 - \rho_3)\rho_3^n \quad k, m, n \geq 0, \end{aligned}$$

where $\rho_1 = \lambda_1/\mu_1$, $\rho_2 = \lambda_1/2\mu_2$, and $\rho_3 = (\lambda_1/2 + \lambda_2)/\mu_3$, and where we have assumed that all of the queues are stable.

*12.8.1 Proof of Burke's Theorem Using Time Reversibility

Consider the sample path of an M/M/1, M/M/c, or M/M/ ∞ system as shown in Fig. 12.24(a). Note that the arrivals in the forward process correspond to the departures in the time-reversed process. In Section 11.5, we showed that birth-and-death Markov chains in steady state are time-reversible processes; that is, the sample functions of the process played backward in time have the same statistics as the forward process. Since M/M/1, M/M/c, and M/M/ ∞ systems are birth-and-death Markov chains, we

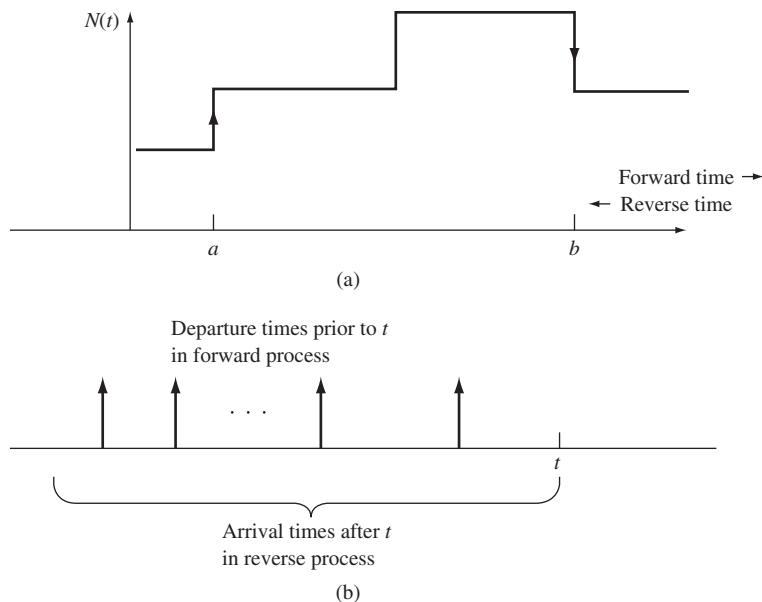


FIGURE 12.24

(a) Time instant a is an arrival time in the forward process and a departure time in the reverse process. Time instant b is a departure in the forward process and an arrival in the reverse process. (b) The departure times prior to time t in the forward process correspond exactly to the arrival times after time t in the reverse process.

This is sometimes called the *Erlang-C* formula, which we denote by $C(c, r)$. This formula gives the probability that an arriving customer is delayed in the queue (i.e., has positive, nonzero wait in the queue), as a function of the parameters c and r . The formula is based on the $M/M/c$ model and thus carries with it all of the corresponding assumptions. In particular, the model ignores complexities such as abandonments, retrials, and nonstationary arrivals – factors that may be important in modeling call centers. Also, the model assumes an infinite queue size. For call centers, this corresponds to an infinite number of available access lines into the center.

■ EXAMPLE 2.3

Calls to a technical support center arrive according to a Poisson process with rate 30 per hour. The time for a support person to serve one customer is exponentially distributed with a mean of 5 minutes. The support center has 3 technical staff to assist callers. What is the probability that a customer is able to immediately access a support staff, without being delayed on hold? (Assume that customers do not abandon their calls.)

For this problem, $\lambda = 30$, $\mu = 12$, and $c = 3$. Then $r = 2.5$ and $\rho = 5/6$. From (2.38),

$$C(c, r) = \frac{2.5^3}{3!(1 - 5/6)} \Big/ \left(\frac{2.5^3}{3!(1 - 5/6)} + 1 + \frac{2.5}{1!} + \frac{2.5^2}{2!} \right) \doteq .702.$$

Since $C(c, r)$ represents the probability of positive delay, the probability of no delay is .298.

Now suppose that the call center wishes to increase the probability of non-delayed calls to 90%. How many servers are needed? To answer this, we incrementally increase c until $1 - C(c, r) \geq .90$. It is found that $c = 6$ servers is the minimum number of servers that satisfies this requirement.

We now obtain the complete probability distributions of the waiting times, $W(t)$ and $W_q(t)$, in a manner similar to that of Section 2.2.4. [Note that $W(t)$ and $W_q(t)$ were not needed to obtain the average values W and W_q .] For $T_q > 0$ and assuming FCFS,

$$\begin{aligned} W_q(t) &= \Pr\{T_q \leq t\} = W_q(0) \\ &+ \sum_{n=c}^{\infty} \Pr\{n - c + 1 \text{ completions in } \leq t \mid \text{arrival found } n \text{ in system}\} \cdot p_n. \end{aligned}$$

Now when $n \geq c$, the system output is Poisson with mean rate $c\mu$, so that the time between successive completions is exponential with mean $1/c\mu$, and the distribution

c is reached. This method avoids numerical overflow that may be encountered with direct application of (2.53).

Although the Erlang-B formula applies to the $M/M/c/c$ (or $M/G/c/c$) queue, it can also be used in the computation of congestion measures for the $M/M/c$ queue. For example, let $C(c, r) = 1 - W_q(0)$ be the Erlang-C probability of delay in an $M/M/c$ queue (2.38). Then, $C(c, r)$ can be written as a function of $B(c, r)$ as follows (Problem 2.38):

$$C(c, r) = \frac{cB(c, r)}{c - r + rB(c, r)}. \quad (2.55)$$

Thus, to calculate $C(c, r)$, one can first calculate $B(c, r)$ iteratively using (2.54) and then apply (2.55) to get $C(c, r)$. Furthermore, $C(c, r)$ can be used in the computation of L_q , L , W_q , and W for the $M/M/c$ queue. For example, (2.33) can be rewritten:

$$L_q = C(c, r) \frac{\rho}{1 - \rho} = C(c, r) \frac{r}{c - r}. \quad (2.56)$$

Similarly, W_q , W , and L (2.34)–(2.36) can all be expressed in terms of $C(c, r)$.

■ EXAMPLE 2.7

With $\lambda = 6$, $\mu = 3$, and $c = 4$, calculate the fraction of customers blocked for an $M/M/c/c$ queue; calculate $1 - W_q(0)$ and L_q for an $M/M/c$ queue. The offered load is $r = 6/3 = 2$ and the traffic intensity is $\rho = 1/2$. The fraction of customers blocked for an $M/M/c/c$ queue is $B(4, 2)$. This can be calculated iteratively using (2.54):

$$\begin{aligned} B(0, 2) &= 1 \\ B(1, 2) &= 2 \cdot B(0, 2)/(1 + 2 \cdot B(0, 2)) = 2/3, \\ B(2, 2) &= 2 \cdot (2/3)/(2 + 2 \cdot (2/3)) = 2/5, \\ B(3, 2) &= 2 \cdot (2/5)/(3 + 2 \cdot (2/5)) = 4/19, \\ B(4, 2) &= 2 \cdot (4/19)/(4 + 2 \cdot (4/19)) = 2/21. \end{aligned}$$

Alternatively, $B(4, 2)$ can be calculated from (2.53):

$$\begin{aligned} B(4, 2) &= \frac{2^4/4!}{1 + 2 + 2^2/2! + 2^3/3! + 2^4/4!} \\ &= \frac{16/24}{(24 + 48 + 48 + 32 + 16)/24} = \frac{2}{21}. \end{aligned}$$

The probability of delay $1 - W_q(0)$ for an $M/M/c$ queue is $C(4, 2)$, which can be calculated using (2.55):

$$C(4, 2) = 4 \cdot (2/21)/(4 - 2 + 2 \cdot (2/21)) = 4/23.$$

The average number in queue L_q can be calculated using (2.56):

$$L_q = C(4, 2) \cdot 2/(4 - 2) = 4/23.$$

Alternatively, $1 - W_q(0)$ and L_q can be calculated using (2.32), (2.33), and (2.38).

$$\frac{1}{p_0} = 1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!(1-0.5)} = \frac{92}{12} = \frac{23}{3},$$

$$1 - W_q(0) = \frac{2^4}{4!(0.5)} \cdot \frac{3}{23} = \frac{4}{23}, \quad L_q = \frac{2^4 \cdot 0.5}{4!(0.5)^2} \cdot \frac{3}{23} = \frac{16}{12} \cdot \frac{3}{23} = \frac{4}{23}.$$

2.7 Queues with Unlimited Service ($M/M/\infty$)

We now treat a queueing model for which there is unlimited service, that is, an infinite number of servers available. This model is often referred to as the ample-server problem. A self-service situation is a good example of the use of such a model.

We make use of the general birth-death results with $\lambda_n = \lambda$ and $\mu_n = n\mu$, for all n , which yields

$$p_n = \frac{r^n}{n!} p_0, \quad p_0 = \left(\sum_{n=0}^{\infty} \frac{r^n}{n!} \right)^{-1}.$$

The infinite series in the expression for p_0 is equal to e^r . Therefore

$$p_n = \frac{r^n e^{-r}}{n!} \quad (n \geq 0),$$

(2.57)

which is a Poisson distribution with mean $r = \lambda/\mu$. The value of λ/μ is not restricted in any way for the existence of a steady-state solution. It also turns out (we show this in Section 5.2.3) that (2.57) is valid for *any* $M/G/\infty$ model. That is, p_n depends only on the mean service time and not on the form of the service-time distribution. It is not surprising that this is true here in light of a similar result we mentioned previously for $M/M/c/c$, since p_n of (2.57) could have been obtained from (2.52) by taking the limit as $c \rightarrow \infty$.

The expected system size is the mean of the Poisson distribution of (2.57) and is thus found as $L = r = \lambda/\mu$. Since we have as many servers as customers in the system, $L_q = 0 = W_q$. The average waiting time in the system is merely the average service time, so that $W = 1/\mu$, and the waiting-time distribution function $W(t)$ is identical to the service-time distribution, namely, exponential with mean $1/\mu$.

■ EXAMPLE 2.8

Television station KCAD in a large western metropolitan area wishes to know the average number of viewers it can expect on a Saturday evening prime-time program. It has found from past surveys that people turning on their television sets on Saturday evening during prime time can be described rather well by a Poisson distribution with a mean of 100,000/h. There are five major TV stations in the area, and it is believed that a given person chooses among these

when the capacity constraint is removed (i.e., K goes to ∞). Thus it follows for $M/M/c/\infty$ that $q_n = p_n$.

Finally, to get the CDF $W_q(t)$ for the line delays, we note, in a fashion similar to the derivation leading to (2.39), that

$$W_q(t) = \Pr\{T_q \leq t\} = W_q(0) + \sum_{n=c}^{K-1} \Pr\{n - c + 1 \text{ completions in } \leq t | \text{arrival found } n \text{ in system}\} \cdot q_n,$$

since there cannot be arrivals joining the system whenever they encounter K customers. It follows that

$$\begin{aligned} W_q(t) &= W_q(0) + \sum_{n=c}^{K-1} q_n \int_0^t \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx \\ &= W_q(0) + \sum_{n=c}^{K-1} q_n \left(1 - \int_t^\infty \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx \right). \end{aligned}$$

For the simplification of (1.15), Section 1.7, we have shown that

$$\int_t^\infty \frac{\lambda(\lambda x)^m}{m!} e^{-\lambda x} dx = \sum_{i=0}^m \frac{(\lambda t)^i e^{-\lambda t}}{i!}.$$

Letting $m = n - c$ and $\lambda = c\mu$ gives

$$\int_t^\infty \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx = \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!}$$

and hence

$$\begin{aligned} W_q(t) &= W_q(0) + \sum_{n=c}^{K-1} q_n - \sum_{n=c}^{K-1} q_n \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!} \\ &= 1 - \sum_{n=c}^{K-1} q_n \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!} \end{aligned}$$

■ EXAMPLE 2.6

Consider an automobile emission inspection station with three inspection stalls, each with room for only one car. It is reasonable to assume that cars wait in such a way that when a stall becomes vacant, the car at the head of the line pulls up to it. The station can accommodate at most four cars waiting (seven in the station) at one time. The arrival pattern is Poisson with a mean of one car every minute during the peak periods. The service time is exponential with mean

6 min. I. M. Fussy, the chief inspector, wishes to know the average number in the system during peak periods, the average wait (including service), and the expected number per hour that cannot enter the station because of full capacity.

Using minutes as the basic time unit, $\lambda = 1$ and $\mu = \frac{1}{6}$. Thus we have $r = 6$ and $\rho = 2$ for this $M/M/3/7$ system. We first calculate p_0 from (2.46) and find that

$$\begin{aligned} p_0 &= \left(\sum_{n=0}^2 \frac{6^n}{n!} + \frac{6^3}{3!} \frac{1 - 2^5}{1 - 2} \right)^{-1} \\ &= \frac{1}{1141} \doteq 0.00088. \end{aligned}$$

From (2.47) we then get

$$L_q = \frac{p_0(6^3)(2)}{3!} [1 - 2^5 + 5(2^4)] = \frac{3528}{1141} \doteq 3.09 \text{ cars.}$$

Then $L = L_q + r(1 - p_K)$, so

$$L = \frac{3528}{1141} + 6 \left(1 - \frac{6^7}{(3^4)(3!)(1141)} \right) = \frac{9606}{1141} \doteq 6.06 \text{ cars.}$$

To find the average wait during peak periods, (2.48) gives that

$$W = \frac{L}{\lambda_{\text{eff}}} = \frac{L}{\lambda(1 - p_7)} = \frac{L}{1 - p_0 6^7 / (3^4 3!)} \doteq 12.3 \text{ min.}$$

The expected number of cars per hour that cannot enter the station is given by

$$60\lambda p_k = 60p_7 = \frac{60p_0 6^7}{3^4 3!} \doteq 30.4 \text{ cars/h.}$$

This might suggest an alternative setup for the inspection station.

2.6 Erlang's Loss Formula ($M/M/c/c$)

The special case of the truncated queue $M/M/c/K$ with $K = c$, that is, where no line is allowed to form, gives rise to a stationary distribution known as Erlang's first formula. This stationary distribution can be obtained from (2.45) and (2.46) with $K = c$ as

$$p_n = \frac{(\lambda/\mu)^n}{n!} \Bigg/ \left(\sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} \right), \quad (0 \leq n \leq c). \quad (2.52)$$

When $n = c$, the resultant formula for p_c is called *Erlang's loss formula* or the *Erlang-B formula*. This is the probability of a full system at any time in steady state.

and the rest of the analysis is identical.

There is an important *invariance* result for finite-source queues, similar in importance to the fact that $M/G/c/c$ steady-state probabilities are independent of the form of G . It is that (2.58) is valid for any finite-source system with exponential service, independent of the nature of the distribution of time to breakdown, as long as the lifetimes are independent with mean $1/\lambda$. The interested reader is referred to Bunday and Scraton (1980) for details of the proof. Furthermore, the $M/G/c/c$ -type result also holds in that if the number of repair technicians equal the number of machines, the repair distribution can be G , as long as the failure times are exponential.

■ EXAMPLE 2.9

The Train SemiConductor Company uses five robots in the manufacture of its circuit boards. The robots break down periodically, and the company has two repair people to do service when robots fail. When one is fixed, the time until the next breakdown is thought to be exponentially distributed with a mean of 30 h. The shop always has enough of a work backlog to ensure that all robots in operating condition will be working. The repair time for each service is thought to be exponentially distributed with a mean of 3 h. The shop manager wishes to know the average number of robots operational at any given time, the expected downtime of a robot that requires repair, and the expected percentage of idle time of each repairer.

To answer any of these questions, we must first calculate p_0 . In this example, $M = 5$, $c = 2$, $\lambda = \frac{1}{30}$, and $\mu = \frac{1}{3}$, and thus $r = \lambda/\mu = \frac{1}{10}$. We use (2.58) and obtain the five $\{a_n\}$ multipliers as $a_1 = 5/10 = \frac{1}{2}$; $a_2 = \frac{1}{10}$; $a_3 = 15/1000 = \frac{3}{200}$; $a_4 = 15/10,000 = \frac{3}{2000}$; and $a_5 = 15/200,000 = \frac{3}{40000}$. It thus follows that

$$p_0 = (1 + \frac{1}{2} + \frac{1}{10} + \frac{3}{200} + \frac{3}{2000} + \frac{3}{40000})^{-1} = \frac{40000}{64663} \doteq 0.619.$$

The average number of operational robots is $M - L$, where

$$L = p_0(1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{10} + 3 \cdot \frac{3}{200} + 4 \cdot \frac{3}{2000} + 5 \cdot \frac{3}{40000}) = \frac{30055}{64663} \doteq 0.465.$$

Thus $5 - 0.465$, or 4.535, robots are in operating condition on average. The expected downtime can be found from (2.61) and is

$$W = \frac{\frac{30055}{64663}}{\frac{1}{30}(5 - \frac{30055}{64663})} = \frac{901650}{29326} \doteq 3.075 \text{ h.}$$

The average fraction of idle time of each server is

$$p_0 + \frac{1}{2}p_1 = p_0(1 + \frac{1}{2}a_1) = p_0(1 + \frac{1}{4}) = \frac{50000}{64663} \doteq 0.773,$$

so each repair person is idle approximately 77% of the time.

The manager, because of the long idle time, is interested in knowing the answer to the same questions if the repair force is reduced to one person. The

results are

$$\begin{aligned} p_0 &\doteq 0.564, \quad L \doteq 0.640, \\ M - L &\doteq 4.360, \quad W \doteq 4.400 \text{ h}. \end{aligned}$$

Since over four robots are expected operational at any time under both situations and the increase in downtime is about one hour with only one repair person, the manager might well decide to move one of them to work elsewhere.

2.8.0.1 Models with Spares. The finite-source model can be generalized to include the use of spares. We assume now that there are M machines in operation plus an additional Y spares. When a machine in operation fails, a spare is immediately substituted for it (if available). If no spare is available when the failure occurs, then the system becomes short. Once a machine is repaired, it becomes a spare, unless the system is short, in which case the repaired machine goes immediately into service. At any given time, there are at most M machines in operation, so the rate of failures is at most $M\lambda$ (i.e., spares that are not in operation do not contribute to the failure rate). For this model, λ_n is different from before and is given by

$$\lambda_n = \begin{cases} M\lambda & (0 \leq n < Y), \\ (M - n + Y)\lambda & (Y \leq n < Y + M), \\ 0 & (n \geq Y + M), \end{cases}$$

where n represents the number of failed machines. Again considering c technicians, we have

$$\mu_n = \begin{cases} n\mu & (0 \leq n < c), \\ c\mu & (n \geq c). \end{cases}$$

We first assume $c \leq Y$ and use (2.3) to find that (with $r = \lambda/\mu$)

$$p_n = \begin{cases} \frac{M^n}{n!} r^n p_0 & (0 \leq n < c), \\ \frac{M^n}{c^{n-c} c!} r^n p_0 & (c \leq n < Y), \\ \frac{M^Y M!}{(M - n + Y)! c^{n-c} c!} r^n p_0 & (Y \leq n \leq Y + M). \end{cases} \quad (2.62)$$

If Y is very large, we essentially have an infinite calling population with mean arrival rate $M\lambda$. Letting Y go to infinity in (2.62) yields the $M/M/c/\infty$ results of (2.31) with $M\lambda$ for λ .

convolution of (a) an exponential distribution with mean $1/(c\mu - \lambda)$ and (b) an exponential distribution with mean $1/\mu$. The first distribution is the conditional distribution of T_q given that $T_q > 0$ (see earlier in this section). This convolution can also be written as the difference of the two exponential functions (see Problem 2.19),

$$\Pr\{T \leq t\} = \frac{c(1-\rho)}{c(1-\rho)-1}(1-e^{-\mu t}) - \frac{1}{c(1-\rho)-1}(1-e^{-(c\mu-\lambda)t}).$$

Thus the overall CDF of the $M/M/c$ system waits may be written as

$$\begin{aligned} W(t) &= W_q(0)[1 - e^{-\mu t}] + [1 - W_q(0)] \\ &\quad \times \left(\frac{c(1-\rho)}{c(1-\rho)-1}(1-e^{-\mu t}) - \frac{1}{c(1-\rho)-1}(1-e^{-(c\mu-\lambda)t}) \right) \\ &= \frac{c(1-\rho) - W_q(0)}{c(1-\rho)-1}(1-e^{-\mu t}) - \frac{1 - W_q(0)}{c(1-\rho)-1}(1-e^{-(c\mu-\lambda)t}). \end{aligned}$$

We now illustrate these developments with an example.

■ EXAMPLE 2.4

City Hospital's eye clinic offers free vision tests every Wednesday evening. There are three ophthalmologists on duty. A test takes, on average, 20 min, and the actual time is found to be approximately exponentially distributed around this average. Clients arrive according to a Poisson process with a mean of 6/h, and patients are taken on a first-come, first-served basis. The hospital planners are interested in knowing (1) the average number of people waiting; (2) the average amount of time a patient spends at the clinic; and (3) the average percentage idle time of each of the doctors. Thus we wish to calculate L_q , W , and the percentage idle time of a server.

We begin by calculating p_0 , since this factor appears in all the formulas derived for the measures of effectiveness. We have that $c = 3$, $\lambda = 6/h$, and $\mu = 1/(20 \text{ min}) = 3/h$. Thus $r = \lambda/\mu = 2$, $\rho = \frac{2}{3}$, and, from (2.32),

$$p_0 = \left(1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!(1-\frac{2}{3})} \right)^{-1} = \frac{1}{9}.$$

From (2.33), we find that

$$L_q = \left(\frac{(2^3)(\frac{2}{3})}{3!(1-\frac{2}{3})^2} \right) \left(\frac{1}{9} \right) = \frac{8}{9},$$

and from (2.33) and (2.35) that

$$W = \frac{1}{\mu} + \frac{L_q}{\lambda} = \frac{1}{3} + \frac{\frac{8}{9}}{6} = \frac{13}{27} \text{ h} \doteq 28.9 \text{ min}.$$

Next, we have already shown (see Table 1.2) that the long-term average fraction of idle time for any server in an $M/M/c$ is equal to $1 - \rho$. For this problem,

therefore, each physician is idle $\frac{1}{3}$ of the time, since the traffic intensity is $\rho = \frac{2}{3}$. Given the three servers on duty, two of them will be busy at any time (on average), since $r = 2$. Furthermore, the fraction of time that there is at least one idle doctor can be computed here as $p_0 + p_1 + p_2 = \Pr\{T_q = 0\} = \frac{5}{9}$.

2.4 Choosing the Number of Servers

In managing a queueing system, it is often desirable to determine an appropriate number of servers c for the system. A larger number of servers improves quality of service to the customers but incurs a higher cost to the queue owner. The problem is to find the number of servers that adequately balances the quality and cost of service. This section gives a simple approximation that can be helpful in choosing the number of servers for an $M/M/c$ queue. The approximation works for queues with a large number of servers.

Before discussing the approximation, we first observe that in steady state the number of servers must be greater than the offered load r . Otherwise, the queue is unstable. Thus we can write

$$c = r + \Delta,$$

where $\Delta > 0$ is the number of additional servers used in excess of the offered load (Δ may need to be a fraction in order to make c an integer). Thus the problem of choosing the number of servers c is equivalent to choosing the number of servers Δ in excess of the offered load.

To motivate ideas for choosing c (or Δ), consider an $M/M/c$ queue with offered load $r = 9$, $c = 12$ servers, and traffic intensity $\rho = 0.75$. Suppose that the owner of the queue has observed the system for a long time and is satisfied with its overall performance, considering both the congestion experienced by the customers and the cost of paying the servers.

Now suppose that the offered load quadruples to $r = 36$. How many new servers should the owner hire? There are several lines of reasoning that can be taken to answer this question.

1. Choose c to maintain (approximately) a constant traffic intensity ρ . In the baseline case, there are 4 servers available for every 3 customers in service, on average ($\rho = 0.75$). It may seem reasonable to keep this ratio constant. So, if the traffic level quadruples, the number of servers should also quadruple to $c = 48$.
2. Choose c to maintain (approximately) a constant measure of congestion. An example congestion measure is $1 - W_q(0)$, the probability that a customer is delayed in the queue, which can be obtained from the Erlang-C formula (2.38). In the baseline case, this turns out to be $1 - W_q(0) = 0.266$. If the traffic quadruples to $r = 36$, then the minimum number of servers needed to maintain (or improve) this level of service is $c = 42$. More generally, if α is the maximum desired fraction of callers delayed in the queue, c is chosen as

since we are in steady state, so that the input rate must equal the output rate; but it is not quite so intuitive that the variances and indeed the distributions are identical. Nevertheless, it is true and proves extremely useful in analyzing series queues, where the initial input rate is Poisson, the service at all stations is exponential, and there is no restriction on queue size between stations.

We now illustrate a series queueing situation of the type above with an example.

■ EXAMPLE 4.1

Cary Meback, the president of a large Virginia supermarket chain, is experimenting with a new store design and has remodeled one of his stores as follows. Instead of the usual checkout-counter design, the store has been remodeled to include a checkout “lounge.” As customers complete their shopping, they enter the lounge with their carts and, if all checkers are busy, receive a number. They then park their carts and take a seat. When a checker is free, the next number is called and the customer with that number enters the available checkout counter. The store has been enlarged so that for practical purposes, there is no limit on either the number of shoppers that can be in the food aisles or the number that can wait in the lounge, even during peak periods.

The management estimates that during peak hours customers arrive according to a Poisson process at a mean rate of 40/h and it takes a customer, on average, $\frac{3}{4}$ h to fill a shopping cart, the filling times being approximately exponentially distributed. Furthermore, the checkout times are also approximately exponentially distributed with a mean of 4 min, regardless of the particular checkout counter (during peak periods each counter has a cashier and bagger, hence the low mean checkout time). Meback wishes to know the following: (1) What is the minimum number of checkout counters required in operation during peak periods? (2) If it is decided to add one more than the minimum number of counters required in operation, what is the average waiting time in the lounge? How many people, on average, will be in the lounge? How many people, on average, will be in the entire supermarket?

This situation can be modeled by a two-station series queue. The first is the food portion of the supermarket. Since it is self-service and arrivals are Poisson, we have an $M/M/\infty$ model with $\lambda = 40$ and $\mu = \frac{4}{3}$. The second station is an $M/M/c$ model, since the output of an $M/M/\infty$ queue is identical to its input. Hence the input to the checkout lounge is Poisson with a mean of 40/h also. Since $c\mu > \lambda$ for steady-state convergence, the minimum number of checkout counters, c_m , must be greater than $\lambda/\mu = 40/15 \doteq 2.67$; hence c_m must be 3.

If it is decided to have four counters in operation, we have an $M/M/4$ model at the checkout stations with $\lambda = 40$ and $\mu = 15$. Meback desires to know W_q and L_q for the $M/M/4$ model, as well as the average total number in the supermarket, which is the sum of the L 's for both models. Using (2.32)

and (2.34) for an $M/M/c$ model, we get

$$p_0 = \left[\sum_{n=0}^3 \frac{1}{n!} \left(\frac{8}{3} \right)^n + \frac{1}{4!} \left(\frac{8}{3} \right)^4 \left(\frac{4}{4 - \frac{8}{3}} \right) \right]^{-1} \doteq 0.06$$

and

$$W_q = \frac{\left(\frac{8}{3}\right)^4 15}{3!(60-40)^2} (0.06) \doteq 0.019 \text{ h} \doteq 1.14 \text{ min.}$$

To get L_q we use Little's formula and find

$$L_q = \lambda W_q \doteq 40(0.019) = 0.76,$$

so that, on average, less than one person will be waiting in the lounge for a checker to become free.

The total number of people in the system, on average, is the L for this $M/M/4$ model plus the L for the $M/M/\infty$ model. For the checkout station we get

$$L = \lambda W = \lambda \left(W_q + \frac{1}{\mu} \right) \doteq 40 \left(0.019 + \frac{4}{60} \right) \doteq 3.44.$$

For the supermarket proper we have, from the $M/M/\infty$ model results, that $L = \lambda/\mu = 40/(\frac{4}{3}) = 30$. Hence the average number of customers in the store during peak hours is 33.44 if Meback decides on four checkout counters in operation. He might do well to perform similar calculations for three checkout counters operating to see how much the congestion increases (see Problem 4.3).

For series queues, therefore, as long as there are no capacity limitations between stations and the input is Poisson, results can be rather easily obtained. Furthermore, it can be shown (see Problem 4.4) that the joint probability that there are n_1 at station 1, n_2 at station 2, ..., and n_j at station j is merely the product $p_{n_1} p_{n_2} \cdots p_{n_j}$. This product-form type of result is quite typical of those available for Jackson networks, as we shall see in subsequent sections.

The analysis for series queues when there are limits on the capacity at a station (except for the case where the only limit is at the last station in a pure series flow situation and arriving customers who exceed the capacity are shunted out of the system—see Problem 4.5) is much more complex. This results from the blocking effect; that is, a station downstream comes up to capacity and thereby prevents any further processing at upstream stations that feed it. We treat some of these types of models in the next section.

4.1.2 Series Queues with Blocking

We consider first a simple sequential two-station, single-server-at-each-station model, where no queue is allowed to form at either station. If a customer is in station 2 and

and (2.34) for an $M/M/c$ model, we get

$$p_0 = \left[\sum_{n=0}^3 \frac{1}{n!} \left(\frac{8}{3} \right)^n + \frac{1}{4!} \left(\frac{8}{3} \right)^4 \left(\frac{4}{4 - \frac{8}{3}} \right) \right]^{-1} \doteq 0.06$$

and

$$W_q = \frac{\left(\frac{8}{3}\right)^4 15}{3!(60-40)^2} (0.06) \doteq 0.019 \text{ h} \doteq 1.14 \text{ min.}$$

To get L_q we use Little's formula and find

$$L_q = \lambda W_q \doteq 40(0.019) = 0.76,$$

so that, on average, less than one person will be waiting in the lounge for a checker to become free.

The total number of people in the system, on average, is the L for this $M/M/4$ model plus the L for the $M/M/\infty$ model. For the checkout station we get

$$L = \lambda W = \lambda \left(W_q + \frac{1}{\mu} \right) \doteq 40 \left(0.019 + \frac{4}{60} \right) \doteq 3.44.$$

For the supermarket proper we have, from the $M/M/\infty$ model results, that $L = \lambda/\mu = 40/(\frac{4}{3}) = 30$. Hence the average number of customers in the store during peak hours is 33.44 if Meback decides on four checkout counters in operation. He might do well to perform similar calculations for three checkout counters operating to see how much the congestion increases (see Problem 4.3).

For series queues, therefore, as long as there are no capacity limitations between stations and the input is Poisson, results can be rather easily obtained. Furthermore, it can be shown (see Problem 4.4) that the joint probability that there are n_1 at station 1, n_2 at station 2, ..., and n_j at station j is merely the product $p_{n_1} p_{n_2} \cdots p_{n_j}$. This product-form type of result is quite typical of those available for Jackson networks, as we shall see in subsequent sections.

The analysis for series queues when there are limits on the capacity at a station (except for the case where the only limit is at the last station in a pure series flow situation and arriving customers who exceed the capacity are shunted out of the system—see Problem 4.5) is much more complex. This results from the blocking effect; that is, a station downstream comes up to capacity and thereby prevents any further processing at upstream stations that feed it. We treat some of these types of models in the next section.

4.1.2 Series Queues with Blocking

We consider first a simple sequential two-station, single-server-at-each-station model, where no queue is allowed to form at either station. If a customer is in station 2 and

Table 4.1 Possible System States

n_1, n_2	Description
0,0	System empty
1,0	Customer in process at 1 only
0,1	Customer in process at 2 only
1,1	Customers in process at 1 and 2
$b,1$	Customer in process at 2 and a customer finished at 1 but waiting for 2 to become available (i.e., system is blocked)

service is completed at station 1, the station-1 customers must wait there until the station-2 customer is completed; that is, the system is *blocked*. Arrivals at station 1 when the system is blocked are turned away. Also, if a customer is in process at station 1, then even if station 2 is empty, arriving customers are turned away, since the system is a sequential one; that is, all customers require service at 1 and then service at 2.

We wish to find the steady-state probability p_{n_1, n_2} of n_1 in the first station and n_2 in the second station. For this model, the possible states are given in Table 4.1.

Assuming arrivals at the system (station 1) are Poisson with parameter λ and service is exponential with parameters μ_1 and μ_2 , respectively, the usual procedure leads to the steady-state equations for this multidimensional Markov chain:

$$\begin{aligned} 0 &= -\lambda p_{0,0} + \mu_2 p_{0,1}, \\ 0 &= -\mu_1 p_{1,0} + \mu_2 p_{1,1} + \lambda p_{0,0}, \\ 0 &= -(\lambda + \mu_2)p_{0,1} + \mu_1 p_{1,0} + \mu_2 p_{b,1}, \\ 0 &= -(\mu_1 + \mu_2)p_{1,1} + \lambda p_{0,1}, \\ 0 &= -\mu_2 p_{b,1} + \mu_1 p_{1,1}. \end{aligned} \quad (4.7)$$

Using the boundary equation $\sum \sum p_{n_1, n_2} = 1$, we have six equations in five unknowns [there is some redundancy in (4.7); hence we can solve for the five steady-state probabilities]. Equation (4.7) can be used to get all probabilities in terms of $p_{0,0}$, and the boundary condition can be used to find $p_{0,0}$. If we let $\mu_1 = \mu_2$, the results are (see Problem 4.6)

$$\begin{aligned} p_{1,0} &= \frac{\lambda(\lambda + 2\mu)}{2\mu^2} p_{0,0}, & p_{0,1} &= \frac{\lambda}{\mu} p_{0,0}, & p_{1,1} &= \frac{\lambda^2}{2\mu^2} p_{0,0}, \\ p_{b,1} &= \frac{\lambda^2}{2\mu^2} p_{0,0}, & p_{0,0} &= \frac{2\mu^2}{3\lambda^2 + 4\mu\lambda + 2\mu^2}. \end{aligned} \quad (4.8)$$

It is easy to see how the problem expands if one allows limits other than zero on queue length or considers more stations. For example, if one customer is allowed to wait between stations, this results in seven state probabilities for which to solve,

To get the average length of the busy period, $E[T_{bp}]$, we simply find the value of the negative of the derivative of the transform of $p'_0(t)$, $s\bar{p}_0(s)$, evaluated at $s = 0$. But an attractive alternative way to find the mean length of the busy period is to use the simple steady-state ratio argument that

$$\frac{1 - p_0}{p_0} = \frac{E[T_{bp}]}{E[T_{idle}]} = \frac{E[T_{bp}]}{1/\lambda}.$$

Since $p_0 = 1 - \lambda/\mu$, it follows that the expected lengths of the busy period and busy cycle, respectively, are

$$E[T_{bp}] = \frac{1}{\mu - \lambda} \quad \text{and} \quad E[T_{bc}] = \frac{1}{\lambda} + \frac{1}{\mu - \lambda}. \quad (2.79)$$

Equation (2.79) holds for all $M/G/1$ -type queues, since the exponential service property played no role in the derivation.

It is not too difficult to extend the notion of the busy period conceptually to the multichannel case. Recall that for one channel a busy period is defined to begin with the arrival of a customer at an idle channel and to end when the channel next becomes idle. In an analogous fashion, let us define an i -channel busy period for $M/M/c$ ($0 \leq i \leq c$) to begin with an arrival at the system at an instant when there are $i - 1$ in the system and to end at the very next point in time when the system size dips to $i - 1$. Let us say that the case where $i = 1$ (an arrival to an empty system) defines the system busy period. In fashion similar to that for $M/M/1$, use $T_{b,i}$ to denote the random variable “length of the i -channel busy period.” Then the CDF of $T_{b,i}$ is determined by considering the original $M/M/c$ differential-difference equations of (2.72) with an absorbing barrier imposed at a system size of $i - 1$ and an initial size of i . Then it should be clear that $p_{i-1}(t)$ will, in fact, be the required CDF, and its derivative the density. The necessary equations are

$$\begin{aligned} p'_{i-1}(t) &= i\mu p_i(t) \quad [\text{because of absorbing barrier}], \\ p'_i(t) &= -(\lambda + i\mu)p_i(t) + (i+1)\mu p_{i+1}(t) \quad [\text{because of absorbing barrier}], \\ p'_n(t) &= -(\lambda + n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t) \quad (i < n < c), \\ p'_n(t) &= -(\lambda + c\mu)p_n(t) + \lambda p_{n-1}(t) + c\mu p_{n+1}(t) \quad (n \geq c). \end{aligned}$$

Proceeding further gets us bogged down in great algebraic detail. Any resultant CDF will be in terms of modified Bessel functions, but with enough time and patience, $p'_{i-1}(t)$, $\bar{p}_{i-1}(s)$, and $E[T_{b,i}]$ can be obtained.

PROBLEMS

- 2.1.** You are told that a small single-server, birth-death-type queue with finite capacity cannot hold more than three customers. The three arrival or birth rates are $(\lambda_0, \lambda_1, \lambda_2) = (3, 2, 1)$, while the service or death rates are $(\mu_1, \mu_2, \mu_3) = (1, 2, 2)$. Find the steady-state probabilities $\{p_i, i =$

$0, 1, 2, 3\}$ and L . Then determine the average or effective arrival rate $\lambda_{\text{eff}} = \sum \lambda_i p_i$, and the expected system waiting time W .

- 2.2.** The finite-capacity constraint of Problem 2.1 has been pushed up to 10 now, with the arrival rates known to be $(4, 3, 2, 2, 3, 1, 2, 1, 2, 1)$ and service rates to be $(1, 1, 1, 2, 2, 2, 3, 3, 3, 4)$. Do as before and find p_i , $i = 0, 1, 2, \dots, 10$, the mean system size L , the effective arrival rate $\lambda_{\text{eff}} = \sum \lambda_i p_i$, and the expected system waiting time W .
- 2.3.** For an $M/M/1$ queue, derive the variance of the number of customers in the system in steady state.
- 2.4.** For the $M/M/1$ and $M/M/c$ queues, find $E[T_q | T_q > 0]$, that is, the expected time one must wait in the queue, given that one must wait at all.
- 2.5.** Derive $W(t)$ and $w(t)$ (the total-waiting-time CDF and its density) as given by the equations (2.29).
- 2.6.** Equation (1.4) is valid for any queueing system and can be used to prove Little's formula for the $G/G/1$ queue in a somewhat different manner than that used in Section 1.5.1. The approach is to plot the cumulative count of arrivals on the same graph as the cumulative count of departures. Then it can be seen that the area between these two step functions from the beginning of a busy period to the beginning of the next (a busy cycle) is the accumulated total of the system waiting times of all the customers who have entered into the system during this busy cycle. Use this argument to derive an empirical version of Little's formula over a busy cycle.
- 2.7.** What effect does doubling λ and μ have on L , L_q , W , and W_q in an $M/M/1$ model?
- 2.8.** A graduate research assistant “moonlights” in the food court in the student union in the evenings. He is the only one on duty at the counter during the hours he works. Arrivals to the counter seem to follow the Poisson distribution with mean of 10/h. Each customer is served one at a time and the service time is thought to follow an exponential distribution with a mean of 4 min. Answer the following questions.
- What is the probability of having a queue?
 - What is the average queue length?
 - What is the average time a customer spends in the system?
 - What is the probability of a customer spending more than 5 min in the queue before being waited on?
 - The graduate assistant would like to spend his idle time grading papers. If he can grade 22 papers an hour on average when working continuously, how many papers per hour can he average while working his shift?

- 2.9.** A rent-a-car maintenance facility has capabilities for routine maintenance (oil change, lubrication, minor tune-up, wash, etc.) for only one car at a time. Cars arrive there according to a Poisson process at a mean rate of three per day, and service time to perform this maintenance seems to have an exponential distribution with a mean of $\frac{7}{24}$ day. It costs the company a fixed \$375 a day to operate the facility. The company estimates loss in profit on a car of \$25/day for every day the car is tied up in the shop. The company, by changing certain procedures and hiring faster mechanics, can decrease the mean service time to $\frac{1}{4}$ day. This also increases their operating costs. Up to what value can the operating cost increase before it is no longer economically attractive to make the change?
- 2.10.** Parts arrive at a painting machine according to a Poisson process with rate λ/h . The machine can paint one part at a time. The painting process appears to be exponential with an average service time of $1/\mu h$. It costs the company about $\$C_1$ per part per hour spent in the system (i.e., being painted or waiting to be painted). The cost of owning and operating the painting machine is strictly a function of its speed. In particular, a machine that works at an average rate of μ costs $\$C_2/h$, whether or not it is always in operation. Determine the value of μ that minimizes the cost of the painting operation.
- 2.11.** Find the probability that k or more are in an $M/M/c/\infty$ system for any $k \geq c$.
- 2.12.** For the $M/M/c/\infty$ model, give an expression for p_n in terms of p_c instead of p_0 , and then derive L_q in terms of ρ and p_c .
- 2.13.**
 - Our local fast-food emporium, Burger Bliss, has yet to learn a lot about queueing theory. So it does not require that all waiting customers form a single line, and instead they make every arrival randomly choose one of *three* lines formed before each server during the weekday lunch period. But they are so traditional about managing their lines that barriers have been placed between the lines to prevent jockeying. Suppose that the overall stream of incoming customers has settled in at a constant rate of 60/h (Poisson-distributed) and that the time to complete a customer's order is well described by an exponential distribution of mean 150 seconds, independent and identically from one customer to the next. Assuming steady state, what is the average total system size?
 - The BB has now agreed that it is preferable to have one line feeding the three servers, so the barriers have been removed. What is the expected steady-state system size now?
- 2.14.** Verify the formula for the mean line delay W_q of the $M/M/c/\infty$ queue.
- 2.15.** The Outfront BBQ Rib Haven does carry out only. During peak periods, two servers are on duty. The owner notices that during these periods, the

servers are almost never idle. She estimates the percent time idle of each server to be 1%. Ideally, the percent idle time would be 10% to allow time for important breaks.

- (a) If the owner decides to add a third server during these times, how much idle time would each server have then?
- (b) Suppose that by adding the third server, the pressure on the servers is reduced, so they can work more carefully, but their service output rate is reduced by 20%. What now is the percent time each would be idle?
- (c) Suppose, instead, the owner decides to hire an aid (at a much lower salary) who servers as a gofer for the two servers, rather than hiring another full server. This allows the two servers to decrease their average service time by 20% (relative to the original service rate). What now is the percent idle time of each of the two servers?

2.16. In the spring of 2006, George Mason University's basketball team advanced to the Final Four of the NCAA tournament – only the second double-digit seed ever to do so. To celebrate the event, the campus book store ordered tee shirts. On the day of the sale, demand for shirts was steady throughout the day and fairly well described by a Poisson process with a rate of 66 per hour. There were four cash registers in operation and the average time of a transaction was 3.5 minutes. Service times were approximately exponentially distributed.

- (a) What was the average length of the line for shirts?
- (b) How long, on average, did customers wait in line to get a shirt?
- (c) What fraction of customers spent more than 30 minutes in the store to get a shirt?

2.17. You are the owner of a small book store. You have two cash registers. Customers wait in a single line to purchase books at one of the two registers. Customers arrive according to a Poisson process with rate $\lambda = 30$ per hour. The time to complete the purchase transactions for one customer follows an exponential distribution with mean 3 minutes.

- (a) Determine W , W_q , L , and L_q for this system.
- (b) Suppose that you pay the register clerks \$10 per hour and that each customer on average purchases books that give you a net \$2 profit. What is the hourly rate that you make money?
- (c) Now suppose that you offer a \$2 rebate to every customer who joins the queue and finds 4 or more people *in the queue*. Now, what is the hourly rate that you make money?

2.18. For an $M/M/2$ queue with $\lambda = 60/\text{h}$ and $\mu = 0.75/\text{min}$, calculate L , L_q , W , W_q , $\Pr\{N \geq k\}$, and $\Pr\{T_q > t\}$ for $k = 2, 4$ and $t = 0.01, 0.03 \text{ h}$.

2.19. For $M/M/c/\infty$, derive the distribution function of the system waiting time for those customers for whom $T_q > 0$, remembering that their wait is the sum of a line delay and a service time.

- 2.20.** For the $M/M/c/K$ queue, calculate L , L_q , W , W_q , p_K , $\Pr\{N \geq k\}$, and $\Pr\{T_q \geq t\}$ for $\lambda = 2/\text{min}$, $\mu = 45/\text{h}$, $c = 2$, $K = 6$, $k = 2$ and 4 , and $t = 0.01$ and 0.02 h.
- 2.21.** The office of the Deputy Inspector General for Inspection and Safety administers the Air Force Accident and Incident Investigation and Reporting Program. It has established 25 investigation teams to analyze and evaluate each accident or incident to make sure it is properly reported to accident investigation boards. Each of these teams is dispatched to the locale of the accident or incident as each requirement for such support occurs. Support is only rendered those commands that have neither the facilities nor qualified personnel to conduct such services. Each accident or incident will require a team being dispatched for a random amount of time, apparently exponential with mean of 3 weeks. Requirements for such support are received by the Deputy Inspector General's office as a Poisson process with mean rate of 347/yr. At any given time, two teams are not available due to personnel leaves, sickness, and so on. Find the expected time spent by an accident or incident in and waiting for evaluation.
- 2.22.** An organization is presently involved in the establishment of a telecommunication center so that it may provide a more rapid outgoing message capability. Overall, the center is responsible for the transmission of outgoing messages and receives and distributes incoming messages. The center manager at this time is primarily concerned with determining the number of transmitting personnel required at the new center. Outgoing message transmitters are responsible for making minor corrections to messages, assigning numbers when absent from original message forms, maintaining an index of codes and a 30-day file of outgoing messages, and actually transmitting the messages. It has been predetermined that this process is exponential and requires a mean time of 28 min/message. Transmission personnel will operate at the center 7 h/day, 5 days/week. All outgoing messages will be processed in the order they are received and follow a Poisson process with a mean rate of 21 per 7-h day. Processing on messages requiring transmission must be started within an average of 2 h from the time they arrive at the center. Determine the minimum number of transmitting personnel to accomplish this service criterion. If the service criterion were to require the probability of any message waiting for the start of processing for more than 3 h to be less than .05, how many transmitting personnel would be required?
- 2.23.** A small branch bank has two tellers, one for receipts and one for withdrawals. Customers arrive to each teller's cage according to a Poisson distribution with a mean of 20/h. (The total mean arrival rate at the bank is 40/h.) The service time of each teller is exponential with a mean of 2 min. The bank manager is considering changing the setup to allow each teller to handle both withdrawals and deposits to avoid the situations that arise from time to time when the queue is sizable in front of one teller while the other

is idle. However, since the tellers would have to handle both receipts and withdrawals, their efficiency would decrease to a mean service time of 2.4 min. Compare the present system with the proposed system with respect to the total expected number of people in the bank, the expected time a customer would have to spend in the bank, the probability of a customer having to wait more than 5 min, and the average idle time of the tellers.

- 2.24.** The Hott Too Trott Heating and Air Conditioning Company must choose between operating two types of service shops for maintaining its trucks. It estimates that trucks will arrive at the maintenance facility according to a Poisson distribution with mean rate of one every 40 min and believes that this rate is independent of which facility is chosen. In the first type of shop, there are dual facilities operating in parallel; each facility can service a truck in 30 min on average (the service time follows an exponential distribution). In the second type there is a single facility, but it can service a truck in 15 min on average (service times are also exponential in this case). To help management decide, they ask their operations research analyst to answer the following questions:
- (a) How many trucks, on average, will be in each of the two types of facilities?
 - (b) How long, on average, will a truck spend in each of the two types of facilities?
 - (c) Management calculates that each minute a truck must spend in the shop reduces contribution to profit by two dollars. They also know from previous experience in running dual-facility shops that the cost of operating such a facility is one dollar per minute (including labor, overhead, etc.). What would the operating cost per minute have to be for operating the single-facility shop in order for there to be no difference between the two types of shops?
- 2.25.** The ComPewter Company, which leases out high-end computer workstations, considers it necessary to overhaul its equipment once a year. Alternative 1 is to provide two separate maintenance stations where all work is done by hand (one machine at a time) for a total annual cost of \$750,000. The maintenance time for a machine has an exponential distribution with a mean of 6 h. Alternative 2 is to provide one maintenance station with mostly automatic equipment involving an annual cost of \$1 million. In this case, the maintenance time for a machine has an exponential distribution with a mean of 3 h. For both alternatives, the machines arrive according to a Poisson input with a mean arrival rate of one every 8 h (since the company leases such a large number of machines, we can consider the machine population as infinite). The cost of downtime per machine is \$150/h. Which alternative should the company choose? Assume that the maintenance facilities are always open and that they work $(24)(365) = 8760$ h/yr.

- 2.26.** Show the following:
- An $M/M/1$ is always better with respect to L than an $M/M/2$ with the same ρ .
 - An $M/M/2$ is always better than two independent $M/M/1$ queues with the same service rate but each getting half of the arrivals.
- 2.27.** For Problem 2.26(a), show that the opposite is true when considering L_q . In other words, faced with a choice between two M/M systems with identical arrival rates, one with two servers and one with a single server who can work twice as fast as each of the two servers, which is the preferable system?
- 2.28.** Show for the $M/M/c/K$ model that taking the limit for p_n and p_0 as $K \rightarrow \infty$ and restricting $\lambda/c\mu < 1$ in (2.45), (2.46), and (2.47) yield the results obtained for the $M/M/c/\infty$ model.
- 2.29.** Show that the $M/M/c/K$ equations (2.45)–(2.47) reduce to those for $M/M/1/K$ when $c = 1$.
- 2.30.** For the $M/M/3/K$ model, compute L_q as K goes from 3 to “ ∞ ” for each of the following ρ values: 1.5, 1, 0.8, 0.5. Comment.
- 2.31.** Find the probability that a customer’s wait in queue exceeds 20 min for an $M/M/1/3$ model with $\lambda = 4/\text{h}$ and $1/\mu = 15 \text{ min}$.
- 2.32.** A small drive-it-through-yourself car wash, in which the next car cannot go through the washing procedure until the car in front is completely finished, has a capacity to hold on its grounds a maximum of 10 cars (including the one in wash). The company has found its arrivals to be Poisson with mean rate of 20 cars/h, and its service times to be exponential with a mean of 12 min. What is the average number of cars lost to the firm every 10-h day as a result of its capacity limitations?
- 2.33.** Under the assumption that customers will not wait if no seats are available, Example 2.1’s hair salon proprietor Cutt can rent, on Saturday, the conference room of a small computer software firm adjacent to her shop for \$30.00 (cost of cleanup on a Saturday). Her shop is open on Saturdays from 8:00 A.M. to 2:00 P.M., and her marginal profit is \$6.75 per customer. This office can seat an additional four people. Should Cutt rent?
- 2.34.** The Fowler-Heir Oil Company operates a crude-oil unloading port at its major refinery. The port has six unloading berths and four unloading crews. When all berths are full, arriving ships are diverted to an overflow facility 20 miles down river. Tankers arrive according to a Poisson process with a mean of one every 2 h. It takes an unloading crew, on average, 10 h to unload a tanker, the unloading time following an exponential distribution. Tankers waiting for unloading crews are served on a first-come, first-served basis. Company management wishes to know the following:
- On average, how many tankers are at the port?

- (b) On average, how long does a tanker spend at the port?
- (c) What is the average arrival rate at the overflow facility?
- (d) The company is considering building another berth at the main port. Assume that construction and maintenance costs would amount to X dollars per year. The company estimates that to divert a tanker to the overflow port when the main port is full costs Y dollars. What is the relation between X and Y for which it would pay for the company to build an extra berth at the main port?
- 2.35.** Fly-Bynite Airlines has a telephone exchange with three lines, each manned by a clerk during its busy periods. During their peak three hours per 24-h period, many callers are unable to get into the exchange (there is no provision for callers to hold if all servers are busy). The company estimates, because of severe competition, that 60% of the callers not getting through use another airline. If the number of calls during these peak periods is roughly Poisson with a mean of 20 calls/h and each clerk spends on average 6 min with a caller, his service time being approximately exponentially distributed, and the average customer spends \$210/trip, what is the average daily loss due to the limited service facilities? (We may assume that the number of people not getting through during off-peak hours is negligible.) If a clerk's pay and fringe benefits cost the company \$24/h and a clerk must work an 8-h shift, what is the optimum number of clerks to employ? The three peak hours occur during the 8-h day shift. At all other times, one clerk can handle all the traffic, and since the company never closes the exchange, exactly one clerk is used on the off shifts. Assume that the cost of adding lines to the exchange is negligible.
- 2.36.** A call center has 24 phone lines and 3 customer service representatives. Suppose that calls arrive to the center according to a Poisson process with rate $\lambda = 15$ per hour. The time to process each call is exponential with a mean of 10 minutes. If all of the service representatives are busy, an arriving customer is placed on hold, but ties up one of the phone lines. If all of the phone lines are tied up, the customer receives a busy signal and the call is lost.
- (a) What is the average time that a customer spends on hold?
- (b) What is the average number of lines busy at one time?
- (c) Suppose that you pay \$0.03 per minute for each call to your center (including the time on hold). Also, the cost for each lost call is estimated at \$20. Fixing the number of service representatives, what is the optimal number of phone lines you should have?
- 2.37.** Prove the iterative relationship in (2.54) for the Erlang-B formula.
- 2.38.** Prove the relationship in (2.55) between the Erlang-B and Erlang-C formulas.

- 2.39.** Show that the steady-state probabilities obtained for the ample-server model ($M/M/\infty$) can also be developed by taking the limit as $c \rightarrow \infty$ in the results for the $M/M/c$ model.
- 2.40.** The Good Writers Correspondence Academy offers a go-at-your-own-pace correspondence course in good writing. New applications are accepted at any time, and the applicant can enroll immediately. Past records indicate applications follow a Poisson distribution with a mean of 8/month. An applicant's mean completion time is found to be 10 weeks, with the distribution of completion times being exponential. On average, how many pupils are enrolled in the school at any given time?
- 2.41.** An application of an $M/M/\infty$ model to the field of *inventory control* is as follows. A manufacturer of a very expensive, rather infrequently demanded item uses the following inventory control procedure. She keeps a safety stock of S units on hand. The customer demand for units can be described by a Poisson process with mean λ . Every time a request for a unit is made (a customer demand), an order is placed at the factory to manufacture another (this is called a one-for-one ordering policy). The amount of time required to manufacture a unit is exponential with mean $1/\mu$. There is a carrying cost for inventory on shelf of $\$h$ per unit per unit time held on shelf (representing capital tied up in inventory which could be invested and earning interest, insurance costs, spoilage, etc.) and a shortage cost of $\$p$ per unit (a shortage occurs when a customer requests a unit and there is none on shelf, i.e., safety stock is depleted to zero). It is assumed that customers who request an item but find that there is none immediately available will wait until stock is replenished by orders due in (this is called backordering or backlogging); thus one can look at the charge $\$p$ as a discount given to the customer because he must wait for his request to be satisfied. The problem, then, becomes one of finding the optimal value of S that minimizes total expected costs per unit time; that is, find the S that minimizes

$$E[C] = h \sum_{z=1}^S z p(z) + p\lambda \sum_{z=-\infty}^0 p(z) \quad (\$/\text{unit time}),$$

where z is the steady-state on-hand inventory level (+ means items on shelf, - means items in backorder) and $p(z)$ is the probability frequency function. Note that $\sum_{z=1}^S z p(z)$ is the average value of the safety stock and $\lambda \sum_{z=-\infty}^0 p(z)$ is the expected number of backorders per unit time, since the second summation is the fraction of time there is no on-shelf safety stock and λ is the average request rate. If $p(z)$ could be determined, one could optimize $E[C]$ with respect to S .

- (a) Show the relationship between Z and N , where N denotes the number of orders outstanding, that is, the number of orders currently being processed at the factory. Hence relate $p(z)$ to p_n .

- (b) Show that the $\{p_n\}$ are the steady-state probabilities of an $M/M/\infty$ queue if one considers the order-processing procedure as the queueing system. State explicitly what the input and service mechanisms are.
- (c) Find the optimum S for $\lambda = 8/\text{month}$, $1/\mu = 3 \text{ days}$, $h = \$50/\text{unit per month held}$, and $p = \$500 \text{ per unit backordered}$.
- 2.42.** Farecard machines that dispense tickets for riding on the subway have a mean operating time to breakdown of 45 h. It takes a technician on average 4 h to repair a machine, and there is one technician at each station. Assume that the time to breakdown and the time to repair are exponentially distributed. What is the number of installed machines necessary to assure that the probability of having at least five operational is greater than .95?
- 2.43.** For the machine repair model with spares, calculate all the usual measures of effectiveness for a problem with $M = 10$, $Y = 2$, $\lambda = 1$, $\mu = 3.5$, $c = 3$. Find the $\Pr\{N \geq k\}$ for $k = 2, 4$.
- 2.44.** Show for the basic machine repair model (no spares) that $q_n(M)$, the failure (arrival) point probabilities for a population of size M , equal $p_n(M - 1)$, the general-time probabilities for a population of size $M - 1$. The $\{q_n\}$ are sometimes referred to as *inside observer probabilities*, while the $\{p_n\}$ are referred to as *outside observer probabilities*.
- 2.45.** Derive $q_n(M)$ given by (2.65) for a machine repair problem with spares, and show that this is *not* equal to $p_n(M - 1)$, but is equal to $p_n(Y - 1)$ (i.e., the steady-state probabilities for a case where M is the same and the number of spares is reduced by one). The algebra is quite messy, so show it only for a numerical example ($M = 2$, $Y = 1$, $c = 1$, $\lambda/\mu = 1$). While that is no proof, the statement can be shown to hold in general (see Sevick and Mitrani, 1979, or Lavenberg and Reiser, 1979).
- 2.46.** A coin-operated dry-cleaning store has five machines. The operating characteristics of the machines are such that any machine breaks down according to a Poisson process with mean breakdown rate of one per day. A repairman can fix a machine according to an exponential distribution with a mean repair time of one-half day. Currently, three repairmen are on duty. The manager, Lew Cendirt, has the option of replacing these three repairmen with a super-repairman whose salary is equal to the total of the three regulars, but who can fix a machine in one-third the time, that is, in one-sixth day. Should he be hired?
- 2.47.** Suppose that each of five machines in a given shop breaks down according to a Poisson law at an average rate of one every 10 h, and the failures are repaired one at a time by two maintenance people operating as two channels, such that each machine has an exponentially distributed servicing requirement of mean 5 h.

- (a) What is the probability that exactly one machine will be up at any one time?
- (b) If performance of the workforce is measured by the ratio of average waiting time to average service time, what is this measure for the current situation?
- (c) What is the answer to (a) if an identical spare machine is put on reserve?
- 2.48.** Find the steady-state probabilities for a machine-repair problem with M machines, Y spares, and c technicians ($c \leq Y$) but with the following discipline: If no spares are on hand and a machine fails ($n = Y + 1$), the remaining $M - 1$ machines running are stopped until a machine is repaired; that is, if the machines are to run, there must be M running simultaneously.
- 2.49.** Very often in real-life modeling, even when the calling population is finite, an infinite-source model is used as an approximation. To compare the two models, calculate L for Example 2.9 assuming the calling population (number of machines) is infinite. Also calculate L for an exact model when the number of machines equals 10 and 5, respectively, for $M\lambda = \frac{1}{3}$ in both cases, and compare to the calculations from an approximate infinite-source model. How do you think ρ affects the approximation? [Hint: When using an infinite-source model as an approximation to a finite-source model, λ must be set equal to $M\lambda$.]
- 2.50.** Find the average operating costs per hour of Example 2.10 when the following conditions prevail:
- (a) C_1 = low-speed cost = \$25/(operating hour); C_2 = high-speed cost = \$50/(operating hour).
- (b) C_1 = \$25/(operating hour); C_2 = \$60/(operating hour).
- (c) Evaluate (b) for $k = 4$. What now is the best policy?
- 2.51.** Assume we have a two-state, state-dependent service model as described in Section 2.9 with $\rho_1 = \frac{4}{3}$ and $\rho = \frac{2}{3}$. Suppose that the customers are lawn-treating machines owned by the Green Thumb Lawn Service Company and these machines require, at random times, greasing on the company's two-speed greasing machine. Furthermore, suppose that the cost per operating hour of the greaser at the lower speed, C_1 , is \$25, and at the high speed, C_2 , is \$110. Also, the company estimates the cost of downtime of a lawn treater to be \$5/h. What is the optimal switch point k ? [Hint: Try several values of k starting at $k = 1$, and compute the total expected cost.]
- 2.52.** Derive the steady-state system-size probabilities for a c -server model with Poisson input and exponential state-dependent service where the mean service rate switches from μ_1 to μ when $k > c$ are in the system.
- 2.53.** Derive the steady-state system-size probabilities for a single-server model with Poisson input and exponential state-dependent service with mean rates $\mu_1 (1 \leq n < k_1)$, $\mu_2 (k_1 \leq n < k_2)$, and $\mu (n \geq k_2)$.

- 2.54.** For the problem treated at the end of Section 2.9 where $\mu_n = n^\alpha \mu$, show for $\alpha \geq 1$ that the tail of the infinite series for calculating p_0 discarded by truncation at some value N is bounded by the tail of the series for e^r . So, given any $\epsilon > 0$, N can be found such that the discarded tail is less than ϵ . Furthermore, show that if $p_0(N)$ is the estimate of p_0 based on N terms (where N is such that the discarded tail is bounded by ϵ), then the error bounds on p_0 become

$$p_0 < p_0(N) < \frac{p_0}{1 - \epsilon} \approx p_0(1 + \epsilon).$$

- 2.55.** It is known for an $M/M/1$ balking situation that the stationary distribution is given by the negative binomial

$$p_n = \binom{N+n-1}{N-1} x^n (1+x)^{-N-n} \quad (n \geq 0, \quad x > 0, \quad N > 1).$$

Find L , L_q , W , W_q , and b_n .

- 2.56.** For an $M/M/1$ balking model, it is known that $b_n = e^{-\alpha n / \mu}$. Find p_n (for all n).
- 2.57.** Consider a single-server queue where the arrivals are Poisson with rate $\lambda = 10$ per hour. The service distribution is exponential with rate $\mu = 5$ per hour. Suppose that customers balk at joining the queue when it is too long. Specifically, when there are n in the system, an arriving customer joins the queue with probability $1/(1+n)$. Determine the steady-state probability that there are n in the system.
- 2.58.** Suppose that the $M/M/1$ reneging model of Section 2.10.2 has the balking function $b_n = 1/n$ for $0 \leq n \leq k$ and 0 for $n > k$, and a reneging function $r(n) = n/\mu$. Find the stationary system-size distribution.
- 2.59.** Derive the steady-state $M/M/\infty$ solution directly from the transient.
- 2.60.** Find the mean number in an $M/M/\infty$ system at time t , assuming the system is empty at $t = 0$.
- 2.61.** For $\rho = 0.5, 0.9$, and 1 in an $M/M/1/\infty$ model with $\lambda = 1$, plot $p_0(t)$ versus t as t goes to infinity. Comment.
- 2.62.** For $\rho = 0.5$ in an $M/M/1/\infty$ model with $\lambda = 1$, find L at $t = 3$.
- 2.63.**
- (a) Prove that if $\bar{f}(s)$ is the Laplace transform of $f(t)$, then $\bar{f}(s+a)$ is the Laplace transform of $e^{-at} f(t)$.
 - (b) Show that the Laplace transform of a linear combination of functions is the same linear combination of the transforms of the functions: symbolically, $\mathcal{L}[\sum_i a_i f_i(t)] = \sum_i a_i \mathcal{L}[f_i(t)]$.

- 2.64.** Use the properties of Laplace transforms to find the functions whose Laplace transforms are the following:
- (a) $(s + 1)/(s^2 + 2s + 2)$.
 - (b) $1/(s^2 - 3s + 2)$.
 - (c) $1/[s^2(s^2 + 1)]$.
 - (d) $e^{-s}/(s + 1)$.
- 2.65.** For the following generating functions (not necessarily *probability* generating functions), write the sequence they generate:
- (a) $G(z) = 1/(1 - z)$.
 - (b) $G(z) = z/(1 - z)$.
 - (c) $G(z) = e^z$.
- 2.66.** Show that the moment generating function of the sum of independent random variables is equal to the product of their moment generating functions.
- 2.67.** Use the result of Problem 2.66 to show the following:
- (a) The sum of two independent Poisson random variables is a Poisson random variable.
 - (b) The sum of two independent and identical exponential random variables has a gamma or Erlang distribution.
 - (c) The sum of two independent but nonidentical exponential random variables has a density that is a linear combination of the two original exponential densities.

7.11 Exercises

EXERCISE 37. (Post office)

At a post office customers arrive according to a Poisson process with a rate of 30 customers per hour. A quarter of the customers wants to cash a cheque. Their service time is exponentially distributed with a mean of 2 minutes. The other customers want to buy stamps and their service times are exponentially distributed with a mean of 1 minute. In the post office there is only one server.

- (i) Determine the generating function $P_L(z)$ of the number of customers in the system.
- (ii) Determine the distribution of the number of customers in the system.
- (iii) Determine the mean number of customers in the system.
- (iv) Determine $\tilde{S}(s)$, the Laplace-Stieltjes transform of the sojourn time.
- (v) Determine the mean and the distribution of the sojourn time.
- (vi) Determine the mean busy period duration.
- (vii) Determine the mean number of customers in the system and the mean sojourn time in case all customers have an exponentially distributed service time with a mean of 75 seconds.

EXERCISE 38.

Consider a single machine where jobs arrive according to a Poisson stream with a rate of 10 jobs per hour. The processing time of a job consists of two phases. Each phase takes an exponential time with a mean of 1 minute.

- (i) Determine the Laplace-Stieltjes transform of the processing time.
- (ii) Determine the distribution of the number of jobs in the system.
- (iii) Determine the mean number of jobs in the system and the mean production lead time (waiting time plus processing time).

EXERCISE 39. (Post office)

At a post office customers arrive according to a Poisson process with a rate of 60 customers per hour. Half of the customers have a service time that is the sum of a fixed time of 15 seconds and an exponentially distributed time with a mean of 15 seconds. The other half have an exponentially distributed service time with a mean of 1 minute.

Determine the mean waiting time and the mean number of customers waiting in the queue.

EXERCISE 40. (Two phase production)

A machine produces products in two phases. The first phase is standard and the same for all products. The second phase is customer specific (the finishing touch). The first (resp.

second) phase takes an exponential time with a mean of 10 (resp. 2) minutes. Orders for the production of one product arrive according to a Poisson stream with a rate of 3 orders per hour. Orders are processed in order of arrival.

Determine the mean production lead time of an order.

EXERCISE 41.

Consider a machine where jobs are being processed. The mean production time is 4 minutes and the standard deviation is 3 minutes. The mean number of jobs arriving per hour is 10. Suppose that the interarrival times are exponentially distributed.

Determine the mean waiting time of the jobs.

EXERCISE 42.

Consider an $M/G/1$ queue, where the server successfully completes a service time with probability p . If a service time is not completed successfully, it has to be repeated until it is successful. Determine the mean sojourn time of a customer in the following two cases:

- (i) The repeated service times are identical.
- (ii) The repeated service times are independent, and thus (possibly) different.

EXERCISE 43.

At the end of a production process an operator manually performs a quality check. The time to check a product takes on average 2 minutes with a standard deviation of 1 minute. Products arrive according to a Poisson stream. One considers to buy a machine that is able to automatically check the products. The machine needs exactly 84 seconds to perform a quality check. Since this machine is expensive, one decides to buy the machine only if it is able to reduce lead time for a quality check (i.e. the time that elapses from the arrival of a product till the completion of its quality check) to one third of the lead time in the present situation.

So only if the arrival rate of products exceeds a certain threshold, one will decide to buy the machine. Calculate the value of this threshold.

EXERCISE 44. (Robotic dairy barn)

In a robotic dairy barn cows are automatically milked by a robot. The cows are lured into the robot by a feeder with nice food that can only be reached by first passing through the robot. When a cow is in the robot, the robot first detects whether the cow has to be milked. If so, then the cow will be milked, and otherwise, the cow can immediately leave the robot and walk to the feeder. A visit to the robot with (resp. without) milking takes an exponential time with a mean of 6 (resp. 3) minutes. Cows arrive at the robot according to a Poisson process with a rate of 10 cows per hour, a quarter of which will be milked.

- (i) Show that the Laplace-Stieltjes transform of the service time in minutes of an arbitrary cow at the robot is given by

$$\tilde{B}(s) = \frac{1}{4} \cdot \frac{4 + 21s}{(1 + 6s)(1 + 3s)}.$$

- (ii) Show that the Laplace-Stieltjes transform of the waiting time in minutes of an arbitrary cow in front of the robot is given by

$$\widetilde{W}(s) = \frac{3}{8} + \frac{9}{16} \cdot \frac{1}{1+12s} + \frac{1}{16} \cdot \frac{1}{1+4s}.$$

- (iii) Determine the fraction of cows for which the waiting time in front of the robot is less than 3 minutes.
- (iv) Determine the mean waiting time in front of the robot.

EXERCISE 45.

Consider a machine processing parts. These parts arrive according to a Poisson stream with a rate of 1 product per hour. The machine processes one part at a time. Each part receives two operations. The first operation takes an exponential time with a mean of 15 minutes. The second operation is done immediately after the first one and it takes an exponential time with a mean of 20 minutes.

- (i) Show that the Laplace-Stieltjes transform of the production lead time (waiting time plus processing time) in hours is given by

$$\widetilde{S}(s) = \frac{5}{4} \cdot \frac{1}{1+s} - \frac{1}{4} \cdot \frac{5}{5+s}.$$

- (ii) Determine the distribution of the production lead time and the mean production lead time.

One uses 3 hours as a norm for the production lead time. When the production lead time of a part exceeds this norm, it costs 100 dollar.

- (iii) Calculate the mean cost per hour.

EXERCISE 46. (Warehouse)

In a warehouse for small items orders arrive according to a Poisson stream with a rate of 6 orders per hour. An order is a list with the quantities of products requested by a customer. The orders are picked one at a time by one order picker. For a quarter of the orders the pick time is exponentially distributed with a mean of 10 minutes and for the other orders the pick time is exponentially distributed with a mean of 5 minutes.

- (i) Show that the Laplace-Stieltjes transform of the pick time in minutes of an arbitrary order is given by

$$\widetilde{B}(s) = \frac{1}{4} \cdot \frac{4+35s}{(1+10s)(1+5s)}.$$

- (ii) Show that the Laplace-Stieltjes transform of the lead time (waiting time plus pick time) in minutes of an arbitrary order is given by

$$\tilde{S}(s) = \frac{5}{32} \cdot \frac{3}{3+20s} + \frac{27}{32} \cdot \frac{1}{1+20s}.$$

- (iii) Determine the fraction of orders for which the lead time is longer than half an hour.
(iv) Determine the mean lead time.

EXERCISE 47. (Machine with breakdowns)

Consider a machine processing parts. Per hour arrive according to a Poisson process on average 5 orders for the production of a part. The processing time of a part is exactly 10 minutes. During processing, however, tools can break down. When this occurs, the machine immediately stops, broken tools are replaced by new ones and then the machine resumes production again. Replacing tools takes exactly 2 minutes. The time that elapses between two breakdowns is exponentially distributed with a mean of 20 minutes. Hence, the total production time of a part, B , consists of the processing time of 10 minutes plus a random number of interruptions of 2 minutes to replace broken tools.

- (i) Determine the mean and variance of the production time B .
(ii) Determine the mean lead time (waiting time plus production time) of an order.

EXERCISE 48. (Arrival and departure distribution)

Consider a queueing system in which customers arrive one by one and leave one by one. So the number of customers in the system only changes by $+1$ or -1 . We wish to prove that $a_n = d_n$. Suppose that the system is empty at time $t = 0$.

- (i) Show that if $L_{k+1}^d \leq n$, then $L_{k+n+1}^a \leq n$.
(ii) Show that if $L_{k+n+1}^a \leq n$, then $L_k^d \leq n$.
(iii) Show that (i) and (ii) give, for any $n \geq 0$,

$$\lim_{k \rightarrow \infty} P(L_k^d \leq n) = \lim_{k \rightarrow \infty} P(L_k^a \leq n).$$

EXERCISE 49. (Number served in a busy period)

Let the random variable N_{bp} denote the number of customers served in a busy period. Define the probabilities f_n by

$$f_n = P(N_{bp} = n).$$

We now wish to find its generating function

$$F_{N_{bp}}(z) = \sum_{n=1}^{\infty} f_n z^n.$$

(i) Show that $F_{N_{bp}}(z)$ satisfies

$$F_{N_{bp}}(z) = z\tilde{B}(\lambda - \lambda F_{N_{bp}}(z)). \quad (7.24)$$

(ii) Show that the mean number of customers served in a busy period is given by

$$E(N_{bp}) = \frac{1}{1 - \rho}.$$

(iii) Solve equation (7.24) for the $M/M/1$ queue.

(iv) Solve equation (7.24) for the $M/D/1$ queue (*Hint:* Use Lagrange inversion formula, see, e.g., [30]).

9.5 Exercises

EXERCISE 56.

Customers arrive to a single-server queue according to a Poisson process with a rate of 10 customers per hour. Half of the customers has an exponential service time with a mean of 3 minutes. The other half has an exponential service time with a mean of 6 minutes. The first half are called type 1 customers, the second half type 2 customers. Determine in each of the following three cases the mean waiting time:

- (i) Service in order of arrival (FCFS).
- (ii) Non-preemptive priority for type 1 customers.
- (iii) Non-preemptive priority for type 2 customers.

EXERCISE 57.

Consider an $M/D/1$ queue with three types of customers. The customers arrive according to a Poisson process. Per hour arrive on average 1 type 1 customer, 2 type 2 customers and also 2 type 3 customers. Type 1 customers have preemptive resume priority over type 2 and 3 customers and type 2 customers have preemptive resume priority over type 3 customers. The service time of each customer is 10 minutes.

Calculate the mean sojourn time for each of the three types of customers.

EXERCISE 58.

Consider a machine where jobs arrive according to a Poisson stream with a rate of 4 jobs per hour. Half of the jobs have a processing time of exactly 10 minutes, a quarter of the jobs have a processing time of exactly 15 minutes and the remaining quarter have a processing time of 20 minutes. The jobs with a processing time of 10 minutes are called type 1 jobs, the ones with a processing time of 15 minutes type 2 jobs and the rest type 3 jobs. The jobs are processed in order of arrival.

- (i) Determine the mean sojourn time (waiting time plus processing time) of a type 1, 2 and 3 job and also of an arbitrary job.

One decides to process smaller jobs with priority. So type 1 orders have highest priority, type 2 orders second highest priority and type 3 orders lowest priority.

Answer question (i) for the following two cases:

- (ii) Jobs processed at the machine may not be interrupted.
- (iii) Type 1 and type 2 jobs may interrupt the processing of a type 3 job. Type 1 jobs may not interrupt the processing of a type 2 job.

EXERCISE 59.

A machine produces a specific part type. Orders for the production of these parts arrive according to a Poisson stream with a rate of 10 orders per hour. The number of parts that has to be produced for an order is equal to n with probability 0.5^n , $n = 1, 2, \dots$. The production of one part takes exactly 2 minutes. Orders are processed in order of arrival.

- (i) Denote by B the production time in minutes of an arbitrary order. Show that

$$E(B) = 4, \quad \sigma^2(B) = 8.$$

- (ii) Determine the mean production lead time (waiting time plus production time) of an arbitrary order.

The management decides to give orders for the production of 1 part priority over the other orders. But the production of an order may not be interrupted.

- (iii) Determine the mean production lead of an order for the production of 1 part, and of an order for the production of at least 2 parts.
(iv) Determine the mean production lead time of an arbitrary order.

EXERCISE 60.

Consider a machine for mounting electronic components on printed circuit boards. Per hour arrive according to a Poisson process on average 30 printed circuit boards. The time required to mount all components on a printed circuit board is uniformly distributed between 1 and 2 minutes.

- (i) Calculate the mean sojourn time of an arbitrary printed circuit board.

One decides to give printed circuit boards with a mounting time between 1 and x ($1 \leq x \leq 2$) minutes non-preemptive priority over printed circuit boards the mounting time of which is greater than x minutes.

- (ii) Determine the mean sojourn time of an arbitrary printed circuit board as a function of x .
(iii) Determine the value of x for which the mean sojourn time of an arbitrary printed circuit board is minimized, and calculate for this specific x the relative improvement with respect to the mean sojourn time calculated in (i).

EXERCISE 61.

A machine produces 2 types of products, type A and type B . Production orders (for one product) arrive according to a Poisson process. For the production of a type A product arrive on average 105 orders per day (8 hours), and for a type B product 135 orders per day. The processing times are exponentially distributed. The mean processing time for a type A order is 1 minute and for a type B order it is 2 minutes. Orders are processed in order of arrival.

- (i) Calculate the mean production lead time for a type A product and for a type B product.
(ii) Determine the Laplace-Stieltjes transform of the waiting time.

- (iii) Determine the Laplace-Stieltjes transform of the production lead time of a type A product and determine the distribution of the production lead time of a type A product.

For the production lead time of a type A product one uses a norm of 10 minutes. Each time the production lead time of a type A product exceeds this norm, a cost of 100 dollar is charged.

- (iv) Calculate the average cost per day.

To reduce the cost one decides to give type A products preemptive resume priority over type B products.

- (v) Determine the mean production lead time for a type A product and for a type B product.

- (vi) Calculate the average cost per day.

EXERCISE 62.

A machine mounts electronic components on three different types of printed circuit boards, type A , B and C boards say. Per hour arrive on average 60 type A boards, 18 type B boards and 48 type C boards. The arrival streams are Poisson. The mounting times are exactly 20 seconds for type A , 40 seconds for type B and 30 seconds for type C . The boards are processed in order of arrival.

- (i) Calculate for each type of printed circuit board the mean waiting time and also calculate the mean overall waiting time.

Now suppose that the printed circuit boards are processed according to the SPTF rule.

- (ii) Calculate for each type of printed circuit board the mean waiting time and also calculate the mean overall waiting time.

11.5 Exercises

EXERCISE 74.

Prove that the fact that (11.1) holds for discrete service time distributions implies that it also holds for general service time distributions.

(*Hint:* Approximate the service time distribution from below and from above by discrete distributions and then consider the probability that there are n or more customers in the system.)

EXERCISE 75.

In a small restaurant there arrive according to a Poisson process on average 5 groups of customers. Each group can be accommodated at one table and stays for an Erlang-2 distributed time with a mean of 36 minutes. Arriving groups who find all tables occupied leave immediately.

How many tables are required such that at most 7% of the arriving groups is lost?

EXERCISE 76.

For a certain type of article there is a stock of at most 5 articles to satisfy customer demand directly from the shelf. Customers arrive according to a Poisson process at a rate of 2 customers per week. Each customer demands 1 article. The ordering policy is as follows. Each time an article is sold to a customer, an order for 1 article is immediately placed at the supplier. The lead time (the time that elapses from the moment the order is placed until the order arrives) is exponentially distributed with a mean of 1 week. If on arrival of a customer the shelf is empty, the customer demand will be lost. The inventory costs are 20 guilders per article per week. Each time an article is sold, this yields a reward of 100 guilders.

- (i) Calculate the probability distribution of the number of outstanding orders.
- (ii) Determine the mean number of articles on stock.
- (iii) What is the average profit (reward - inventory costs) per week?

EXERCISE 77.

In our library there are 4 VUBIS terminals. These terminals can be used to obtain information about the available literature. If all terminals are occupied when someone wants information, then that person will not wait but leave immediately (to look for the required information somewhere else). A user session on a VUBIS terminal takes on average 2.5 minutes. Since the number of potential users is large, it is reasonable to assume that users arrive according to a Poisson stream. On average 72 users arrive per hour.

- (i) Determine the probability that i terminals are occupied, $i = 0, 1, \dots, 4$.
- (ii) What is the fraction of arriving users finding all terminals occupied?
- (iii) How many VUBIS terminals are required such that at most 5% of the arriving users find all terminals occupied?

EXERCISE 78.

Consider a machine continuously processing parts (there is always raw material available). The processing time of a part is exponentially distributed with a mean of 20 seconds. A finished part is transported immediately to an assembly cell by an automatic conveyor system. The transportation time is exactly 3 minutes.

- (i) Determine the mean and variance of the number of parts on the conveyor.

To prevent that too many parts are simultaneously on the conveyor one decides to stop the machine as soon as there are N parts on the conveyor. The machine is turned on again as soon as this number is less than N .

- (ii) Determine the throughput of the machine as a function of N .
- (iii) Determine the smallest N for which the throughput is at least 100 parts per hour.

EXERCISE 79.

A small company renting cars has 6 cars available. The costs (depreciation, insurance, maintenance, etc.) are 60 guilders per car per day. Customers arrive according to a Poisson process with a rate of 5 customers per day. A customer rents a car for an exponential time with a mean of 1.5 days. Renting a car costs 110 guilders per day. Arriving customers for which no car is available are lost (they will go to another company).

- (i) Determine the fraction of arriving customers for which no car is available.
- (ii) Determine the mean profit per day.

The company is considering to buy extra cars.

- (iii) How many cars should be bought to maximize the mean profit per day?

4.8 Exercises

EXERCISE 13. (bulk arrivals)

In a work station orders arrive according to a Poisson arrival process with arrival rate λ . An order consists of N independent jobs. The distribution of N is given by

$$P(N = k) = (1 - p)p^{k-1}$$

with $k = 1, 2, \dots$ and $0 \leq p < 1$. Each job requires an exponentially distributed amount of processing time with mean $1/\mu$.

- (i) Derive the distribution of the total processing time of an order.
- (ii) Determine the distribution of the number of orders in the system.

EXERCISE 14. (variable production rate)

Consider a work station where jobs arrive according to a Poisson process with arrival rate λ . The jobs have an exponentially distributed service time with mean $1/\mu$. So the service completion rate (the rate at which jobs depart from the system) is equal to μ .

If the queue length drops below the threshold Q_L the service completion rate is lowered to μ_L . If the queue length reaches Q_H , where $Q_H \geq Q_L$, the service rate is increased to μ_H . (L stands for low, H for high.)

Determine the queue length distribution and the mean time spent in the system.

EXERCISE 15.

A repair man fixes broken televisions. The repair time is exponentially distributed with a mean of 30 minutes. Broken televisions arrive at his repair shop according to a Poisson stream, on average 10 broken televisions per day (8 hours).

- (i) What is the fraction of time that the repair man has no work to do?
- (ii) How many televisions are, on average, at his repair shop?
- (iii) What is the mean throughput time (waiting time plus repair time) of a television?

EXERCISE 16.

In a gas station there is one gas pump. Cars arrive at the gas station according to a Poisson process. The arrival rate is 20 cars per hour. Cars are served in order of arrival. The service time (i.e. the time needed for pumping and paying) is exponentially distributed. The mean service time is 2 minutes.

- (i) Determine the distribution, mean and variance of the number of cars at the gas station.
- (ii) Determine the distribution of the sojourn time and the waiting time.
- (iii) What is the fraction of cars that has to wait longer than 2 minutes?

An arriving car finding 2 cars at the station immediately leaves.

- (iv) Determine the distribution, mean and variance of the number of cars at the gas station.
- (v) Determine the mean sojourn time and the mean waiting time of all cars (including the ones that immediately leave the gas station).

EXERCISE 17.

A gas station has two pumps, one for gas and the other for LPG. For each pump customers arrive according to a Poisson process. On average 20 customers per hour for gas and 5 customers for LPG. The service times are exponential. For both pumps the mean service time is 2 minutes.

- (i) Determine the distribution of the number of customers at the gas pump, and at the LPG pump.
- (ii) Determine the distribution of the *total* number of customers at the gas station.

EXERCISE 18.

Consider an $M/M/1$ queue with two types of customers. The mean service time of all customers is 5 minutes. The arrival rate of type 1 customers is 4 customers per hour and for type 2 customers it is 5 customers per hour. Type 1 customers are treated with priority over type 2 customers.

- (i) Determine the mean sojourn time of type 1 and 2 customers under the preemptive-resume priority rule.
- (ii) Determine the mean sojourn time of type 1 and 2 customers under the non-preemptive priority rule.

EXERCISE 19.

Consider an $M/M/1$ queue with an arrival rate of 60 customers per hour and a mean service time of 45 seconds. A period during which there are 5 or more customers in the system is called crowded, when there are less than 5 customers it is quiet. What is the mean number of crowded periods per day (8 hours) and how long do they last on average?

EXERCISE 20.

Consider a machine where jobs arrive according to a Poisson stream with a rate of 20 jobs per hour. The processing times are exponentially distributed with a mean of $1/\mu$ hours. The processing cost is 16μ dollar per hour, and the waiting cost is 20 dollar per order per hour.

Determine the processing speed μ minimizing the average cost per hour.

Chapter 5

$M/M/c$ queue

In this chapter we will analyze the model with exponential interarrival times with mean $1/\lambda$, exponential service times with mean $1/\mu$ and c parallel identical servers. Customers are served in order of arrival. We suppose that the occupation rate per server,

$$\rho = \frac{\lambda}{c\mu},$$

is smaller than one.

5.1 Equilibrium probabilities

The state of the system is completely characterized by the number of customers in the system. Let p_n denote the equilibrium probability that there are n customers in the system. Similar as for the $M/M/1$ we can derive the equilibrium equations for the probabilities p_n from the flow diagram shown in figure 5.1.

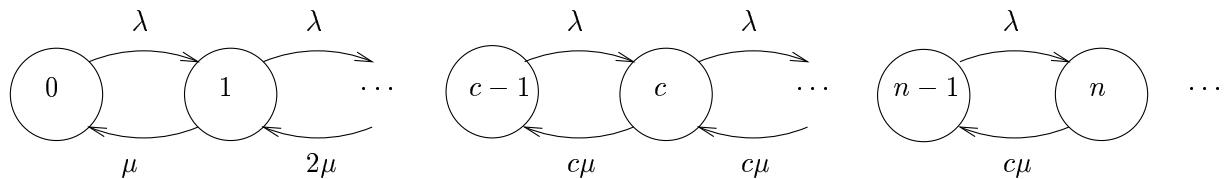


Figure 5.1: Flow diagram for the $M/M/c$ model

Instead of equating the flow into and out of a single state n , we get simpler equations by equating the flow out of and into the set of states $\{0, 1, \dots, n-1\}$. This amounts to equating the flow between the two neighboring states $n-1$ and n yielding

$$\lambda p_{n-1} = \min(n, c)\mu p_n, \quad n = 1, 2, \dots$$

Iterating gives

$$p_n = \frac{(c\rho)^n}{n!} p_0, \quad n = 0, \dots, c$$

5.5 Exercises

EXERCISE 21. (a fast and a slow machine)

Consider two parallel machines with a common buffer where jobs arrive according to a Poisson stream with rate λ . The processing times are exponentially distributed with mean $1/\mu_1$ on machine 1 and $1/\mu_2$ on machine 2 ($\mu_1 > \mu_2$). Jobs are processed in order of arrival. A job arriving when both machines are idle is assigned to the fast machine. We assume that

$$\rho = \frac{\lambda}{\mu_1 + \mu_2}$$

is less than one.

- (i) Determine the distribution of the number of jobs in the system.
- (ii) Use this distribution to derive the mean number of jobs in the system.
- (iii) When is it better to not use the slower machine at all?
- (iv) Calculate for the following two cases the mean number of jobs in the system with and without the slow machine.
 - (a) $\lambda = 2, \mu_1 = 5, \mu_2 = 1;$
 - (b) $\lambda = 3, \mu_1 = 5, \mu_2 = 1.$

Note: In [21] it is shown that one should not remove the slow machine if $r > 0.5$ where $r = \mu_2/\mu_1$. When $0 \leq r < 0.5$ the slow machine should be removed (and the resulting system is stable) whenever $\rho \leq \rho_c$, where

$$\rho_c = \frac{2 + r^2 - \sqrt{(2 + r^2)^2 + 4(1 + r^2)(2r - 1)(1 + r)}}{2(1 + r^2)}.$$

EXERCISE 22.

One is planning to build new telephone boxes near the railway station. The question is how many boxes are needed. Measurements showed that approximately 80 persons per hour want to make a phone call. The duration of a call is approximately exponentially distributed with mean 1 minute. How many boxes are needed such that the mean waiting time is less than 2 minutes?

EXERCISE 23.

An insurance company has a call center handling questions of customers. Nearly 40 calls per hour have to be handled. The time needed to help a customer is exponentially distributed with mean 3 minutes. How many operators are needed such that only 5% of the customers has to wait longer than 2 minutes?

EXERCISE 24.

In a dairy barn there are two water troughs (i.e. drinking places). From each trough only

one cow can drink at the same time. When both troughs are occupied new arriving cows wait patiently for their turn. It takes an exponential time to drink with mean 3 minutes. Cows arrive at the water troughs according to a Poisson process with rate 20 cows per hour.

- (i) Determine the probability that there are i cows at the water troughs (waiting or drinking), $i = 0, 1, 2, \dots$.
- (ii) Determine the mean number of cows waiting at the troughs and the mean waiting time.
- (iii) What is the fraction of cows finding both troughs occupied on arrival?
- (iv) How many troughs are needed such that at most 10% of the cows find all troughs occupied on arrival?

EXERCISE 25.

A computer consists of three processors. Their main task is to execute jobs from users. These jobs arrive according to a Poisson process with rate 15 jobs per minute. The execution time is exponentially distributed with mean 10 seconds. When a processor completes a job and there are no other jobs waiting to be executed, the processor starts to execute maintenance jobs. These jobs are always available and they take an exponential time with mean 5 seconds. But as soon as a job from a user arrives, the processor interrupts the execution of the maintenance job and starts to execute the new job. The execution of the maintenance job will be resumed later (at the point where it was interrupted).

- (i) What is the mean number of processors busy with executing jobs from users?
- (ii) How many maintenance jobs are on average completed per minute?
- (iii) What is the probability that a job from a user has to wait?
- (iv) Determine the mean waiting time of a job from a user.

Solutions to Exercises

Exercise 1.

(i) Use that

$$P(Y_n > x) = \prod_{i=1}^n P(X_i > x)$$

and

$$P(Z_n \leq x) = \prod_{i=1}^n P(X_i \leq x).$$

(ii) It follows that

$$\begin{aligned} P(X_i = \min(X_1, \dots, X_n)) &= \int_0^\infty \prod_{j \neq i} P(X_j > x) f_{X_i}(x) dx \\ &= \int_0^\infty \mu_i e^{-(\mu_1 + \dots + \mu_n)x} dx = \frac{\mu_i}{(\mu_1 + \dots + \mu_n)}. \end{aligned}$$

Exercise 1

Exercise 2.

Use Laplace-Stieltjes transforms to prove that

$$\begin{aligned} E(e^{-sS}) &= \sum_{k=1}^{\infty} P(N = k) E(e^{-sS} \mid N = k) \\ &= \sum_{k=1}^{\infty} (1-p)p^{k-1} \left(\frac{\mu}{\mu+s} \right)^k = \frac{\mu(1-p)}{\mu(1-p)+s}. \end{aligned}$$

Exercise 2

Exercise 3.

Use that the random variable

$$X = \begin{cases} 1/\mu_1, & \text{with probability } p_1, \\ \vdots & \vdots \\ 1/\mu_k, & \text{with probability } p_k, \end{cases}$$

has variance ≥ 0 , and hence that

$$\sum_{i=1}^k p_i (1/\mu_i)^2 \geq \left(\sum_{i=1}^k p_i (1/\mu_i) \right)^2.$$

Exercise 3

Exercise 4.

(i) $p_0(t) = P(A_1 > t) = e^{-\lambda t}.$

(ii) Use that

$$p_n(t + \Delta t) = \lambda \Delta t p_{n-1}(t) + (1 - \lambda \Delta t) p_n(t) + o(\Delta t),$$

and let Δt tend to zero.

(iii) Prove by induction that

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

is the solution of the differential equation in (ii).

Exercise 4

Exercise 5.

(i) $p_0(t) = P(A_1 > t) = e^{-\lambda t}.$

(ii) Use that

$$P(N(t) = n) = \int_0^\infty P(N(t) = n \mid A_1 = x) f_{A_1(x)} dx.$$

(iii) Prove by induction that

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

is the solution of the integral equations in (ii).

Exercise 5

Exercise 6.

Merging property: Use that the minimum of independent exponential random variables is again an exponential random variable. (See also Exercise 1)

Splitting property: Use the result of Exercise 2.

[Exercise 6](#)

Exercise 7.

Choose $p = (18 - 4\sqrt{14})/25 = 0.1213$ and $\mu = (2 - p)/4 = 0.4697$.

Exercise 7

Exercise 8.

- (i) For a Coxian-2 distribution we have

$$E(X^2) = \frac{2}{\mu_1^2} + \frac{2p_1}{\mu_1\mu_2} + \frac{2p_1}{\mu_2^2}, \quad \text{and} \quad E(X) = \frac{1}{\mu_1} + \frac{p_1}{\mu_2}.$$

Use this to show that $E(X^2) \geq \frac{3}{2}E(X)^2$ and hence that $c_X^2 \geq \frac{1}{2}$.

- (ii) Show that both distributions have the same Laplace-Stieltjes transform. Try to understand why these distributions are equivalent! (cf. Exercise 7)

Exercise 8

Exercise 9. Use Laplace-Stieltjes transforms, or use the formula

$$X_1 = \min(X_1, X_2) + (X_1 - \min(X_1, X_2)),$$

where X_1 is an exponential random variable with parameter λ and X_2 is an exponential random variable, independent of X_1 , with parameter $\mu - \lambda$. Exercise 9

Exercise 10. Use Exercise 9 with $\mu = \mu_1$ and $\lambda = \mu_2$.

Exercise 10

Exercise 11. Use generating functions and the fact that for the sum $Z = X + Y$ of *independent* discrete random variables X and Y , it holds that (see subsection 2.2)

$$P_Z(z) = P_X(z) \cdot P_Y(z).$$

Exercise 11

Exercise 12. As time unit we take 1 minute.

- (i) Solve the (global) balance equations

$$\lambda q_n p_n = \mu p_{n+1}, \quad n = 0, 1, 2, 3,$$

where $\lambda = \mu = 1/3$, together with the normalization equation. This gives

$$p_0 = \frac{32}{103}, \quad p_1 = \frac{32}{103}, \quad p_2 = \frac{24}{103}, \quad p_3 = \frac{12}{103}, \quad p_4 = \frac{3}{103}.$$

(ii) $E(L) = 128/103 \approx 1.24$.

(iii) $E(S) = 384/71 \approx 5.41$ minutes.

(iv) $E(S) = 384/103 \approx 3.73$ minutes.

$E(W) = 171/103 \approx 1.66$ minutes.

Exercise 12

Exercise 13.

- (i) Exponential with parameter $\mu^* = \mu(1 - p)$ (see Exercise 2).
- (ii) $P(L = n) = (1 - \rho)\rho^n$ for $n = 0, 1, 2, \dots$, where $\rho = \lambda/\mu^*$.

Exercise 13

Exercise 14. We have that

$$\text{service completion rate} = \begin{cases} \mu_L, & \text{if nr. of customers } < Q_L, \\ \mu, & \text{if } Q_L \leq \text{nr. of customers} < Q_H, \\ \mu_H, & \text{if nr. of customers } \geq Q_H. \end{cases}$$

The (global) balance equations are

$$\begin{aligned} \lambda p_n &= \mu_L p_{n+1}, & \text{if } n + 1 < Q_L, \\ \lambda p_n &= \mu p_{n+1}, & \text{if } Q_L \leq n + 1 < Q_H, \\ \lambda p_n &= \mu_H p_{n+1}, & \text{if } n + 1 \geq Q_H. \end{aligned}$$

The solution of these equations is given by

$$p_n = \begin{cases} p_0 \left(\frac{\lambda}{\mu_L} \right)^n, & \text{if } n < Q_L, \\ p_0 \left(\frac{\lambda}{\mu_L} \right)^{Q_L-1} \left(\frac{\lambda}{\mu} \right)^{n-Q_L+1}, & \text{if } Q_L \leq n < Q_H, \\ p_0 \left(\frac{\lambda}{\mu_L} \right)^{Q_L-1} \left(\frac{\lambda}{\mu} \right)^{Q_H-Q_L} \left(\frac{\lambda}{\mu_H} \right)^{n-Q_H+1}, & \text{if } n \geq Q_H. \end{cases}$$

Finally, p_0 follows from the normalization equation.

Exercise 14

Exercise 15.

- (i) $3/8$
- (ii) $5/3$
- (iii) 80 minutes

Exercise 15

Exercise 16.

(i) $P(L = n) = \frac{1}{3} \left(\frac{2}{3}\right)^n, \quad n = 0, 1, 2, \dots$

and hence $E(L) = 2$ and $\sigma^2(L) = 6$.

(ii) $P(S \leq t) = 1 - e^{-t/6}, \quad t \geq 0,$

$$P(W \leq t) = 1 - \frac{2}{3}e^{-t/6}, \quad t \geq 0.$$

(iii) $\frac{2}{3}e^{-1/3} \approx 0.48.$

(iv) $p_0 = \frac{9}{19}, \quad p_1 = \frac{6}{19}, \quad p_2 = \frac{4}{19},$

hence $E(L) = 14/19 \approx 0.737$ and $\sigma^2(L) = 22/19 - (14/19)^2 \approx 0.615$.

(v) $E(S) = 42/19 \approx 2.21$ minutes and $E(W) = 12/19 \approx 0.63$ minutes.

Exercise 16

Exercise 17.

(i) It holds that

$$P(L^{(Gas)} = n) = \frac{1}{3} \left(\frac{2}{3}\right)^n, \quad n = 0, 1, 2, \dots$$

and

$$P(L^{(LPG)} = n) = \frac{5}{6} \left(\frac{1}{6}\right)^n, \quad n = 0, 1, 2, \dots$$

(ii) Use

$$P(L = n) = \sum_{k=0}^n P(L^{(Gas)} = k, L^{(LPG)} = n - k)$$

to show that

$$P(L = n) = \frac{20}{54} \left(\frac{2}{3}\right)^n - \frac{5}{54} \left(\frac{1}{6}\right)^n, \quad n = 0, 1, 2, \dots$$

Exercise 17

Exercise 18.

(i) $E(S_1) = 7.5$ minutes.

$E(S_2) = 30$ minutes.

(ii) $E(S_1) = 10.625$ minutes.

$E(S_2) = 27.5$ minutes.

Exercise 18

Exercise 19. Fraction of time that it is crowded: $\rho^5 \approx 0.24$.

Number of crowded periods (per 8 hours = 480 minutes): $\lambda p_4 = 480(1 - \rho)\rho^4 \approx 38$.

E(crowded period) = E(busy period) = 3 minutes.

Exercise 19

Exercise 20. We have

$$\begin{aligned}\text{average costs per hour} &= 16\mu + 20E(L^q) \\ &= 16\mu + 20 \frac{\rho^2}{(1-\rho)} \\ &= 16\mu + \frac{8000}{\mu(\mu-20)}.\end{aligned}$$

For $\mu > 20$, this function is minimal for $\mu = \mu^* \approx 25$.

Exercise 20

Exercise 21. As state description we use the number of jobs in the system, and if this number of jobs is equal to 1, we distinguish between state $(1, f)$ in which the fast server is working and state $(1, s)$ in which the slow server is working.

(i) As solution of the balance equations we find

$$\begin{aligned} p_0 &= \frac{1 - \rho}{1 - \rho + C}, \\ p_1 &= p_{1,f} + p_{1,s} = Cp_0, \\ p_n &= \rho^{n-1}p_1, \quad n > 1, \end{aligned}$$

where we used the notation $\mu = \mu_1 + \mu_2$, $\rho = \lambda/\mu$ and

$$C = \frac{\lambda\mu(\lambda + \mu_2)}{\mu_1\mu_2(2\lambda + \mu)}.$$

(ii) For the mean number of jobs in the system we find

$$E(L) = \sum_{n=1}^{\infty} np_n = \frac{C}{(1 - \rho)(1 - \rho + C)}.$$

(iii) It is better not to use the slower machine at all if $E(L^f)$, the expected number of jobs in the system when you only use the fast server is smaller than $E(L)$. This is the case if $\mu_1 > \lambda$ and

$$\frac{\lambda}{\mu_1 - \lambda} < \frac{C}{(1 - \rho)(1 - \rho + C)}.$$

(iv) In case (a) we have

$$E(L^f) = \frac{2}{3} < \frac{81}{104} = E(L).$$

In case (b) we have

$$E(L^f) = \frac{3}{2} > \frac{24}{17} = E(L).$$

Exercise 21

Exercise 22. As time unit we choose 1 minute: $\lambda = 4/3$ and $\mu = 1$. In order to have $\rho < 1$ we need that $c \geq 2$. Hence, we first try $c = 2$. This gives $\Pi_W = 8/15 \approx 0.533$ and $E(W) = 24/30 = 0.8$ minutes. Hence, we conclude that 2 boxes is enough. Exercise 22

Exercise 23. As time unit we choose 1 minute: $\lambda = 2/3$ and $\mu = 1/3$. In order to have $\rho < 1$ we need that $c \geq 3$. Hence, we first try $c = 3$. This gives $\Pi_W = 4/9 \approx 0.444$ and

$$P(W > 2) = \Pi_W \cdot e^{-2/3} \approx 0.228 > 0.05.$$

Similarly, for $c = 4$ we find $\Pi_W = 4/23 \approx 0.174$ and

$$P(W > 2) = \Pi_W \cdot e^{-4/3} \approx 0.046 < 0.05.$$

Hence, we need at least 4 operators.

Exercise 23

Exercise 24. As time unit we choose 1 minute: $\lambda = 1/3$ and $\mu = 1/3$.

(i) We have

$$p_0 = \frac{1}{3}, \quad p_n = \frac{1}{3} \left(\frac{1}{2}\right)^{n-1}, \quad n \geq 1.$$

(ii) Using (5.2) and (5.3) we have

$$E(L^q) = \Pi_W \cdot \frac{\rho}{1-\rho} = 1/3, \quad E(W) = \Pi_W \cdot \frac{1}{1-\rho} \cdot \frac{1}{c\mu} = 1 \text{ minute.}$$

(iii) $\Pi_W = 1/3$.

(iv) For $c = 2$ we have $\Pi_W = 1/3 > 1/10$. Similarly, we can find for $c = 3$ that $\Pi_W = 1/11 < 1/10$. Hence, we need 3 troughs.

Exercise 24

Exercise 25. As time unit we choose 1 minute: $\lambda = 15$ and $\mu = 6$.

- (i) $c \cdot \rho = \lambda/\mu = 2.5$.
- (ii) $c \cdot (1 - \rho) \cdot 12 = 6$ maintenance jobs per minute.
- (iii) We have

$$p_0 = \frac{8}{178}, \quad p_1 = \frac{20}{178}, \quad p_n = \frac{25}{178} \left(\frac{5}{6}\right)^{n-2}, \quad n \geq 2,$$

and hence (see (5.1))

$$\Pi_W = \frac{125}{178} \approx 0.702.$$

- (iv) Using (5.3), we have

$$E(W) = \Pi_W \cdot \frac{1}{1 - \rho} \cdot \frac{1}{c\mu} = \frac{1}{3} \cdot \Pi_W \approx 0.234 \text{ minutes.}$$

Exercise 25

Exercise 30.

- (i) The distribution of the number of uncompleted tasks in the system is given by

$$p_n = \frac{7}{24} \left(\frac{2}{3}\right)^n + \frac{7}{40} \left(-\frac{2}{5}\right)^n, \quad n = 0, 1, 2, \dots$$

- (ii) The distribution of the number of jobs in the system is given by

$$q_n = \frac{35}{48} \left(\frac{4}{9}\right)^n - \frac{21}{80} \left(\frac{4}{25}\right)^n, \quad n = 0, 1, 2, \dots$$

- (iii) The mean number of jobs equals $E(L) = \sum_{n=1}^{\infty} nq_n = 104/105$.

- (iv) The mean waiting time of a job equals

$$E(W) = \frac{\rho}{1-\rho} E(R_B) = 8/7 \cdot 3/2 = 12/7 \text{ minutes.}$$

(Check: Little's formula $E(L) = \lambda E(S)$ is satisfied, with $E(L) = 104/105$ job, $\lambda = 4/15$ job per minute and $E(S) = 26/7$ minutes.)

Exercise 30

Exercise 31. Define T_i as the mean time till the first customer is rejected if we start with i phases work in the system at time $t = 0$. Then we have

$$\begin{aligned} T_0 &= 1 + T_2, \\ T_1 &= \frac{1}{2} + \frac{1}{2}T_0 + \frac{1}{2}T_3, \\ T_2 &= \frac{1}{2} + \frac{1}{2}T_1 + \frac{1}{2}T_4, \\ T_3 &= \frac{1}{2} + \frac{1}{2}T_2, \\ T_4 &= \frac{1}{2} + \frac{1}{2}T_3. \end{aligned}$$

The solution of this set of equations is given by

$$(T_0, T_1, T_2, T_3, T_4) = (4, 3\frac{1}{2}, 3, 2, 1\frac{1}{2}).$$

Hence, if at time $t = 0$ the system is empty, the mean time till the first customer is rejected is equal to 4.

Exercise 31

Exercise 32. The distribution of the number of phases work in the system is given by

$$p_n = \frac{7}{24} \left(\frac{2}{3}\right)^n + \frac{7}{40} \left(-\frac{2}{5}\right)^n, \quad n = 0, 1, 2, \dots$$

(i) The distribution of the waiting time (in minutes) is given by

$$P(W \leq t) = 1 - \frac{7}{12}e^{-\frac{1}{12}t} + \frac{1}{20}e^{-\frac{7}{20}t}.$$

(ii) The fraction of customers that has to wait longer than 5 minutes is given by

$$P(W > 5) = \frac{7}{12}e^{-\frac{5}{12}} - \frac{1}{20}e^{-\frac{7}{4}} \approx 0.376.$$

Exercise 32

Exercise 33.

- (i) The distribution of the number of customers in the system is given by

$$p_n = \frac{3}{7} \left(\frac{1}{2}\right)^n + \frac{6}{35} \left(-\frac{1}{5}\right)^n, \quad n = 0, 1, 2, \dots$$

- (ii) The mean number of customers equals $E(L) = \sum_{n=1}^{\infty} np_n = 5/6$. Now, either use the PASTA property

$$E(S) = (5/6) \cdot 6 + (1/4) \cdot 6 + 6$$

or use Little's formula

$$E(S) = \frac{5/6}{1/15}$$

to conclude that the mean sojourn time of an arbitrary customer is equal to 12.5 minutes.

Exercise 33

Exercise 34.

- (i) See Example 6.2.1.
- (ii) Use PASTA and/or Little to conclude that $E(S) = 11/12$ week.
- (iii) Because $p_0 + p_1 + p_2 < 0.99$ and $p_0 + p_1 + p_2 + p_3 > 0.99$ we need at least 4 spare engines.

Exercise 34

Exercise 35.

- (i) The distribution of the number of uncompleted tasks at the machine is given by

$$p_n = \frac{9}{39} \left(\frac{3}{4}\right)^n + \frac{4}{39} \left(-\frac{1}{3}\right)^n, \quad n = 0, 1, 2, \dots$$

- (ii) The mean number of uncompleted tasks equals $E(L_{task}) = \sum_{n=1}^{\infty} np_n = 11/4$. Hence, using PASTA, the mean waiting time of a job is $E(W) = E(L_{task}) \cdot 1 = 11/4$ minutes.
- (iii) The mean sojourn time of a job equals $E(S) = E(W) + E(B) = 11/4 + 8/5 = 87/20$ minutes. Hence, using Little's formula, we have $E(L_{job}) = 5/12 \cdot 87/20 = 29/16$.

Exercise 35

Exercise 36.

- (i) The distribution of the number of customers in the system is given by

$$p_n = \frac{2}{5} \left(\frac{1}{2}\right)^n + \frac{4}{15} \left(-\frac{1}{3}\right)^n, \quad n = 0, 1, 2, \dots$$

- (ii) For the mean number of customers in the system we have $E(L) = \sum_{n=1}^{\infty} np_n = 3/4$.
Hence, using PASTA, the mean waiting time of the first customer in a group equals
 $E(W_1) = E(L) \cdot 5 = 15/4$ minutes.
- (iii) The mean waiting time of the second customer in a group equals $E(W_2) = E(W_1) + 5 = 35/4$ minutes.

(Check: Little's formula $E(L^q) = \lambda E(W)$ is satisfied, with $E(L^q) = 3/4 - 1/3 = 5/12$ customer, $\lambda = 1/15$ customer per minute and $E(W) = 25/4$ minutes.)

Exercise 36

Exercise 37. As time unit we take 1 minute. Hence, $\lambda = 1/2$ and

$$\tilde{B}(s) = \frac{1}{4} \cdot \frac{\frac{1}{2}}{\frac{1}{2} + s} + \frac{3}{4} \cdot \frac{1}{1 + s}.$$

(i) From (7.6) we have

$$P_L(z) = \frac{(1 - \rho)\tilde{B}(\lambda - \lambda z)(1 - z)}{\tilde{B}(\lambda - \lambda z) - z} = \frac{3}{8} \frac{15 - 7z}{(3 - 2z)(5 - 2z)} = \frac{\frac{9}{32}}{1 - \frac{2}{3}z} + \frac{\frac{3}{32}}{1 - \frac{2}{5}z}.$$

(ii) The distribution of the number of customers is given by

$$p_n = \frac{9}{32} \left(\frac{2}{3}\right)^n + \frac{3}{32} \left(\frac{2}{5}\right)^n, \quad n = 0, 1, 2, \dots$$

(iii) $E(L) = \sum_{n=1}^{\infty} np_n = 43/24$.

(iv) From (7.7) we have

$$\tilde{S}(s) = \frac{(1 - \rho)\tilde{B}(s)s}{\lambda\tilde{B}(s) + s - \lambda} = \frac{3}{4} \frac{4 + 7s}{(1 + 4s)(3 + 4s)} = \frac{27}{32} \cdot \frac{\frac{1}{4}}{\frac{1}{4} + s} + \frac{5}{32} \cdot \frac{\frac{3}{4}}{\frac{3}{4} + s}.$$

(v) The distribution function of the sojourn time is given by

$$F_S(x) = \frac{27}{32} (1 - e^{-\frac{1}{4}x}) + \frac{5}{32} (1 - e^{-\frac{3}{4}x}),$$

and the mean sojourn time by

$$E(S) = \frac{27}{32} \cdot 4 + \frac{5}{32} \cdot \frac{4}{3} = \frac{43}{12} \text{ minutes.}$$

(vi) From (7.21) we have

$$E(BP) = \frac{E(B)}{1 - \rho} = \frac{10}{3} \text{ minutes.}$$

(vii) For the $M/M/1$ queue we have

$$E(L) = \frac{\rho}{1 - \rho} = \frac{5}{3},$$

and

$$E(S) = \frac{E(L)}{\lambda} = \frac{10}{3} \text{ minutes.}$$

Exercise 37

Exercise 38. As time unit we take 1 minute, so $\lambda = 1/6$.

(i) $\tilde{B}(s) = \left(\frac{1}{1+s}\right)^2$.

(ii) $p_n = \frac{6}{5} \left(\frac{1}{4}\right)^n - \frac{8}{15} \left(\frac{1}{9}\right)^n, \quad n = 0, 1, 2, \dots$

(iii) $E(L) = 11/24$ and $E(S) = 11/4$ minutes.

Exercise 38

Exercise 39. As time unit we take 1 minute, so $\lambda = 1$. Let X be exponential with parameter 4 and Y be exponential with parameter 1. Then,

$$\begin{aligned} E(B) &= \frac{1}{2} \cdot E\left(\frac{1}{4} + X\right) + \frac{1}{2} \cdot E(Y) = \frac{3}{4} \text{ minutes,} \\ E(B^2) &= \frac{1}{2} \cdot E\left([\frac{1}{4} + X]^2\right) + \frac{1}{2} \cdot E(Y^2) = \frac{1}{2} \cdot \frac{5}{16} + \frac{1}{2} \cdot 2 = \frac{37}{32} \text{ minutes,} \end{aligned}$$

and so $E(R) = 37/48$ minutes. Hence, $E(W) = 37/16$ minutes and $E(L^q) = 37/16$ customers.

Exercise 39

Exercise 40. As time unit we take 1 minute, so $\lambda = 1/20$. Furthermore, $E(B) = 12$ minutes and $E(R) = 31/3$ minutes. Hence, $E(S) = 55/2 = 27.5$ minutes. Exercise 40

Exercise 41. The mean waiting time of jobs is given by

$$E(W) = \frac{\rho}{1 - \rho} \cdot E(R) = \frac{25}{4} \text{ minutes.}$$

Exercise 41

Exercise 44. The time unit is 1 minute: $\lambda = 1/6$, $E(B) = 15/4$, $\rho = 5/8$, $E(R) = (2/5) \cdot 6 + (3/5) \cdot 3 = 21/5$.

(i) The service time is hyperexponentially distributed with parameters $p_1 = 1/4$, $p_2 = 3/4$, $\mu_1 = 1/6$ and $\mu_2 = 1/3$.

(ii) From (7.9) we have

$$\widetilde{W}(s) = \frac{(1 - \rho)s}{\lambda \widetilde{B}(s) + s - \lambda} = \frac{1 + 9s + 18s^2}{(1 + 12s)(1 + 4s)} = \frac{3}{8} + \frac{9}{16} \cdot \frac{1}{1 + 12s} + \frac{1}{16} \cdot \frac{1}{1 + 4s}.$$

(iii) The distribution function of the waiting time is given by

$$F_W(x) = \frac{3}{8} + \frac{9}{16} (1 - e^{-\frac{1}{12}x}) + \frac{1}{16} (1 - e^{-\frac{1}{4}x}).$$

Hence, the fraction of cows for which the waiting time is less than 3 minutes equals

$$F_W(3) = \frac{3}{8} + \frac{9}{16} (1 - e^{-\frac{1}{4}}) + \frac{1}{16} (1 - e^{-\frac{3}{4}}) \approx 0.532.$$

(iv) The mean waiting time is given by

$$E(W) = \frac{9}{16} \cdot 12 + \frac{1}{16} \cdot 4 = 7 \text{ minutes.}$$

Alternatively, from a mean value analysis we have

$$E(W) = \frac{\rho}{1 - \rho} E(R) = \frac{5}{3} \cdot \frac{21}{5} = 7 \text{ minutes.}$$

Exercise 44

Exercise 45. The time unit is 1 hour: $\lambda = 1$, $E(B) = 7/12$, $\rho = 7/12$, $E(R) = (3/7) \cdot (7/12) + (4/7) \cdot (1/3) = 37/84$.

(i) The Laplace-Stieltjes transform of the processing time is given by

$$\tilde{B}(s) = \frac{4}{4+s} \cdot \frac{3}{3+s}.$$

From (7.7) we have

$$\tilde{S}(s) = \frac{(1-\rho)\tilde{B}(s)s}{\lambda\tilde{B}(s)+s-\lambda} = \frac{5}{(1+s)(5+s)} = \frac{5}{4} \cdot \frac{1}{1+s} - \frac{1}{4} \cdot \frac{5}{5+s}.$$

(ii) The distribution function of the production lead time is given by

$$F_S(x) = \frac{5}{4} (1 - e^{-x}) - \frac{1}{4} (1 - e^{-5x}).$$

The mean production lead time is given by

$$E(S) = \frac{5}{4} \cdot 1 - \frac{1}{4} \cdot \frac{1}{5} = \frac{6}{5} \text{ hours.}$$

Alternatively, from a mean value analysis we have

$$E(S) = \frac{\rho}{1-\rho} E(R) + E(B) = \frac{7}{5} \cdot \frac{37}{84} + \frac{7}{12} = \frac{6}{5} \text{ hours.}$$

(iii) The mean cost per hour equals

$$\lambda \cdot (1 - F_S(3)) \cdot 100 = \left(\frac{5}{4} \cdot e^{-3} - \frac{1}{4} \cdot e^{-15} \right) \cdot 100 \approx 6.22 \text{ dollar.}$$

Exercise 45

Exercise 46. The time unit is 1 minute: $\lambda = 1/10$, $E(B) = 25/4$, $\rho = 5/8$, $E(R) = (2/5) \cdot 10 + (3/5) \cdot 5 = 7$.

(i) The pick time is hyperexponentially distributed with parameters $p_1 = 1/4$, $p_2 = 3/4$, $\mu_1 = 1/10$ and $\mu_2 = 1/5$.

(ii) From (7.7) we have

$$\tilde{S}(s) = \frac{(1 - \rho)\tilde{B}(s)s}{\lambda\tilde{B}(s) + s - \lambda} = \frac{12 + 105s}{4(3 + 20s)(1 + 20s)} = \frac{5}{32} \cdot \frac{3}{3 + 20s} + \frac{27}{32} \cdot \frac{1}{1 + 20s}.$$

(iii) The distribution function of the sojourn time is given by

$$F_S(x) = \frac{5}{32} (1 - e^{-\frac{3}{20}x}) + \frac{27}{32} (1 - e^{-\frac{1}{20}x}).$$

Hence, the fraction of orders for which the lead time is longer than half an hour is given by

$$1 - F_S(30) = \frac{5}{32} \cdot e^{-\frac{9}{2}} + \frac{27}{32} \cdot e^{-\frac{3}{2}} \approx 0.190.$$

(iv) The mean lead time is given by

$$E(S) = \frac{5}{32} \cdot \frac{20}{3} + \frac{27}{32} \cdot 20 = \frac{215}{12} \text{ minutes.}$$

Alternatively, from a mean value analysis we have

$$E(S) = \frac{\rho}{1 - \rho} E(R) + E(B) = \frac{5}{3} \cdot 7 + \frac{25}{4} = \frac{215}{12} \text{ minutes.}$$

Exercise 46

Exercise 51. As time unit we choose 1 minute: $\mu = 1$. The Laplace-Stieltjes transform of the interarrival time distribution is given by

$$\tilde{A}(s) = \frac{1}{3} \cdot \frac{1}{1+s} + \frac{2}{3} \cdot \frac{1}{1+3s}.$$

(i) $a_n = (1 - \sigma) \sigma^n$, where σ , the solution in $(0,1)$ of $\sigma = \tilde{A}(\mu - \mu\sigma)$, is given by

$$\sigma = \frac{21 - \sqrt{153}}{18} \approx 0.48.$$

(ii) $E(L^a) = \frac{\sigma}{1-\sigma} \approx 0.92$.

(iii) $\tilde{S}(s) = \frac{1-\sigma}{1-\sigma+s}$.

(iv) $E(S) = \frac{1}{1-\sigma} \approx 1.92$.

(v) $E(L) = \lambda \cdot E(S) = \frac{3}{7} \cdot E(S) \approx 0.82$.

Exercise 51

Exercise 52. We have $\mu = 6$ and the Laplace-Stieltjes transform of the interarrival time distribution is given by

$$\tilde{A}(s) = \frac{13}{24} \cdot \frac{3}{3+s} + \frac{11}{24} \cdot \frac{2}{2+s}.$$

(i) $a_n = (1 - \sigma) \sigma^n$, where σ , the solution in $(0,1)$ of $\sigma = \tilde{A}(\mu - \mu\sigma)$, is given by $\sigma = \frac{5}{12}$.

(ii) $F_W(t) = 1 - \sigma e^{-\mu(1-\sigma)t} = 1 - \frac{5}{12} e^{-\frac{7}{2}t}$.

Exercise 52

Exercise 53. The sojourn time is exponentially distributed with parameter $\mu(1 - \sigma)$, where $\mu = 1$ and

$$\sigma = \frac{7 - \sqrt{17}}{8} \approx 0.36.$$

Exercise 53

Exercise 54.

(i) The solution in $(0,1)$ of $\sigma = e^{-2(1-\sigma)}$, is given by $\sigma \approx 0.203$.

(ii) $F_S(t) = 1 - e^{-\mu(1-\sigma)t} \approx 1 - e^{-(0.4)t}$.

Exercise 54

Exercise 55.

(i) $3/8$.

(ii) Let p_n denote the probability that there are n cars waiting for the ferry. Then,

$$\begin{aligned} p_0 &= \frac{1}{4} + \frac{3}{4} \cdot \left(\frac{1}{2}\right)^2 = \frac{7}{16}, \\ p_1 &= \frac{3}{4} \cdot \left(\frac{1}{2}\right)^1 + \frac{3}{4} \cdot \left(\frac{1}{2}\right)^3 = \frac{15}{32}, \\ p_n &= \frac{3}{4} \cdot \left(\frac{1}{2}\right)^{n+2}, \quad n \geq 2. \end{aligned}$$

(iii) $E(L^q) = \sum_{n=0}^{\infty} np_n = \frac{3}{4}$, and hence using Little's formula we have $E(W) = 3$ minutes.

Exercise 55

Exercise 56. As time unit we choose 1 minute:

$$\lambda = 1/6, E(B) = 9/2, \rho = 3/4 \text{ and } E(R) = 5,$$

$$\lambda_1 = 1/12, E(B_1) = 3, \rho_1 = 1/4 \text{ and } E(R_1) = 3,$$

$$\lambda_2 = 1/12, E(B_2) = 6, \rho_2 = 1/2 \text{ and } E(R_2) = 6.$$

$$(i) \quad E(W) = \frac{\rho}{1-\rho} E(R) = 15 \text{ minutes.}$$

(ii) Use formula (9.3) on page 89:

$$E(W_1) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{1 - \rho_1} = 5 \text{ minutes ,}$$

$$E(W_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} = 20 \text{ minutes ,}$$

$$E(W) = \frac{1}{2} E(W_1) + \frac{1}{2} E(W_2) = 12.5 \text{ minutes .}$$

(iii) Similar to formula (9.3), we now have

$$E(W_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{1 - \rho_2} = \frac{15}{2} = 7.5 \text{ minutes ,}$$

$$E(W_1) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{(1 - \rho_2)(1 - \rho_1 - \rho_2)} = 30 \text{ minutes ,}$$

$$E(W) = \frac{1}{2} E(W_1) + \frac{1}{2} E(W_2) = 18.75 \text{ minutes .}$$

Exercise 56

Exercise 57. As time unit we choose 1 minute:

$$\lambda_1 = 1/60, E(B_1) = 10, \rho_1 = 1/6 \text{ and } E(R_1) = 5,$$

$$\lambda_2 = 1/30, E(B_2) = 10, \rho_2 = 1/3 \text{ and } E(R_2) = 5,$$

$$\lambda_3 = 1/30, E(B_3) = 10, \rho_3 = 1/3 \text{ and } E(R_3) = 5.$$

Now, use formula (9.5) for $E(S_i)$ on page 90:

$$E(S_1) = \frac{\rho_1 E(R_1)}{1 - \rho_1} + E(B_1) = 11 \text{ minutes ,}$$

$$E(S_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{E(B_2)}{1 - \rho_1} = 18 \text{ minutes ,}$$

$$E(S_3) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2) + \rho_3 E(R_3)}{(1 - \rho_1 - \rho_2)(1 - \rho_1 - \rho_2 - \rho_3)} + \frac{E(B_3)}{1 - \rho_1 - \rho_2} = 70 \text{ minutes ,}$$

Exercise 57

Exercise 58. As time unit we choose 1 minute:

$$\lambda = 1/15, E(B) = 55/4, \rho = 11/12 \text{ and } E(R) = 15/2,$$

$$\lambda_1 = 1/30, E(B_1) = 10, \rho_1 = 1/3 \text{ and } E(R_1) = 5,$$

$$\lambda_2 = 1/60, E(B_2) = 15, \rho_2 = 1/4 \text{ and } E(R_2) = 15/2,$$

$$\lambda_3 = 1/60, E(B_3) = 20, \rho_3 = 1/3 \text{ and } E(R_3) = 10.$$

(i) $E(W_1) = E(W_2) = E(W_3) = E(W) = \frac{\rho}{1-\rho}E(R) = 165/2 = 82.5$ minutes. Hence, $E(S_1) = 92.5$ minutes, $E(S_2) = 97.5$ minutes, $E(S_3) = 102.5$ minutes and $E(S) = 96.25$ minutes.

(ii) Use formula (9.4) for $E(S_i)$ on page 89:

$$E(S_1) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2) + \rho_3 E(R_3)}{1 - \rho_1} + E(B_1) = \frac{325}{16} \approx 20.31 \text{ minutes ,}$$

$$E(S_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2) + \rho_3 E(R_3)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + E(B_2) = \frac{159}{4} = 39.75 \text{ minutes ,}$$

$$E(S_3) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2) + \rho_3 E(R_3)}{(1 - \rho_1 - \rho_2)(1 - \rho_1 - \rho_2 - \rho_3)} + E(B_3) = 218 \text{ minutes ,}$$

$$E(S) = \frac{1}{2}E(S_1) + \frac{1}{4}E(S_2) + \frac{1}{4}E(S_3) \approx 74.59 \text{ minutes .}$$

(iii) Combine the arguments of Sections 9.1 and 9.2:

$$E(S_1) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{1 - \rho_1} + E(B_1) = \frac{245}{16} \approx 15.31 \text{ minutes ,}$$

$$E(S_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + E(B_2) = \frac{111}{4} = 27.75 \text{ minutes ,}$$

$$E(S_3) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2) + \rho_3 E(R_3)}{(1 - \rho_1 - \rho_2)(1 - \rho_1 - \rho_2 - \rho_3)} + \frac{E(B_3)}{1 - \rho_1 - \rho_2} = 246 \text{ minute s ,}$$

$$E(S) = \frac{1}{2}E(S_1) + \frac{1}{4}E(S_2) + \frac{1}{4}E(S_3) \approx 76.09 \text{ minutes .}$$

Exercise 58

Exercise 59. As time unit we choose 1 minute: $\lambda = 1/6$.

(i) For N , the number of parts that has to be produced for an order, we have

$$P(N = n) = \left(\frac{1}{2}\right)^n, \quad n = 1, 2, 3, \dots$$

and hence $E(N) = 2$ and $\sigma^2(N) = 2$ (see also Section 2.4.1). From $B = 2N$, it now follows that $E(B) = 4$ and $\sigma^2(B) = 8$.

(ii) Using that $\rho = 2/3$ and $E(R) = 3$, we have

$$E(S) = \frac{\rho}{1 - \rho} E(R) + E(B) = 10 \text{ minutes} .$$

(iii) We now have

$$\lambda_1 = 1/12, E(B_1) = 2, \rho_1 = 1/6 \text{ and } E(R_1) = 1,$$

$$\lambda_2 = 1/12, E(B_2) = 6, \rho_2 = 1/2 \text{ and } E(R_2) = 11/3.$$

Hence,

$$E(S_1) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{1 - \rho_1} + E(B_1) = \frac{22}{5} = 4.4 \text{ minutes} ,$$

$$E(S_2) = \frac{\rho_1 E(R_1) + \rho_2 E(R_2)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} + E(B_2) = \frac{66}{5} = 13.2 \text{ minutes} ,$$

(iv) $E(S) = \frac{1}{2}E(S_1) + \frac{1}{2}E(S_2) = \frac{44}{5} = 8.8 \text{ minutes.}$

Exercise 59

Exercise 63.

$$E(S) = \frac{\rho}{1 - \rho} \cdot E(R_B) + \frac{T}{T + \frac{1}{\lambda} e^{-\lambda T}} \cdot \frac{T}{2} + E(B).$$

Exercise 63

Exercise 64.

As time unit we choose 1 second:

$$\lambda = 1/6.$$

- (i) The mean waiting time satisfies

$$E(W) = 2.5 + E(L^q) \cdot 5.$$

Together with Little's formula, $E(L^q) = \lambda E(W)$, this yields $E(W) = 15$ seconds.

- (ii) The time elapsing from entering the carrier till the departure of that bin is 4 cycles ($= 4 \cdot 5 = 20$ seconds) plus moving out of the carrier ($= 2$ seconds), so 22 seconds. Hence, the mean sojourn time is equal to $15 + 22 = 37$ seconds.

Exercise 64

Exercise 65.

(i) The mean number of orders in the system is given by

$$E(L) = \frac{\rho}{1 - \rho} + \frac{N - 1}{2}.$$

(ii) From Little's formula we obtain

$$E(S) = \frac{1/\mu}{1 - \rho} + \frac{N - 1}{2\lambda}.$$

(iii) The average cost (setup cost + machine cost + waiting cost) per minute equals

$$\frac{6}{N} + 8 + \left(4 + \frac{3}{2}(N - 1)\right) = \frac{6}{N} + \frac{21}{2} + \frac{3N}{2}.$$

(iv) $N = 2$.

Exercise 65

Exercise 66. See exercise 4 of the exam of June 21, 1999.

Exercise 66

Exercise 69. As time unit we choose 1 hour:

$$\lambda = 1, E(B) = 1/2, \rho = 1/2 \text{ and } E(R_B) = 1/2.$$

- (i) The fraction of time that the machine processes orders is $1/2$. The mean duration of a period that the machine is switched off equals 1, the mean duration of a switch-on period equals T . Hence, the mean duration of a period that the machine processes orders equals $1+T$. Hence, both the mean number of orders processed in a production cycle and the mean duration of a production cycle equals $2 + 2T$. The mean waiting time of an order equals

$$E(W) = E(L^q) \cdot 1/2 + \frac{1}{2 + 2T} \cdot T + \frac{T}{2 + 2T} \cdot \frac{T}{2} + \frac{1+T}{2 + 2T} \cdot 1/2.$$

Together, with Little's formula $E(L^q) = 1 \cdot E(W)$ this gives

$$E(W) = \frac{T^2 + 3T + 1}{2 + 2T}.$$

Hence, the mean production lead time of an order equals

$$E(S) = \frac{T^2 + 4T + 2}{2 + 2T}.$$

- (ii) The average cost per hour equals

$$\frac{17}{2 + 2T} + \frac{T^2 + 3T + 1}{2 + 2T} = \frac{T^2 + 3T + 18}{2 + 2T}$$

which is minimal for $T = 3$.

Exercise 69

Exercise 71.

(i) $E(S) = 105/2 = 52.5$ minutes .

(ii) $E(L) = 21/6.$

Exercise 71

Exercise 73. As time unit we choose 1 minute:

$$\lambda = 1/10, E(B) = 15/2, \rho = 3/4 \text{ and } E(R_B) = 35/9.$$

- (i) The fraction of time that the server serves customers is $3/4$. The mean duration of a period that the server is away equals $10 + 10 + 5 = 25$ minutes. Hence, the mean duration of a busy period equals 75 minutes.
- (ii) $75/7.5 = 10$ customers.
- (iii) The mean waiting time of a customer equals

$$E(W) = E(L^q) \cdot 15/2 + 1/10 \cdot 5 + 1/20 \cdot 5/2 + 3/4 \cdot 35/9.$$

Together, with Little's formula $E(L^q) = 1/10 \cdot E(W)$ this gives $E(W) = 121/6 = 20.17$ minutes. Hence, the mean sojourn time of a customer equals $E(S) = 27.67$ minutes.

Exercise 73

Exercise 77.

(i) For the probability that i terminals are occupied we have

$$p_i = \frac{\frac{3^i}{i!}}{\sum_{n=0}^4 \frac{3^n}{n!}} = \frac{8}{131} \frac{3^i}{i!}.$$

Hence,

$$(p_0, p_1, p_2, p_3, p_4) = \left(\frac{8}{131}, \frac{24}{131}, \frac{36}{131}, \frac{36}{131}, \frac{27}{131} \right).$$

(ii) $B(4, 3) = p_4 = \frac{27}{131} = 0.2061$.

(iii) Use the recursion (11.3):

$$B(4, 3) = 0.2061, \quad B(5, 3) = 0.11005, \quad B(6, 3) = 0.05215, \quad B(7, 3) = 0.0219.$$

So, we need at least 7 terminals.

Exercise 77

Exercise 79.

- (i) $B(6, 7.5) = 0.3615$.
- (ii) The mean profit per day equals

$$5 \cdot 110 \cdot 1.5 \cdot (1 - B(6, 7.5)) - 6 \cdot 60 = 166.7 \text{ guilders} .$$

- (iii) When the company has c cars, the mean profit per day equals

$$5 \cdot 110 \cdot 1.5 \cdot (1 - B(c, 7.5)) - c \cdot 60 \text{ guilders} .$$

So, if the company buys 1 extra car, the mean profit becomes 174.7 guilders, if the company buys 2 extra cars, it becomes 173.8 guilders, if the company buys 3 extra cars, it becomes 163.4 guilders, and so on. The mean profit per day is maximized when the company buys 1 extra car.

Exercise 79

5. Mean number of jobs in the system: $E[n] = \rho + \rho^2(1 + C_s^2)/[2(1 - \rho)]$ This equation is known as the Pollaczek-Khinchin (P-K) mean-value formula. Note that the mean number in the queue grows linearly with the variance of the service time distribution.

6. Variance of number of jobs in the system:

$$\text{Var}[n] = E[n] + \lambda^2 \text{Var}[s] + \frac{\lambda^3 E[s^3]}{3(1-\rho)} + \frac{\lambda^4 (E[s^2])^2}{4(1-\rho)^2}$$

7. Mean number of jobs in the queue: $E[n] = \rho^2(1 + C_s^2)/[2(1 - \rho)]$

8. Variance of number of jobs in the queue: $\text{Var}[n_q] = \text{Var}[n] - \rho + \rho^2$

9. Mean response time:

$$E[r] = E[n]/\lambda = E[s] + \rho E[s](1 + C_s^2)/[2(1 - \rho)]$$

10. Variance of the response time: $\text{Var}[r] = \text{Var}[s] + \lambda E[s^3]/[3(1 - \rho)] + \lambda^2 (E[s^2])^2/[4(1 - \rho)^2]$

11. Mean waiting time: $E[w] = \rho E[s](1 + C_s^2)/[2(1 - \rho)]$

12. Variance of the waiting time: $\text{Var}[w] = \text{Var}[r] - \text{Var}[s]$

13. Idle time distribution: $F(I) = 1 - e^{-\lambda I}$. The idle time is exponentially distributed.

14. Mean number of jobs served in one busy period: $1/(1 - \rho)$

15. Variance of number of jobs served in one busy period: $\rho(1 - \rho) + \lambda^2 E[s^2]/(1 - \rho)^3$

16. Mean busy period duration: $E[s]/(1 - \rho)$

17. Variance of the busy period: $E[s^2]/(1 - \rho)^3 - (E[s])^2/(1 - \rho)^2$

For last come, first served (LCFS) or service in, random order (SIRO), the expressions for $E[n]$ and $E[r]$ are the same as above for FCFS. The variance expressions are different:

$$\text{Var}[r_{\text{SIRO}}] = \text{Var}[s] + \frac{2\lambda E[s^3]}{3(1-\rho)(2-\rho)} + \frac{\lambda^2(2+\rho)(E[s^2])^2}{4(1-\rho)^2(2-\rho)}$$

$$\text{Var}[r_{\text{LCFS}}] = \text{Var}[s] + \frac{\lambda E[s^3]}{3(1-\rho)^2} + \frac{\lambda^2(1+\rho)(E[s^2])^2}{4(1-\rho)^3}$$

Notice that $\text{Var}[r_{\text{FCFS}}] \leq \text{Var}[r_{\text{SIRO}}] \leq \text{Var}[r_{\text{LCFS}}]$

Box 31.6 M/G/1 Queue with Processor Sharing (PS)

1. Parameters:

λ = arrival rate in jobs per unit time

$E[s]$ = mean service time per job

2. Traffic intensity: $\rho = \lambda E[s] < 1$

3. The system is stable if the traffic intensity ρ is less than 1.

4. Probability of n jobs in the system $p_n = (1 - \rho)\rho^n$, $n = 0, 1, \dots, \infty$

5. Mean number in the system: $E[n] = \rho/(1 - \rho)$

6. Variance of the number in system: $\text{Var}[n] = \rho/(1 - \rho)^2$

7. Mean response time: $E[r] = E[s]/(1 - \rho)$

Notice that the expressions given here are the same as those for the M/M/1 queue. The distributions are however different. Processor sharing approximates round-robin scheduling with small quantum size and negligible overhead.

Box 31.7 M/D/1 Queue

1. Parameters:

λ = arrival rate in jobs per unit time

$E[s]$ = service time per job, s is constant

Substituting $E[s^k] = (E[s])^k$, $k = 2, 3, \dots$, in the results for M/G/1, we obtain the results listed here for M/D/1.

2. Traffic intensity: $\rho = \lambda E[s]$
3. The system is stable if the traffic intensity is less than 1.
4. Probability of n jobs in system:

$$p_n = \begin{cases} 1 - \rho, & n = 0 \\ (1 - \rho)(e^\rho - 1), & n = 1 \\ (1 - \rho) \sum_{j=0}^{n-1} \frac{(-1)^{n-j}(j\rho)^{n-j-1}(j\rho + n - j)e^{j\rho}}{(n-j)!}, & n \geq 2 \end{cases}$$

5. Mean number of jobs in the system: $E[n] = \rho + \rho^2/[2(1 - \rho)]$
6. Variance of number of jobs in the system: $\text{Var}[n] = E[n] + \rho^3/[3(1 - \rho)] + \rho^4/[4(1 - \rho)^2]$
7. Cumulative distribution function for response time:

$$F(r) = p_n \frac{(r - nE[s])}{E[s]} + \sum_{j=0}^{n-1} p_j, \quad r \geq E[s] \text{ and } n = \left\lfloor \frac{r}{E[s]} \right\rfloor$$

8. Mean response time: $E[r] = E[s] + \rho E[s]/[2(1 - \rho)]$
9. Variance of response time: $\text{Var}[r] = \rho(E[s])^2/[3(1 - \rho)] + \rho^2(E[s])^2/[4(1 - \rho)^2]$
10. Mean number of jobs in the queue: $E[n_q] = \rho^2/[2(1 - \rho)]$
11. Variance of number of jobs in the queue:

$$\text{Var}[n_q] = \rho^2 + \frac{\rho^2}{2(1 - \rho)} + \frac{\rho^3}{3(1 - \rho)} + \frac{\rho^4}{4(1 - \rho)^2}$$

12. Mean waiting time: $E[w] = \rho E[s]/[2(1 - \rho)]$
13. Variance of waiting time: $\text{Var}[w] = \text{Var}[r]$
14. Probability of serving n jobs in one busy period:

$$P(n) = \frac{(n\rho)^{n-1}}{n!} e^{-n\rho}$$

15. The cumulative distribution function of the busy period:

$$F(b) = \sum_{j=1}^n \frac{(j\rho)^{j-1}}{j!} e^{-j\rho}, \quad n = \left\lfloor \frac{b}{E[s]} \right\rfloor$$

Here, $[x]$ is the largest integer not exceeding x .

Box 31.8 M/G/ ∞ Queue

1. Parameters:

λ = arrival rate in jobs per unit time
 $E[s]$ = mean service time per jobs

2. Traffic intensity: $\rho = \lambda E[s]$
3. The system is always stable: $\rho < \infty$ is less than 1.
4. Probability of no jobs in the system: $\rho_0 = e^{-\rho}$

$$p_n = (e^{-\rho}/n!) \rho^n, \quad n = 0, 1, \dots, \infty$$

Example 12.16

A computer handles two types of jobs. Type 1 jobs require a constant service time of 1 ms, and type 2 jobs require an exponentially distributed amount of time with mean 10 ms. Find the mean waiting time if the system operates as follows: (1) an ordinary M/G/1 system and (2) a two-priority M/G/1 system with priority given to type 1 jobs. Assume that the arrival rates of the two classes are Poisson with the same rate.

The first two moments of the service time are

$$\begin{aligned} E[\tau] &= \frac{1}{2}E[\tau_1] + \frac{1}{2}E[\tau_2] = 5.5 \\ E[\tau^2] &= \frac{1}{2}E[\tau_1^2] + \frac{1}{2}E[\tau_2^2] = \frac{1}{2}(1^2 + 2(10^2)) = 100.5. \end{aligned}$$

The traffic intensity for each class and the total traffic intensity are

$$\begin{aligned} \rho_1 &= 1 \frac{\lambda}{2}, \quad \rho_2 = 10 \frac{\lambda}{2}, \quad \text{and} \\ \rho &= \lambda E[\tau] = 5.5\lambda, \end{aligned}$$

where λ is the total arrival rate. The mean residual service time is then

$$E[R] = \frac{\lambda E[\tau^2]}{2} = 50.25\lambda.$$

From Eq. (12.92), the mean waiting time for an M/G/1 system is

$$E[W] = \frac{E[R]}{1 - \rho} = \frac{50.25\lambda}{1 - 5.5\lambda}. \quad (12.107)$$

For the priority system we have

$$E[W_1] = \frac{E[R]}{1 - \rho_1} = \frac{50.25\lambda}{1 - 0.5\lambda} \quad (12.108)$$

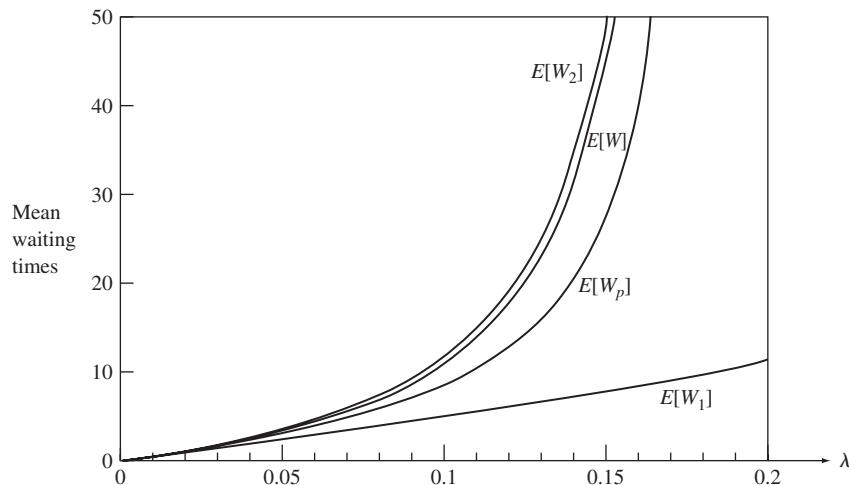
and

$$E[W_2] = \frac{E[R]}{(1 - \rho_1)(1 - \rho)} = \frac{50.25\lambda}{(1 - 0.5\lambda)(1 - 5.5\lambda)}. \quad (12.109)$$

Comparison of Eqs. (12.108) and (12.109) with Eq. (12.107) shows that the waiting time of type 1 customers is improved by a factor of $(1 - \rho)/(1 - \rho_1)$ and that of type 2 is worsened by the factor $1/(1 - \rho_1)$.

The overall mean waiting for the priority system is

$$\begin{aligned} E[W_p] &= \frac{1}{2}E[W_1] + \frac{1}{2}E[W_2] = \frac{1}{2}\left(\frac{E[R]}{1 - \rho_1}\right)\left(1 + \frac{1}{1 - \rho}\right) \\ &= \left(\frac{1 - \rho/2}{1 - \rho_1}\right)\left(\frac{E[R]}{1 - \rho}\right) \\ &= \frac{1 - 2.75\lambda}{1 - 0.5\lambda} E[W], \end{aligned}$$

**FIGURE 12.18**

Relative mean waiting times in priority and nonpriority M/G/1 systems: $E[W]$, mean waiting time in M/G/1 system; $E[W_1]$, $E[W_2]$, mean waiting time for type 1 and type 2 customers in priority system; $E[W_p]$, overall mean waiting time in priority system.

where $E[W]$ is the mean waiting time of the M/G/1 system without priorities. Figure 12.18 shows $E[W]$, $E[W_p]$, $E[W_1]$, and $E[W_2]$. It can be seen that the discipline “short-job type first” used here improves the average waiting time. The graphs for $E[W_1]$ and $E[W_2]$ also show that at $\lambda = 2/11$ the lower-priority queue becomes unstable but the higher-priority remains stable up to $\lambda = 2$.

12.7 M/G/1 ANALYSIS USING EMBEDDED MARKOV CHAINS

In the previous section we noted that the state of an M/G/1 queueing system is given by the number of customers in the system $N(t)$ and the residual service time of the customer in service. Suppose we observe $N(t)$ at the instants when the residual service time becomes zero (i.e., at the instants D_j when the j th service completion occurs); then all of the information relevant to the probability of future events is embodied in $N_j = N(D_j)$, *the number of customers left behind by the j th departing customer*. We will show that the sequence N_j is a discrete-time Markov chain and that the steady state pmf at customer departure instants is equal to the steady state pmf of the system at arbitrary time instants. Thus we can find the steady state pmf of $N(t)$ if we can find the steady state pmf for the chain N_j .

12.7.1 The Embedded Markov Chain

First we show that the sequence $N_j = N(D_j)$ is a Markov chain. Consider the relation between N_j and N_{j-1} . If $N_{j-1} \geq 1$, then a customer enters service immediately at time D_j , as shown in Fig. 12.19(a), and N_j equals N_{j-1} , minus the customer that is served in

If we substitute Eqs. (12.123) and (12.121b) into Eq. (12.120), we obtain the **Pollaczek–Khinchin transform equation**,

$$G_N(z) = \frac{(1 - \rho)(z - 1)\hat{\tau}(\lambda(1 - z))}{z - \hat{\tau}(\lambda(1 - z))}. \quad (12.125)$$

Note that $G_N(z)$ depends on the utilization ρ , the arrival rate λ , and the Laplace transform of the service time pdf.

Example 12.17 M/M/1 System

Use the Pollaczek–Khinchin transform formula to find the pmf for $N(t)$ for an M/M/1 system.

The Laplace transform for the pdf of an exponential service of mean $1/\mu$ is

$$\hat{\tau}(s) = \frac{\mu}{s + \mu}.$$

Thus the Pollaczek–Khinchin transform formula is

$$\begin{aligned} G_N(z) &= \frac{(1 - \rho)(z - 1)[\mu/(\lambda(1 - z) + \mu)]}{z - [\mu/(\lambda(1 - z) + \mu)]} \\ &= \frac{(1 - \rho)(z - 1)\mu}{(\lambda - \lambda z + \mu)z - \mu} = \frac{1 - \rho}{1 - \rho z}, \end{aligned}$$

where we canceled the $z - 1$ term from the numerator and denominator and noted that $\rho = \lambda/\mu$. By expanding $G_N(z)$ in a power series, we have

$$G_N(z) = \sum_{k=0}^{\infty} (1 - \rho)\rho^k z^k = \sum_{k=0}^{\infty} P[N = k]z^k,$$

which implies that the steady state pmf is

$$P[N = k] = (1 - \rho)\rho^k \quad k = 0, 1, 2, \dots,$$

which is in agreement with our previous results for the M/M/1 system.

Example 12.18 M/H₂/1 System

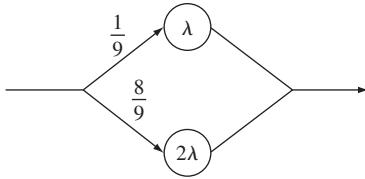
Find the pmf for the number of customers in an M/G/1 system that has arrivals of rate λ and where the service times are hyperexponential random variables of degree two, as shown in Fig. 12.20. In other words, with probability 1/9 the service time is exponentially distributed with mean $1/\lambda$, and with probability 8/9 the service time is exponentially distributed with mean $1/2\lambda$.

In order to find $\hat{\tau}(s)$ we note that the pdf of τ is

$$f_{\tau}(x) = \frac{1}{9}\lambda e^{-\lambda x} + \frac{8}{9}2\lambda e^{-2\lambda x} \quad x > 0.$$

Thus the mean service time is

$$E[\tau] = \frac{1}{9\lambda} + \frac{8}{9(2\lambda)} = \frac{5}{9\lambda},$$

**FIGURE 12.20**

A hyperexponential service time results if we select an exponential service time of rate λ with probability $1/9$ and an exponential service time of rate 2λ with probability $8/9$.

and the server utilization is $\rho = \lambda E[\tau] = 5/9$. The Laplace transform of $f_\tau(x)$ is

$$\hat{\tau}(s) = \frac{1}{9} \frac{\lambda}{s + \lambda} + \frac{8}{9} \frac{2\lambda}{s + 2\lambda} = \frac{18\lambda^2 + 17\lambda s}{9(s + \lambda)(s + 2\lambda)}.$$

Substitution of $\hat{\tau}(\lambda(1 - z))$ into Eq. (12.125) gives

$$\begin{aligned} G_N(z) &= \frac{(1 - \rho)(z - 1)(18\lambda^2 + 17\lambda^2(1 - z))}{9(\lambda - \lambda z + \lambda)(\lambda - \lambda z + 2\lambda)z - (18\lambda^2 + 17\lambda^2(1 - z))} \\ &= \frac{(1 - \rho)(z - 1)(35 - 17z)}{9(2 - z)(3 - z)z - (35 - 17z)}, \end{aligned}$$

where we have canceled λ^2 from the numerator and denominator. If we factor the denominator we obtain

$$\begin{aligned} G_N(z) &= \frac{(1 - \rho)(35 - 17z)(z - 1)}{9(z - 1)(z - 7/3)(z - 5/3)} \\ &= (1 - \rho) \left\{ \frac{1/3}{1 - 3z/7} + \frac{2/3}{1 - 3z/5} \right\}, \end{aligned}$$

where we have carried out a partial fraction expansion. Finally we note that since $G_N(z)$ converges for $|z| < 1$, we can expand $G_N(z)$ as follows:

$$G_N(z) = (1 - \rho) \left\{ \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{3}{7}\right)^k z^k + \frac{2}{3} \sum_{k=0}^{\infty} \left(\frac{3}{5}\right)^k z^k \right\}.$$

Since the coefficient of z^k is $P[N = k]$, we finally have that

$$P[N = k] = \frac{4}{27} \left(\frac{3}{7}\right)^k + \frac{8}{27} \left(\frac{3}{5}\right)^k \quad k = 0, 1, \dots,$$

where we used the fact that $\rho = 5/9$.

$$E[n_q] = \sum_{n=1}^{\infty} (n-1)p_n = \sum_{n=1}^{\infty} (n-1)(1-\rho)\rho^n = \frac{\rho^2}{1-\rho}$$

When there are no jobs in the system, the server is said to be idle; at all other times the server is busy. The time interval between two successive idle intervals is called **busy period**. All results for M/M/1 queues including some for the busy period are summarized in Box 31.1. The following example illustrates the application of these results in modeling a network gateway.

Example 31.1 On a network gateway, measurements show that the packets arrive at a mean rate of 125 packets per second (**pps**) and the gateway takes about 2 milliseconds to forward them. Using an M/M/1 model, analyze the gateway. What is the probability of buffer overflow if the gateway had only 13 buffers? How many buffers do we need to keep packet loss below one packet per million?

Arrival rate $\lambda = 125$ pps

Service rate $\mu = 1/0.002 = 500$ pps

Gateway utilization $\rho = \lambda/\mu = 0.25$

Probability of n packets in gateway $= (1 - \rho)\rho^n = 0.75(0.25)^n$

$$\frac{\rho}{1-\rho} = \frac{0.25}{0.75}$$

Mean number of packets in gateway $= \frac{1/\mu}{1-\rho} = \frac{1/500}{1-0.25}$

Mean time spent in gateway $= \frac{1/\mu}{1-\rho} = \frac{1/500}{1-0.25}$

Probability of buffer overflow $= \rho$ (more than 13 packets in gateway)

$$\rho^{13} = 0.25^{13} = 1.49 \times 10^{-8}$$

≈ 15 packets per billion packets

To limit the probability of loss to less than 10^{-6} ,

$$\rho^n \leq 10^{-6}$$

Box 31.1 M/M/1 Queue

1. Parameters:

λ = arrival rate in jobs per unit time

μ = service rate in jobs per unit time

2. Traffic intensity: $\rho = \lambda/\mu$

3. Stability condition: Traffic intensity ρ must be less than 1.

4. Probability of zero jobs in the system: $p_0 = 1 - \rho$

5. Probability of n jobs in the system: $p_n = (1 - \rho)\rho^n, n = 0, 1, \dots, \infty$

6. Mean number of jobs in the system: $E[n] = \rho/(1 - \rho)$

7. Variance of number of jobs in the system: $\text{Var}[n] = \rho/(1 - \rho)^2$

8. Probability of k jobs in the queue:

$$P(n_q = k) = \begin{cases} 1 - \rho^2, & k = 0 \\ (1 - \rho)\rho^{k+1}, & k > 0 \end{cases}$$

9. Mean number of jobs in the queue:

$$E[n_q] = \rho^2/(1 - \rho)$$

10. Variance of number of jobs in the queue:

$$\text{Var}[n_q] = \rho^2(1 + \rho - \rho^2)/(1 - \rho)^2$$

11. Cumulative distribution function of the response time:

$$F(r) = 1 - e^{-r(1-\rho)}$$

12. Mean response time: $E[r] = (1/\mu)/(1 - \rho)$

$$\frac{1/\mu^2}{(1 - \rho)^2}$$

13. Variance of the response time: $\text{Var}[r] =$

14. q -Percentile of the response time: $E[r]\ln[100/(100 - q)]$

15. 90-Percentile of the response time: $2.3E[r]$

16. Cumulative distribution function of waiting time:

$$F(w) = 1 - \rho e^{-\mu w(1-\rho)}$$

$$\rho \frac{1/\mu}{1 - \rho}$$

17. Mean waiting time: $E[w] =$

18. Variance of the waiting time: $\text{Var}[w] = (2 - \rho)\rho/[\mu^2(1 - \rho)^2]$

19. q -Percentile of the waiting time: $\max \left(0, \frac{E[w]}{\rho} \ln[100\rho/(100 - q)] \right)$

$$\left(0, \frac{E[w]}{\rho} \ln[10\rho] \right)$$

20. 90-Percentile of the waiting time: \max

21. Probability of finding n or more jobs in the system: ρ^n

$$\frac{1}{n} \binom{2n-2}{n-1} \frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$

22. Probability of serving n jobs in one busy period:

$$\frac{1}{n} \binom{2n-2}{n-1} \frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$

23. Mean number of jobs served in one busy period: $1/(1 - \rho)$

24. Variance of number of jobs served in one busy period: $\rho(1 + \rho)/(1 - \rho)^3$

25. Mean busy period duration: $1/[\mu(1 - \rho)]$

26. Variance of the busy period: $1/[\mu^2(1 - \rho)^3] - 1/[\mu^2(1 - \rho)^2]$

or

$$n > \log(10^{-6})/\log(0.25) = 9.96$$

We need about 10 buffers.

The last two results about buffer overflow are approximate. Strictly speaking, the gateway should actually be modeled as a finite buffer M/M/1/B queue. However, since the utilization is low and the number of buffers is far above the mean queue length, the results obtained are a close approximation.

Figure 31.3 shows the response time as a function of the utilization at the gateway of Example 31.1. As the rate increases, the utilization approaches 1 and the number of jobs in the system and response time approach infinity. This infinite response time is the key reason for not subjecting a server to 100% utilization. For an M/M/1 queue to be stable, the traffic intensity ρ must be less than 1.

$$F(r) = \begin{cases} 1 - e^{-\mu r} - \frac{\rho}{1-m+m\rho} e^{-m\mu(1-\rho)r} - e^{-\mu r}, & \rho \neq (m-1)/m \\ & r > 0 \\ 1 - e^{-\mu r} - \rho \mu r e^{-\mu r}, & \rho = (m-1)/m \end{cases} \quad (31.6)$$

Notice that the response time r is not exponentially distributed unless $m = 1$. In general, the coefficient of variation, that is, the ratio of the standard deviation to the mean, of r is less than 1.

Similarly, the probability distribution function of the waiting time is

$$F(w) = 1 - \rho e^{-m\mu(1-\rho)w}$$

Since w has a truncated exponential distribution function, the q -percentile can be computed as follows:

$$w_q = \max \left\{ 0, \frac{1}{m\mu(1-\rho)} \ln \left(\frac{100\rho}{100-q} \right) \right\}$$

If the probability of queueing ρ is less than $1 - q/100$, the second term in the equation can be negative. The correct answer in those cases is zero.

Example 31.2 Students arrive at the university computer center in a Poisson manner at an average rate of 10 per hour. Each student spends an average of 20 minutes at the terminal, and the time can be assumed to be exponentially distributed. The center currently has five terminals. Some students have been complaining that waiting times are too long. Let us analyze the center usage using a queueing model.

The center can be modeled as an M/M/5 queueing system with an arrival rate of $\frac{1}{6}$ per minute and a service rate of $\frac{1}{20}$ per minute.

Substituting these into the expressions previously used in this section, we get

$$\text{Traffic intensity } \rho = \frac{\lambda}{m\mu} = \frac{0.167}{5 \times 0.05} = 0.67$$

The probability of all terminals being idle is

$$\begin{aligned} p_0 &= \left[1 + \frac{(5 \times 0.67)^5}{5!(1-0.67)} + \frac{(5 \times 0.67)^1}{1!} \right. \\ &\quad \left. + \frac{(5 \times 0.67)^2}{2!} + \frac{(5 \times 0.67)^3}{3!} + \frac{(5 \times 0.67)^4}{4!} \right]^{-1} \\ &= 0.0318 \end{aligned}$$

The probability of all terminals being busy is

$$\rho = \frac{(m\rho)^m}{m!(1-\rho)} p_0 = \frac{(5 \times 0.67)^5}{5!(1-0.67)} \times 0.0318 = 0.33$$

Average terminal utilization is

$$\rho = 0.67$$

Average number of students in the center is

$$E[n] = m\rho + \frac{\rho\varrho}{1-\rho} = 5 \times 0.67 + \frac{0.67 \times 0.33}{1 - 0.67} = 4.0$$

The average number of students waiting in the queue is

$$E[n_q] = \frac{\rho\varrho}{1-\rho} = \frac{0.67 \times 0.33}{1 - 0.67} = 0.65$$

The average number of students using the terminals is

$$E[n_s] = E[n] - E[n_q] = 4 - 0.65 = 3.35$$

The mean and variance of the time spent in the center are

$$E[r] = \frac{1}{\mu} \left(1 + \frac{\varrho}{m(1-\rho)} \right) = \frac{1}{0.05} \left(1 + \frac{0.33}{5(1-0.67)} \right) = 24$$

$$\text{Var}[r] = \frac{1}{\mu^2} \left(1 + \frac{\varrho(2-\varrho)}{m^2(1-\rho)^2} \right) = \frac{1}{0.05^2} \left(1 + \frac{0.33(2-0.33)}{5^2(1-0.67)^2} \right) = 479$$

Thus, each student spends an average of 24 minutes in the center. Of these, 20 minutes are spent working on the terminal and 4 minutes are spent waiting in the queue. We can further verify this using the formula for the mean waiting time:

$$E[w] = \frac{\varrho}{m\mu(1-\rho)} = \frac{0.33}{5 \times 0.05 \times (1 - 0.67)} = 4$$

The 90-percentile of the waiting time is

$$\max \left\{ 0, \frac{E[w]}{\varrho} \ln(10\varrho) \right\} = \max \left\{ 0, \frac{4}{0.33} \ln(10 \times 0.33) \right\} = 14$$

Thus, 10% of the students have to wait more than 14 minutes.

Queueing models can be used not only to study the current behavior but also to predict what would happen if we made changes to the system. The following examples illustrate this.

Example 31.3 The students would like to limit their waiting time to an average of 2 minutes and no more than 5 minutes in 90% of the cases. Is it feasible? If yes, then how many terminals are required?

Let us analyze the system with $m = 6, 7, \dots$ terminals while keeping the same arrival and service rates of $\lambda = 0.167$ and $\mu = 0.05$, respectively.

With $m = 6$ terminals, we have

$$\text{Traffic intensity } \rho = \frac{0.167}{6 \times 0.05} = 0.556$$

$$\text{Probability of all terminals being idle} = p_0 = 0.0346$$

$$\text{Probability of all terminals being busy} = \ell = 0.15$$

$$\text{Average waiting time} = E[w] = 1.1 \text{ minutes}$$

The 90-percentile of waiting time is

$$\max \left\{ 0, \frac{1.1}{0.15} \ln(10 \times 0.15) \right\} = \max\{0, 3.0\} = 3.0$$

Thus, with just one more terminal we will be able to satisfy the students' demands.

One of the important decisions to be made when there is more than one identical server is whether to keep separate queues for each server or to keep just one queue for all servers. For Poisson arrivals and exponential service times, the first option of separate queues can be modeled using m M/M/1 queues, each with an arrival rate of λ/m . The second option of one queue can be modeled using an M/M/ m queue with an arrival rate of λ . It is easy to show that the single-queue alternative is better. We illustrate this with an example.

Example 31.4 Consider what would have happened if the five terminals in Example 31.2 were located in five different locations on the campus, thereby needing a separate queue for each.

In this case, the system can be modeled as five separate M/M/1 queues. The arrival rate for each terminal would be one-fifth of the total arrival rate. Using $m = 1$, $\lambda = 0.167/5 = 0.0333$, and $\mu = 0.05$, we have

$$\text{Traffic intensity } \rho = \frac{0.0333}{0.05} = 0.67$$

The mean time spent in the terminal room is

$$E[r] = \frac{1/\mu}{1 - \rho} = \frac{1/0.05}{1 - 0.67} = 60$$

The variance of the time spent in the terminal room is

$$\text{Var}[r] = \frac{1/\mu^2}{(1 - \rho)^2} = \frac{1/0.05^2}{(1 - 0.67)^2} = 3600$$

Compare this to the mean of 24 minutes and a variance of 479 in Example 31.2 when all five terminals were

located in one facility. It is clear that the single-queue alternative is better. In this example, we have ignored the difference in the walking time to the terminal room(s) from the dormitory. It is quite possible that having several terminal rooms distributed across the campus may reduce the walking time considerably and may become the preferred solution.

In general, if all jobs are identical, it is better to have just one queue than to have multiple queues. Of course, if some students need very short terminal sessions and others need very long sessions, this recommendation would not apply.

A special case of an $M/M/m$ queue is the $M/M/\infty$ queue with infinite servers. In such a queue, the jobs never have to wait. The response time is equal to the service time. The mean response time is equal to the mean service time regardless of the arrival rate. Such service centers are therefore also called **delay centers**. A delay center is used to represent dedicated resources, such as terminals in timesharing systems. Properties of such queues can be easily derived from those for $M/M/m$ queues. Also results presented for $M/G/\infty$ queues in Box 31.8 apply to delay centers as well.

31.4 M/M/m/B QUEUE WITH FINITE BUFFERS

An $M/M/m/B$ queue is similar to the $M/M/m$ queue except that the number of buffers B is finite. After B buffers are full, all arrivals are lost. We assume that B is greater than or equal to m ; otherwise, some servers will never be able to operate due to a lack of buffers and the system will effectively operate as an $M/M/B/B$ queue.

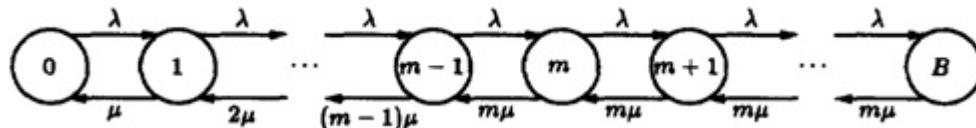


FIGURE 31.5 State transition diagram for an $M/M/m/B$ queue.

The state transition diagram for an $M/M/m/B$ queue is shown in Figure 31.5. The system can be modeled as a birth-death process using the following arrival and service rates:

$$\lambda_n = \lambda, \quad n = 0, 1, 2, \dots, B - 1$$

$$\mu_n = \begin{cases} n\mu, & n = 1, 2, \dots, m - 1 \\ m\mu, & n = m, m + 1, \dots, B \end{cases}$$

Theorem 31.1 gives us the following expression for the probability of n jobs in the system:

$$p_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} p_0, & n = 1, 2, \dots, m - 1 \\ \frac{\lambda^n}{m!m^{n-m}\mu^n} p_0, & n = m, m + 1, \dots, B \end{cases}$$

In terms of the traffic intensity $\rho = \lambda/m\mu$, we have

$$p_n = \begin{cases} \frac{(m\rho)^n}{n!} p_0, & n = 1, 2, \dots, m - 1 \\ \frac{\rho^n m^m}{m!} p_0, & n = m, m + 1, \dots, B \end{cases}$$

The probability of zero jobs in the system is computed by the relationship

$$U = \frac{\text{busy time per server}}{\text{total time}} = \frac{(\lambda' T / \mu) / m}{T} = \frac{\lambda'}{m\mu} = \rho(1 - p_B)$$

The probability of the full system is given by p_B . For a M/M/m/m system, the number of buffers is exactly equal to the number of servers, and the loss probability is

$$p_m = \frac{(m\rho)^m}{m!} p_0 = \frac{(m\rho)^m / m!}{\sum_{j=0}^m [(m\rho)^j / j!]}$$

This formula is called **Erlang's loss formula**. It was originally derived by Erlang to compute the probability of lost phone calls at a telephone exchange. It turns out that the formula is valid not only for an M/M/m/m queue but also for M/G/m/m queues.

The results for the M/M/m/B queues are summarized in Box 31.3. For the special case of a single server, many of the results can be expressed in closed forms. This special case is summarized in Box 31.4. The following example illustrates the application of these results.

Example 31.5 Consider the gateway of Example 31.1 again. Let us analyze the gateway assuming it has only two buffers. The arrival rate and the service rate, as before, are 125 pps and 500 pps, respectively. In this case

$$\begin{aligned}\lambda &= 125, & \mu &= 500, & m &= 1, & B &= 2 \\ \text{Traffic intensity } \rho &= \frac{\lambda}{m\mu} = \frac{125}{1 \times 500} = 0.25\end{aligned}$$

For $n = 1, 2, \dots, B$ are the p_n

$$p_1 = \rho p_0 = 0.25p_0$$

$$p_2 = \rho^2 p_0 = 0.25^2 p_0 = 0.0625p_0$$

Then p_0 is determined by summing all probabilities:

$$p_0 + p_1 + p_2 = 1 \Rightarrow p_0 + 0.25p_0 + 0.0625p_0 = 1 \Rightarrow p_0$$

$$= \frac{1}{1 + 0.25 + 0.0625} = 0.76$$

Substituting for p_0 in p_n , we get

$$p_1 = 0.25p_0 = 0.19$$

$$p_2 = 0.0625p_0 = 0.0476$$

The mean number of jobs in the system is

$$E[n] = \sum_{n=1}^B np_n = 1 \times 0.19 + 2 \times 0.0476 = 0.29$$

The mean number of jobs in the queue is

$$E[n_q] = \sum_{n=m}^B (n-m)p_n = (2-1) \times 0.0476 = 0.0476$$

The effective arrival rate in the system is

$$\lambda' = \lambda(1 - p_B) = 125(1 - p_2) = 125(1 - 0.0476) = 119 \text{ pps}$$

and

$$\text{Packet loss rate } \lambda - \lambda' = 125 - 119 = 6 \text{ pps}$$

Box 31.3 M/M/m/B Queue (B Buffers)

1. Parameters:

λ = arrival rate in jobs per unit time

μ = service rate in jobs per unit time

m = number of servers

B = number of buffers, $B \geq m$

2. Traffic intensity: $\rho = \lambda/(m\mu)$

3. The system is always stable: $\rho < \leq$

4. Probability of zero jobs in the system:

$$p_0 = \left[1 + \frac{(1 - \rho^{B-m+1})(m\rho)^m}{m!(1 - \rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!} \right]^{-1}$$

For $m = 1$:

$$p_0 = \begin{cases} \frac{1 - \rho}{1 - \rho^{B+1}}, & \rho \neq 1 \\ \frac{1}{B+1}, & \rho = 1 \end{cases}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} \frac{1}{n!}(m\rho)^n p_0, & 0 \leq n < m \\ \frac{m^m \rho^m}{m!} p_0, & m \leq n \leq B \end{cases}$$

6. Mean number of jobs in the system: $E[n] = \sum_{n=1}^B np_n$

For $m = 1$:

$$E[n_q] = \frac{\rho}{1-\rho} - \rho \frac{1+B\rho^B}{1-\rho^{B+1}}$$

7. Mean number of jobs in the queue: $E[n_q] = \sum_{n=m+1}^B (n-m)p_n$ For $m = 1$:

$$E[n_q] = \frac{\rho}{1-\rho} - \rho \frac{1+B\rho^B}{1-\rho^{B+1}}$$

8. Effective arrival rate in the system: $\lambda' = \sum_{n=0}^{B-1} \lambda p_n = \lambda(1-p_B)$

9. Average utilization of each server: $U = \lambda'/m\mu = \rho(1-p_B)$

10. Mean response time: $E[r] = E[n]/\lambda' = E[n]/[\lambda(1-p_B)]$

11. Mean waiting time: $E[w] = E[r] - 1/\mu = E[n_q]/[\lambda(1-p_B)]$

12. The loss rate is given by λp_B jobs per unit time.

13. For an M/M/m/m queue, the probability of a full system is given by

$$p_m = \frac{(m\rho)^m / m!}{\sum_{j=0}^m \frac{(m\rho)^j}{j!}}$$

Box 31.4 M/M/1/B Queue (B Buffers)

1. Parameters:

λ = arrival rate in jobs per unit time

μ = service rate in jobs per unit time

B = number of buffers

2. Traffic intensity: $\rho = \lambda/\mu$

3. The system is always stable: $\rho < \infty$

4. Probability of zero jobs in the system:

$$p_0 = \begin{cases} \frac{1-\rho}{1-\rho^{B+1}}, & \rho \neq 1 \\ \frac{1}{B+1}, & \rho = 1 \end{cases}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} \frac{1-\rho}{1-\rho^{B+1}} \rho^n, & \rho \neq 1 \\ \frac{1}{B+1}, & \rho = 1 \\ 0 & n > B \end{cases} \quad 0 \leq n \leq B$$

6. Mean number of jobs in the system:

$$E[n] = \frac{\rho}{1-\rho} - \frac{(B+1)\rho^{B+1}}{1-\rho^{B+1}}$$

7. Mean number of jobs in the queue:

$$E[n_q] = \frac{\rho}{1-\rho} - \rho \frac{1+B\rho^B}{1-\rho^{B+1}}$$

8. Effective arrival rate in the system: $\lambda' = \sum_{n=0}^{B-1} \lambda p_n = \lambda(1-p_B)$

9. Mean response time: $E[r] = E[n_q]/\lambda' = E[n]/[\lambda(1-p_B)]$

10. Mean waiting time: $E[w] = E[r] - 1/\mu = E[n_q]/[\lambda(1-p_B)]$

The mean response time is

$$E[r] = \frac{E[n]}{\lambda'} = \frac{0.29}{119} = 2.40 \times 10^{-3} \text{ second}$$

The mean time waiting in the queue is

$$E[w] = \frac{E[n_q]}{\lambda'} = \frac{0.0476}{119} = 4.0 \times 10^{-4} \text{ second}$$

The variance and other statistics for the number of jobs in the system can also be computed since the complete probability mass function p_n is known. For example,

$$\text{Var}[n] = E[n^2] - (E[n])^2 = (1^2 \times 0.19 + 2^2 \times 0.0476) - (0.29)^2 = 0.2963$$

31.5 RESULTS FOR OTHER QUEUEING SYSTEMS

A majority of queueing models used in computer systems performance analysis assume exponential interarrival times and exponential service times. Therefore, the M/M/m systems discussed so far cover a majority of cases. Also, systems with general arrivals or general service times are sometimes used. These include G/M/1, M/G/1, G/G/1, and G/G/m queueing systems. The key results for these systems and those for M/D/1 systems, which are a special case of M/G/1 systems, are summarized in Boxes 31.5 to 31.10. In particular, Box 31.10 for G/G/m systems summarizes the result presented earlier in Section 30.2. Readers interested in detailed derivation of results in other boxes should refer to one of several books devoted exclusively to queueing theory.

Box 31.5 M/G/1 Queue

1. Parameters:

λ = arrival rate in jobs per unit time

$E[s]$ = mean service time per job

C_s = coefficient of variation of the service time

2. Traffic intensity: $\rho = \lambda E[s]$

3. The system is stable if the traffic intensity ρ is less than 1.

4. Probability of zero jobs in the system: $\rho_0 = 1 - \rho$

7.12 A service station is staffed with two identical servers. Customers arrive according to a $\text{PP}(\lambda)$. The service times are iid with common distribution $\exp(\mu)$ at either server. Consider the following two routing policies

1. Each customer is randomly assigned to one of the two servers with equal probability.
2. Customers are alternately assigned to the two servers.

Once a customer is assigned to a server he stays in that line until served. Let $X_i(t)$ be the number of customers in line for the i th server. Is $\{X_i(t), t \geq 0\}$ the queue-length process of an $M/M/1$ queue or an $G/M/1$ queue under the two routing schemes? Identify the parameters of the queues.

7.13 Consider the following variation of an $M/G/1$ queue: All customers have iid service times with common cdf G , with mean τ_G and variance σ_G^2 . However the customers who enter an empty system have a different service time cdf H with mean τ_H and variance σ_H^2 . Let $X(t)$ be the number of customers at time t . Is $\{X(t), t \geq 0\}$ a CTMC? If yes, give its generator matrix. Let X_n be the number of customers in the system after the n th departure. Is $\{X_n, n \geq 0\}$ a DTMC? If yes, give its transition probability matrix.

7.14 Consider a communication node where packets arrive according to a $\text{PP}(\lambda)$. The node is allowed to transmit packets only at times $n = 0, 1, 2 \dots$, and transmission time of a packet is one unit of time. If a packet arrives at an empty system, it has to wait for the next transmission time to start its transmission. Let $X(t)$ be the number of packets in the system at time t , X_n be the number of packets in the system after the completion of the n th transmission, and \bar{X}_n be the number of packets available for transmission at time n . Is $\{X_n, n \geq 0\}$ a DTMC? If yes, give its transition probabilities. Is $\{\bar{X}_n, n \geq 0\}$ a DTMC? If yes, give its transition probabilities.

7.15 Suppose the customers that cannot enter an $M/M/1/1$ queue (with arrival rate λ and service rate μ) enter service at another single server queue with infinite waiting room. This second queue is called an overflow queue. The service times at the overflow queue are iid $\exp(\theta)$ random variables. Let $X(t)$ be the number of customers at the overflow queue at time t . Model the overflow queue as a $G/M/1$ queue. What is the LST of the interarrival time distribution to the overflow queue?

7.10 Computational Exercises

7.1 Show that the variance of the number of customers in steady state in a stable $M/M/1$ system with arrival rate λ and service rate μ is given by

$$\sigma^2 = \frac{\rho}{(1 - \rho)^2},$$

where $\rho = \lambda/\mu$.

7.2 Let $X^q(t)$ be the number of customers in the queue (not including any in service) at time t in an $M/M/1$ queue with arrival rate λ and service rate μ . Is $\{X^q(t), t \geq 0\}$ a CTMC? Compute the limiting distribution of $X^q(t)$ assuming $\lambda < \mu$. Show that the expected number of customers in the queue (not including the customer in service) is given by

$$L^q = \frac{\rho^2}{1 - \rho}.$$

7.3 Let W_n^q be the time spent in the queue (not including time in service) by the n th arriving customer in an $M/M/1$ queue with arrival rate λ and service rate μ . Compute the limiting distribution of W_n^q assuming $\lambda < \mu$. Compute W^q , the limiting expected value of W_n^q as $n \rightarrow \infty$. Using the results of Computational Exercise 7.2 show that $L^q = \lambda W^q$. Thus little's law holds when applied to the customers in the queue.

7.4 Let $X(t)$ be the number of customers in the system at time t in an $M/M/1$ queue with arrival rate λ and service rate $\mu > \lambda$. Let

$$T = \inf\{t \geq 0 : X(t) = 0\}.$$

T is called the busy period. Compute $E(T|X(0) = i)$.

7.5 Let T be as in Computational Exercise 7.4. Let N be the total number of customers served during $(0, T]$. Compute $E(N|X(0) = i)$.

7.6 Customers arrive according to $PP(\lambda)$ to a queueing system with two servers. The i th server ($i = 1, 2$) needs $\exp(\mu_i)$ amount of time to serve one customer. Each incoming customer is routed to server 1 with probability p_1 or to server 2 with probability $p_2 = 1 - p_1$, independently. Queue jumping is not allowed. Find the optimum routing probabilities that will minimize the expected total number of customers in the system in steady state.

7.7 Consider a stable $M/M/1$ queue with the following cost structure. A customer who sees i customers ahead of him when he joins the system costs $\$c_i$ to the system. The system charges every customer a fee of $\$f$ upon entry. Show that the long run net revenue is given by

$$\lambda(f - \sum_{i=0}^{\infty} c_i \rho^i (1 - \rho)).$$

7.8 This is a generalization of Computational Exercise 7.6. A queueing system consists of K servers, each with its own queue. Customers arrive at the system according to a $PP(\lambda)$. A system controller routes an incoming customer to server k with probability α_k , where $\alpha_1 + \alpha_2 + \dots + \alpha_K = 1$. Customers assigned to server k receive iid $\exp(\mu_k)$ service times. Assume that $\mu_1 + \mu_2 + \dots + \mu_K > \lambda$. It costs h_k dollars to hold a customer for one unit of time in queue k .

1. What are the feasible values of α_k 's so that the resulting system is stable?

2. Compute the the expected holding cost per unit time as a function of the routing probabilities α_k ($1 \leq k \leq K$) in the stable region.
3. Compute the optimal routing probabilities α_k that minimize the holding cost per unit time for the entire system.

7.9 Compute the long run fraction of customers who cannot enter the $M/M/1/K$ system described in Subsection 7.3.2.

7.10 Compute W , the expected time spent in the system by an arriving customers in steady state in an $M/M/1/K$ system, by using Little's Law and Equation 7.20. (If an arriving customer does not enter, his time in the system is zero.) What is the correct value of λ in $L = \lambda W$ as applied to this example?

7.11 Compute W , the expected waiting time of entering customers in steady state in an $M/M/1/K$ system, by using Little's Law and Equation 7.20. What is the correct value of λ in $L = \lambda W$ as applied to this example?

7.12 Suppose there are $0 < i < K$ customers in an $M/M/1/K$ queue at time 0. Compute the expected time when the queue either becomes empty or full.

7.13 Consider the $M/M/1/K$ system of Subsection 7.3.2 with the following cost structure. Each customer waiting in the system costs $\$c$ per unit time. Each customer entering the system pays $\$a$ as an entry fee to the system. Compute the long run rate of net revenue for this system.

7.14 Consider the system of Modeling Exercise 7.2 with production rate of 10 per hour and demand rate of 8 per hour. Suppose the machine is turned off when the number of items in the warehouse reaches K , and is turned on again when it falls to $K - 1$. Any demand that occurs when the warehouse is empty is lost. It costs 5 dollars to produce an item, and 1 dollar to keep an item in the warehouse for one hour. Each item sells for ten dollars.

1. Model this system as an $M/M/1/K$ queue. State the parameters.
2. Compute the long run net income (revenue-production and holding cost) per unit time, as a function of K .
3. Compute numerically the optimal K that maximizes the net income per unit time.

7.15 Consider the $M/M/1$ queue with balking (but no reneging) as described in Subsection 7.3.6. Suppose the limiting distribution of the number of customers in this queue is $\{p_j, j \geq 0\}$. Using PASTA show that in steady state an arriving customer enters the system with probability $\sum_{j=0}^{\infty} \alpha_j p_j$.

7.16 Consider the $M/M/1$ queue with balking (but no reneging) as described in Subsection 7.3.6. Suppose the limiting distribution of the number of customers in this queue is $P(\rho)$, where $\rho = \lambda/\mu$. What balking probabilities will produce this limiting distribution?

7.17 Show that the expected number of busy servers in a stable $M/M/s$ queue is λ/μ .

7.18 Derive Equation 7.21. Hence or otherwise compute the expected waiting time of a customer in the $M/M/s$ system in steady state.

7.19 Show that for a stable $M/M/s$ queue of Subsection 7.3.3

$$L^q = \frac{p_s \rho}{(1 - \rho)^2}.$$

Compute W^q explicitly and show that Little's Law $L^q = \lambda W^q$ is satisfied.

7.20 Compute the limiting distribution of the time spent in the queue by a customer in an $M/M/s$ queue. Hence or otherwise compute the limiting distribution of the time spent in the system by a customer in an $M/M/s$ queue.

7.21 Consider two queueing systems. System 1 has s servers, each serving at rate μ . System 2 has a single server, serving at rate $s\mu$. Both systems are subject to $PP(\lambda)$ arrivals. Show that in steady state, the expected number of customers in the queue (not including those in service) System 2 is less than in System 1. This shows that it is better to have a single efficient server than many inefficient ones.

7.22 Consider the finite population queue of Subsection 7.3.5 with two machines and one repairperson. Suppose every working machine produces revenue at a rate of $\$r$ per unit time. It costs $\$C$ to repair a machine. Compute the long run rate at which the system earns profits (revenue - cost).

7.23 When is the system in Modeling Exercise 7.2 stable? Assuming stability, compute the limiting distribution of the number of items in the warehouse. What fraction of the incoming demands are satisfied in steady state?

7.24 Compute the limiting distribution $\{p_i, 0 \leq i \leq s\}$ of the number of customers in an $M/M/s/s$ queue with arrival rate λ and service rate μ for each server.

7.25 The quantity p_s in the Computational Exercise 7.24 is called the blocking probability, and is denoted by $B(s, \rho)$ where $\rho = \lambda/\mu$. Show that the long run rate at which the customers enter the system is given by $\lambda(1 - B(s, \rho))$. Also, show that $B(s, \rho)$ satisfies the recursion

$$B(s, \rho) = \frac{\rho B(s - 1, \rho)}{s + \rho B(s - 1, \rho)},$$

with initial condition $B(0, \rho) = 1$.

7.26 When is the system in Modeling Exercise 7.3 stable? Assuming stability, compute the limiting distribution of the number of customers in the bank. What is the steady state probability that three tellers are active?

7.27 When is the system in Modeling Exercise 7.4 stable? Assuming stability, compute the limiting distribution of the number of customers in the system.

7.28 When is the system in Modeling Exercise 7.5 stable? Assuming stability, compute the expected number of customers in the system in steady state.

7.29 Consider the single server queue with N -type control described in Modeling Exercise 6.16. Let $X(t)$ be the number of customers in the system at time t , and $Y(t)$ be 1 if the server busy and 0 if it is idle at time t . Show that $\{(X(t), Y(t)), t \geq 0\}$ is a CTMC and that it is stable if $\rho = \lambda/\mu < 1$. Assuming it is stable, show that

$$p_{i,j} = \lim_{t \rightarrow \infty} P(X(t) = i, Y(t) = j), \quad i \geq 0, \quad j = 0, 1,$$

is given by

$$\begin{aligned} p_{i,0} &= \frac{1 - \rho}{N}, \quad 0 \leq i < N, \\ p_{i,1} &= \frac{\rho}{N}(1 - \rho^i), \quad 1 \leq i < N \\ p_{N+n,1} &= \frac{\rho}{N}(1 - \rho^N)\rho^n, \quad n \geq 0. \end{aligned}$$

7.30 Consider the queueing system of Computational Exercise 7.29. Suppose it costs $\$f$ to turn the server on from the off position, while turning the server off is free of cost. It costs $\$c$ to keep one customer in the system for one unit of time. Compute the long run operating cost per unit of the N -type policy. Show how one can optimally choose N to minimize this cost rate.

7.31 Consider the system of Modeling Exercise 6.31. What is the limiting distribution of the number of customers in the system as seen by an arriving customer of type i ? By an entering customer of type i ? ($i = 1, 2$)

7.32 Compute the limiting distribution of the CTMC in modeling Exercise 7.10 for the case of $s = 3$. What fraction of the customers are turned away in steady state?

7.33 Consider the Jackson network of single server queues as shown in Figure 7.9. Derive the stability condition. Assuming stability compute



Figure 7.9 *Queueing network for Computational Exercise 7.33.*

1. the expected number of customers in steady state in the network,
2. the fraction of the time the network is completely empty in steady state.

7.34 Do Computational Exercise 7.33 for the network in Figure 7.10.

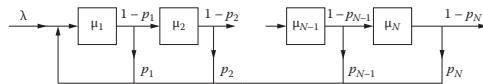


Figure 7.10 *Queueing network for Computational Exercise 7.34.*

7.35 Do Computational Exercise 7.33 for the network in Figure 7.11.

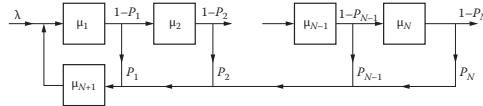


Figure 7.11 *Queueing network for Computational Exercise 7.35.*

7.36 North Carolina State Fair has 35 rides, and it expects to get about 60,000 visitors per day (12 hours) on the average. Each visitor is expected to take 5 rides on the average during his/her visit. Each ride lasts approximately 1 minute and serves an average of 30 riders per batch. Construct an approximate Jackson network model of the rides in the state fair that incorporates all the above data in a judicious fashion. State your assumptions. Is this network stable? Show how to compute the average queue length at a typical ride.

7.37 Consider a network of two nodes in series that operates as follows: customers arrive at the first node from outside according to a $\text{PP}(\lambda)$, and after completing service at node 1 move to node 2, and exit the system after completing service at node 2. The service times at each node are iid $\exp(\mu)$. Node 1 has one server active as long as there are five or fewer customers present at that node, and two servers active otherwise. Node 2 has one server active for up to two customers, two servers for three through ten customers, and three servers for any higher number. If an arriving customer sees a total of i customers at the two nodes, he joins the first node with probability $1/(i + 1)$ and leaves the system without any service with probability $i/(i + 1)$. Compute

1. the condition of stability,
2. the expected number of customers in the network in steady state.

7.38 A 30 mile long stretch of an interstate highway in Montana has no inlets or exits. This stretch is served by 3 cell towers, stationed at milepost numbers 5, 15, and 25. Each tower serves calls in the ten mile section around it. Cars enter the highway at milepost zero according to a $\text{PP}(\lambda)$, with $\lambda = 60/\text{hr}$. (Ignore the traffic in the reverse direction.) They travel at a constant speed of 100 miles per hour. Each entering car

initiates a phone call at rate $\theta = .2$ per minute, i.e., the time until the initiation of a call is an $\exp(\theta)$ random variable. The call duration is exponentially distributed with mean 10 minutes. Once the call is finished the car does not generate any new calls. (Thus each car generates at most one call.) Suppose there is enough channel capacity available that no calls are blocked. When the calling car crosses from the area of one station to the next, the call is seamlessly handed over to the next station. Model this as a Jackson network with five nodes, each having infinite servers. Node 1 is for the first tower, nodes 2 and 3 are for the second tower, and nodes 4 and 5 are for the third tower. Nodes 1, 2, and 4 handle newly initiated calls, while nodes 3 and 5 handle handed-over calls. Tower 1 does not handle any handed-over calls. Note that for infinite server nodes the service time distribution can be general. Let $X_i(t)$ be the number of calls at time t in node i , $1 \leq i \leq 5$. Compute

1. the service time distribution of the calls in node i ,
2. the routing matrix,
3. the expected number of calls handled by the i^{th} station in steady state,
4. the expected number of calls that are handed over from station i to station $i + 1$ per unit time ($i = 1, 2$).

7.39 Consider an open Jackson network with N single-server nodes. Customers arrive from outside the network to the i^{th} node with rate λ_i . A fraction p_i of the customers completing service at node i join the queue at node $i + 1$ and the rest leave the network permanently, $i = 1, 2, \dots, N - 1$. Customers completing service at node N join the queue at node 1 with probability p_N , and the rest leave the network permanently. The service times at node i are $\exp(\mu_i)$ random variables.

1. State the assumptions to model this as a Jackson network.
2. What are the traffic equations for the Jackson network? Solve them.
3. What is the condition of stability?
4. What is the expected number of customers in the network in steady state, assuming the network is stable?

7.40 Show that the probability that a customer in an open Jackson network of Section 7.4 stays in the network forever is zero if $I - R$ is invertible.

7.41 For a closed Jackson network of single server queues, show that

1. $\lim_{t \rightarrow \infty} P(X_i(t) \geq j) = \rho_i^j \frac{G_N(K)}{G_N(K-j)}$, $0 \leq j \leq K$.
2. $L_i = \lim_{t \rightarrow \infty} E(X_i(t)) = \sum_{j=1}^K \rho_i^j \frac{G_N(K)}{G_N(K-j)}$, $0 \leq j \leq K$.

7.42 Generalize the method of computing $G_N(K)$ derived in Example 7.11 to general closed Jackson networks of single-server queues with N nodes and K customers.

7.43 A simple communications network consists of two nodes labeled A and B connected by two one-way communication links: line AB from A to B , and line BA from node B to A . There are N users at each node. The i th user ($1 \leq i \leq N$) at node A (B) is denoted by A_i (B_i). User A_i has an interactive session set up with user B_i and it operates as follows: User A_i sends a message to user B_i . All the messages generated at node A wait in a buffer at node A for transmission to the appropriate user at node B on line AB in an FCFS fashion. When user B_i receives the message from user A_i , she spends a random amount of time, called think time, to generate a response to it. All the messages generated at node B wait in a buffer at node B for transmission to the appropriate user at node A on line BA in an FCFS fashion. When user A_i receives the message from user B_i , she spends a random amount of time to generate a response to it. This process of messages going back and forth between the pairs of users A_i and B_i continues forever. Suppose all the think times are iid $\exp(\theta)$ random variables, and the message transmission times are iid $\exp(\mu)$ random variables. Model this as a closed Jackson network. What is the expected number of messages in the buffers at nodes A and B in steady state?

7.44 For the closed Jackson network of Section 7.5, define the throughput $TH(i)$ of node i as the rate at which customers leave node i in steady state, i.e.,

$$TH(i) = \sum_{n=0}^K \mu_i(n) \lim_{t \rightarrow \infty} P(X_i(t) = n).$$

Show that

$$TH(i) = \pi_i \frac{G_N(K)}{G_N(K-1)}.$$

7.45 When is the system in Modeling Exercise 7.7 stable? Assuming stability, compute the expected number of customers in the system in steady state.

7.46 When is the system in Modeling Exercise 7.6 stable? Assuming stability, compute the generating function of the limiting distribution of the number of customers in the system.

7.47 Compute the expected number of customers in steady state in an $M/G/1$ system where the arrival rate is one customer per hour and the service time distribution is $PH(\alpha, M)$ where

$$\alpha = [0.5 \ 0.5 \ 0]$$

and

$$M = \begin{bmatrix} -3 & 1 & 1 \\ 0 & -3 & 2 \\ 0 & 0 & -3 \end{bmatrix}.$$

7.48 Compute the expected queue length in an $M/G/1$ queue with the following service time distributions (all with mean $1/\mu$):

1. Exponential with parameter μ ,

2. Uniform over $[0, 2/\mu]$,
3. Deterministic with mean $1/\mu$,
4. Erlang with parameters $(k, k\mu)$.

Which distribution produces the largest congestion? Which produces the smallest?

7.49 Consider the $\{X(t), t \geq 0\}$ and the $\{X_n, n \geq 0\}$ processes defined in Modeling Exercise 7.8. Show that the limiting distribution of the two (if they exist) are identical. Let p_n (q_n) be the limiting probability that there are n customers in the system and the server is up (down). Let $p(z)$ and $q(z)$ be the generating functions of $\{p_n, n \geq 0\}$ and $\{q_n, n \geq 0\}$. Show that this system is stable if

$$\frac{\lambda}{\mu} < \frac{\alpha}{\alpha + \theta}.$$

Assuming that the system is stable show that

$$q(z) = \frac{\left(\frac{\mu}{z}\right) \left(\frac{\alpha}{\alpha+\theta} - \frac{\lambda}{\mu}\right)}{\left(\frac{\mu}{z} - \lambda\right) \left(\frac{\alpha}{\theta} + \frac{\lambda}{\theta}(1-z)\right) - \lambda},$$

and

$$p(z) = \left(\frac{\alpha}{\theta} + \frac{\lambda}{\theta}(1-z)\right) q(z).$$

7.50 Show that the DTMC $\{X_n, n \geq 0\}$ in the Modeling Exercise 7.11 is positive recurrent if $\rho = \lambda\tau < 1$, where λ is the arrival rate and τ is the mean service time. Assuming the DTMC is stable, show that the generating function of the limiting distribution of X_n is given by

$$\phi(z) = \frac{1-\rho}{m} \cdot \frac{\tilde{G}(\lambda - \lambda z)}{z - \tilde{G}(\lambda - \lambda z)} \cdot (\psi(z) - 1),$$

where \tilde{G} is the LST of the service time, m is the expected number of arrivals during a single vacation, and $\psi(z)$ is the generating function of the number of arrivals during a single vacation.

7.51 Let $X(t)$ be the number of customers at time t in the system described in Modeling Exercise 7.11. Show that $\{X_n, n \geq 0\}$ and $\{X(t), t \geq 0\}$ have the same limiting distribution, assuming it exists. Using the results of Computational Exercise 7.50 show that the expected number of customers in steady state is given by

$$L = \rho + \frac{1}{2} \cdot \frac{\rho^2}{1-\rho} \left(1 + \frac{\sigma^2}{\tau^2}\right) + \frac{m^{(2)}}{2m},$$

where σ^2 is the variance of the service time, $m^{(2)}$ is the second factorial moment of the number of arrivals during a single vacation.

7.52 Let $X(t)$ be the number of customers at time t in an $M/G/1$ queue under N -type control as explained in Modeling Exercise 6.16 for an $M/M/1$ queue. Using the

results of Computational Exercises 7.50 and 7.51 establish the condition of stability for this system and compute the generating function of the limiting distribution of $X(t)$ as $t \rightarrow \infty$.

7.53 When is the queueing system described in Modeling Exercise 7.12 stable? Assuming stability, compute the expected number of customers in the system in steady state under the two policies. Which policy is better at minimizing the expected number in the system in steady state?

7.54 Analyze the stability of the $\{X(t), t \geq 0\}$ process in Modeling Exercise 7.9. Assuming stability, compute the limiting distribution of the number of items in the warehouse. What fraction of the demands are lost in steady state?

7.55 Show that the DTMC $\{X_n, n \geq 0\}$ in the Modeling Exercise 7.12 is positive recurrent if $\rho = \lambda\tau_G < 1$. Assuming the DTMC is stable, show that the generating function of the limiting distribution of X_n is given by

$$\phi(z) = \frac{1 - \lambda\tau_G}{1 - \lambda\tau_G + \lambda\tau_H} \cdot \frac{z\tilde{H}(\lambda - \lambda z) - \tilde{G}(\lambda - \lambda z)}{z - \tilde{G}(\lambda - \lambda z)}.$$

Hint: Use the results of Computational Exercise 4.24.

7.56 Let $X(t)$ be the number of customers at time t in the system described in Modeling Exercise 7.12. Show that $\{X_n, n \geq 0\}$ and $\{X(t), t \geq 0\}$ have the same limiting distribution, assuming it exists. Using the results of Computational Exercise 7.55 show that the expected number of customers in steady state is given by

$$L = \frac{\lambda\tau_H}{1 - \lambda\tau_G + \lambda\tau_H} + \frac{\lambda^2}{2} \cdot \frac{\sigma_H^2 + \tau_H^2 - \sigma_G^2 - \tau_G^2}{1 - \lambda\tau_G + \lambda\tau_H} + \frac{\lambda^2}{2} \cdot \frac{\sigma_G^2 + \tau_G^2}{1 - \lambda\tau_G}.$$

7.57 Show that the DTMC $\{X_n, n \geq 0\}$ in the Modeling Exercise 7.14 is positive recurrent if $\lambda < 1$. Assuming the DTMC is stable, compute $\phi(z)$, the generating function of the limiting distribution of X_n as $n \rightarrow \infty$.

7.58 Show that the DTMC $\{\bar{X}_n, n \geq 0\}$ in the Modeling Exercise 7.14 is positive recurrent if $\lambda < 1$. Assuming the DTMC is stable, compute $\bar{\phi}(z)$, the generating function of the limiting distribution of \bar{X}_n as $n \rightarrow \infty$.

7.59 In the Modeling Exercise 7.14, is the limiting distribution of $\{X(t), t \geq 0\}$ same as that of $\{X_n, n \geq 0\}$ or $\{\bar{X}_n, n \geq 0\}$? Explain.

7.60 Consider an $M/G/1$ queue where the customers arrive according to a $PP(\lambda)$ and request iid service times with common mean τ , and variance σ^2 . After service completion, a customer leaves with probability p , or returns to the end of the queue with probability $1 - p$, and behaves like a new customer.

1. Compute the mean and variance of the amount of time a customer spends in service during the sojourn time in the system.
2. Compute the condition of stability.
3. Assuming stability, compute the expected number of customers in the system as seen by a departure (from the system) in steady state.
4. Assuming stability, compute the expected number of customers in the system at a service completion (customer may or may not depart at each service completion) in steady state.

7.61 Compute the limiting distribution of the number of customers in a $G/M/1$ queue with the interarrival times

$$G(x) = r(1 - e^{-\lambda_1 x}) + (1 - r)(1 - e^{-\lambda_2 x}),$$

where $0 < r < 1$, $\lambda_1 > 0$, $\lambda_2 > 0$. The service times are iid $\exp(\mu)$.

7.62 Let $X(t)$ be the number of customers in a $G/M/2$ queue at time t . Let X_n^* be the number of customers as seen by the n th arrival. Show that $\{X_n^*, n \geq 0\}$ is a DTMC, and compute its one-step transition probability matrix. Derive the condition of stability and the limiting distribution of X_n^* as $n \rightarrow \infty$.

7.63 Consider the overflow queue of Modeling Exercise 7.15.

1. Compute the condition of stability for the overflow queue.
2. Assuming the overflow queue is stable, compute the pmf of the number of customers in the overflow queue in steady state.

7.64 Consider the following modification to the $M/G/1/1$ retrial queue of Section 7.7. A new customer joins the service immediately if he finds the server free upon his arrival. If the server is busy, the arriving customer leaves immediately with probability c , or joins the orbit with probability $1 - c$, and conducts retrials until he is served. Let X_n and $X(t)$ be as in Section 7.7. Derive the condition of stability and compute the generating function of the limiting distribution of X_n and $X(t)$. Are they the same?

7.65 Consider the retrial queue of Section 7.7 with $\exp(\mu)$ service times. Show that the results of Section 7.7 are consistent with those of Example 6.38.

7.66 A warehouse stocks Q items. Orders for these items arrive according to a $\text{PP}(\mu)$. The warehouse follows a $(Q, Q - 1)$ replenishment policy with back orders as follows: If the warehouse is not empty, the incoming demand is satisfied from the existing stock and an order is placed with the supplier for replenishment. If the warehouse is empty, the incoming demand is back-logged and an order is placed with the supplier for replenishment. The lead time, i.e., the amount of time it takes for the order to reach the warehouse from the supplier, is a random variable with distribution $G(\cdot)$. The lead times are iid, and orders may cross, i.e., the orders placed at the

supplier may be received out of order. Let $X(t)$ be the number of outstanding orders at time t .

1. Model $\{X(t), t \geq 0\}$ as an $M/G/\infty$ queue.
2. Compute the long run fraction of the time the warehouse is empty.

8.24 The station 1 is an $M/M/\infty$ queue with arrival rate λ and service rate μ . Hence $L_1 = \lambda/\mu$. Station i , ($2 \leq i \leq K+1$) is an $M/M/1$ queue with arrival rate λ/K and service rate θ . Hence $L_i = \rho/(1-\rho)$, where $\rho_i = \lambda/K\theta$. Finally, the queue in station $K+2$ is an $M/M/1$ queue with arrival rate $p\lambda$ and service rate α . Hence $L_{K+2} = \rho'/(1-\rho')$, where $\rho' = p\lambda/\alpha$. Hence the total number of customers in the store is

$$L = \frac{\lambda}{\mu} + \frac{K\lambda}{K\theta - \lambda} + \frac{p\lambda}{\alpha - p\lambda}.$$

SOLUTIONS TO COMPUTATIONAL PROBLEMS

8.1 $\lambda = 5$, $\tau = 1.3$. Let s be the number of servers. From Theorem 8.4, the queue is stable if $s > \lambda\tau = 5 * 1.3 = 6.5$. Hence the minimum number of servers needed for stability is 7.

8.2 From Example 8.5, the expected number of busy servers is given by $B = \min(\lambda\tau, s) = \min(6.5, s)$. Thus $B = s$ if $1 \leq s \leq 6$, and $B = 6.5$ if $s \geq 7$.

8.3 Suppose there are s servers. Assuming all servers are equally busy, this means the expected number of busy servers must be at most $.8s$. Thus $.8s > 6.5$, i.e., $s \geq 9$. Thus we need at least 9 servers.

8.4 Let L, λ, W be the usual quantities for the original system, and L', λ', W' be the corresponding quantities for the new system. Then we have $\lambda' = 2\lambda$, $W' = W$. Hence $L' = \lambda'W' = 2\lambda W = 2L$. Thus the mean number in the new system is doubled.

8.5 Doubling arrival rates and halving service times is equivalent to changing time scale. (The clocks in the new system run at twice the speed.) Thus the mean numbers are not affected. Hence L and L_q remain unchanged. Then, due to Little's Law, W and W_q are halved.

8.6 This is an $M/M/1/K$ queue with $\lambda = 8$, $\mu = 4$, $K = 4$. Hence

$$W = 2.1935 \text{ hours.}$$

8.7 The fraction of the customers lost is given by $p_4(4) = .5161$. Hence the fraction of the customers that enter is given by $1 - p_4(4) = .4839$. Hence the rate at which customers enter is $.4839 * 8 = 3.8712$ per hour. Each entering

customer pays 12 dollars. Hence the long run revenue rate is

$$3.8712 * 12 = 46.4544 \text{ dollars/hour.}$$

8.8 This is an $M/M/s/K$ queue with $\lambda = 8$, $\mu = 4$, $s = 2$, $K = 4$. Hence the new rate of revenue is given by

$$12 * \lambda * (1 - p_4(4)) = 12 * 8 * (1 - .2222) = 74.6667 \text{ dollars/hour.}$$

8.9 This is an $M/M/1/K$ queue with $\lambda = 8$, $\mu = 4$, $K = 5$. The rate of revenue is given by

$$12 * \lambda * (1 - p_5(5)) = 12 * 8 * (1 - .5079) = 47.2381. \text{ dollars/hour.}$$

The revenue with 4 chairs was computed to be 46.4544 in Computational Problem 8.7. Hence the fifth chair increases the revenue rate by .7837 dollars per hour!

8.10 This is an $M/M/1/K$ queue with $\lambda = 1$ per hour, $\mu = 20/24 = 5/3$ per hour, $K = 10$. The machine is off whenever the warehouse is full. The long run fraction of the time the machine is off is given by $p_{10}(10) = .1926$.

8.11 This is the same queue as in Computational Problem 8.10. The demands are lost when the warehouse is empty. The demands occur according to a Poisson process. Hence, according to PASTA, the long run probability that a demand sees the warehouse empty is given by $p_0(10)$. Hence the long run fraction of the demands lost are given by $p_0(10) = .0311$.

8.12 This is the same queue as in Computational Problem 8.10. Let W be the expected time an item spends in the warehouse in steady state. Then the expected revenue from the sale is $100 - W$ dollars. Using the parameters given in Computational Problem 8.10, we get $W = 8.3114$ hours. Hence the expected sale price is \$91.6886. Now the rate at which items enter the warehouse is

$$\lambda * (1 - p_K(K)) = 1 * (1 - .1926) = .8074 \text{ per hour.}$$

Hence the revenue rate is $.8074 * 91.6886 = 74.0294$ dollars/hour.

8.13 This is an $M/M/s/K$ queue with $\lambda = 60$, $\mu = 10$, $s = 6$. We need determine K such that $p_K(K) \leq .05$. We have

$$p_{10}(10) = .1286, \quad p_{11}(11) = .1140, \quad \dots, \quad p_{22}(22) = .0506, \quad p_{23}(23) = .0481.$$

Hence the system needs a total of 23 lines. It currently has 10. Hence an additional 13 need to be installed. The expected queueing time increases

from $.0246*60=1.48$ minutes to $.1290*60=7.74$ minutes.

8.14 This is an $M/M/s/K$ queue with $\lambda = 18$, $\mu = 6$, $K = 15$. Let $W_q(s)$ be the expected queueing time (in minutes) when s tellers are used. We have

$$W_q(1) = 135, \quad W_q(2) = 55.21, \quad W_q(3) = 18.72, \quad W_q(4) = 4.38.$$

Hence the bank should use 4 tellers.

8.15 This is an $M/M/s/K$ queue with $\lambda = 18$, $\mu = 6$, $K = 15$. Let $l(s)$ be the fraction lost when s tellers are used. We have

$$l(1) = .6667, \quad l(2) = .3340, \quad l(3) = .0672, \quad l(4) = .0055.$$

Hence the bank should use 4 tellers.

8.16 This is an $M/M/K/K$ queue with $\lambda = 60$, $\mu = .75$, $K = 75$. The fraction of cars turned away is given by $p_{75}(75) = .1256$.

8.17 Using the parameters given in Computational Problem 8.16 we get L . Since each car pays \$3.00 per hour, the long run revenue rate per hour is given by

$$3.00 * L = 3.00 * 69.9543 = \$209.86 \text{ per hour.}$$

8.18 Using the results for an $M/M/1$ queue with $\lambda = 15$, $\mu = 18$, we get the expected number of customers in steady state as 5. The probability that there are at least 3 customers waiting is given by $\rho^3 = .5787$.

8.19 A packet is dropped if there are no tokens in the token pool. The token pool is an $M/M/1$ queue with $\lambda = 150000$, $\mu = 200000$. Hence, the probability that there are no token in token pool is given by $p_0 = 1 - \rho = .25$. Hence 25% of the incoming packets will be dropped.

8.20 The packet buffer is an $M/M/1$ queue with $\lambda = 150000$, $\mu = 200000$. Hence, the expected waiting time is given by $W = 2 * 10^{-5}$ per second. That is the packets wait on the average 20 microseconds before being admitted into the network.

8.21 The warehouse is an $M/M/1$ queue with service rate 12 per hour. The fraction of demands lost is $1 - \rho$. Hence we must have $\rho \geq .9$, i.e., $\lambda \geq .9\mu = 10.8$ per hour. Thus, the mean production time is at the most $60/10.8 = 5.5556$ minutes. The mean number in the warehouse is then $\rho/(1 - \rho) = .9/.1 = 9$.

8.22 Under option 1, we have an $M/M/2$ queue with $\lambda = 25$, $\mu = 15$. Hence, we get $W = .2182$ hours, $W_q = .1515$ hours. Under option 2, we have an $M/M/1$ queue with $\lambda = 25$, $\mu = 30$. Hence, we get $W = .2000$ hours, $W_q = .1667$ hours. Thus option 1 minimizes W , while option 2 minimizes W_q .

8.23 Let $p(s)$ be the probability of waiting in an $M/M/s$ queue. Using the data in Example 8.10 we get

$$p(3) = .9933, \quad p(4) = .5050, \quad p(5) = .2335, \quad p(6) = .0978.$$

Hence the post office needs to keep six windows open.

8.24 Let $C(s)$ be the expected cost rate when s servers are employed. From Conceptual Problem 8.15 we have $C(s) = rs + b(\lambda/\mu) + hL$, where $r = 15$, $b = 5$, $\lambda = 20$, $\mu = 6$, $h = 60$. We need at least 4 servers to ensure stability. For $s \geq 4$ we get

$$C(4) = 473.98, \quad C(5) = 330.87, \quad C(6) = 317.78,$$

$$C(7) = 325.01, \quad C(8) = 337.66,$$

etc. Hence the cost rate is minimized at $s = 6$. Note that we expect the cost rate to keep increasing for $s \geq 8$.

8.25 From the solution to Computational Problem 8.24 we see that the cost rate is 317.78 dollars per hour if we use 6 servers. If we charge f dollars per customer, the revenue will be $20f$ dollars per hour. Hence we must have $20f \geq 317.78$, or $f \geq 15.89$ dollars.

8.26 This is an $M/M/4$ queue with $\lambda = 15$, $\mu = 5$. Hence the expected time in the system is given by $W = .3019$ hours = 18.11 minutes.

8.27 This is an $M/M/20$ queue with $\lambda = 36$, $\mu = 2$. The desired answer is $W_q = .1377$ hours = 8.26 minutes.

8.28 This is an $M/M/\infty$ queue with $\lambda = 40$, $\mu = 1/3$. Hence in steady state the number of cars in the lot is a Poisson random variable with mean $40/(1/3) = 120$.

8.29 This is an $M/M/\infty$ queue with $\lambda = 80$, $\mu = 4$. Hence in steady state the number of customers in the store is a Poisson random variable with parameter $80/4 = 20$. The probability that a $P(20)$ random variable is more than 25 is 0.1122.

8.30 This is an $M/G/1$ queue with $\lambda = 5/6$ per hour, and iid constant service times with $\tau = 1$ hour and $s^2 = 1$ (hours)². Hence, from Equation 8.39, we have $L = 5$.

8.31 A service time is 1 hour with probability .9, and $7/6$ hour with probability .1. Hence $\tau = 1*.9 + (7/6)*.1 = 1.0167$, and $s^2 = (1)^2*.9 + (7/6)^2*.1 = 1.0361$. Hence, from Equation 8.39, we have $L = 5.5511$.

8.32 The service times are iid Erl(2,12). Hence $\tau = 2/12 = 1/6$ hours, $\sigma^2 = 2/(12)^2 = 1/72$ hours². Thus $s^2 = 1/72 + 1/36 = 3/72 = 1/24$. We also have $\lambda = 3$. Hence $L = 1.1250$.

8.33 This is an $M/G/1$ queue. Let T_1 and T_2 be two iid Exp(6) random variables. Then the service time (in hours) of a single customer is $T = \max(T_1, T_2)$. The cdf of T is given by

$$F(x) = \mathbb{P}(T \leq x) = \mathbb{P}(T_1 \leq x, T_2 \leq x) = (1 - e^{-6x})^2.$$

Hence

$$\tau = \mathbb{E}(T) = \int_0^\infty x F'(x) dx = .25 \text{ hrs},$$

and

$$s^2 = \mathbb{E}(T^2) = \int_0^\infty x^2 F'(x) dx = 7/72 \text{ hrs}^2.$$

Using $\lambda = 3$ we get $L = 3.6250$. Thus the congestion is more in this setup. This is because one of the two servers is forced to be idle part of the time.

8.34 We are given $\lambda = 10$ per hour and $\tau = 1/12$ hours.

1. Exp(12): $s^2 = 2/12^2 = 1/72$ hours². Hence $W_q = .6250$ hours.
2. Uniform(0, a) has mean $a/2$. Hence we must choose $a = 1/6$. Hence $s^2 = a^2/12 + (a/2)^2 = 1/108$. and $W_q = .4861$ hours.
3. Deterministic with $\tau = 1/12$, and $s^2 = 1/144$. Hence $W_q = .4167$ hours.

Thus the deterministic distribution produces the smallest W_q and the exponential distribution produces the largest.

8.35 Let T be a typical service time. We are given

$$\mathbb{P}(T = 2) = .5, \quad \mathbb{P}(T = 3) = .2, \quad \mathbb{P}(T = 5) = .3.$$

Hence $\tau = 2*.5 + 3*.2 + 5*.3 = 3.1$ minutes and $s^2 = 4*.5 + 9*.2 + 25*.3 = 11.3$ minutes². Using $\lambda = 18$ per hour, we get $L = 14.3721$.

8.36 Consider the queue in front of server 1. It is a $G/M/1$ queue with iid Erl(2,10) inter arrival times, and Exp(6) service times. The traffic intensity

is $\rho = (10/2)/6 = 5/6 < 1$. The functional equation, using the results of Example 8.14, becomes

$$u = \left(\frac{10}{10 + 6(1-u)} \right)^2,$$

with the required solution $\alpha = .7822$. Using Equation 8.42 we get $L_1 = 3.8259$. Similarly $L_2 = 3.8259$.

Let $X(t)$ be the number of customers in the single line served by the two servers. Then $\{X(t), t \geq 0\}$ is an $M/M/2$ queue with $\lambda = 10$, $\mu = 6$. Hence $L = 5.4545$. This is less than $L_1 + L_2 = 7.6518$. Hence pooling the two servers will reduce congestion.

8.37 This is a $G/M/1$ queue with constant inter arrival times with mean 1 hour, exponential service times with mean $24/30$ hours. Hence $\lambda = 1$, $\mu = 30/24 = 1.25$. Hence the traffic intensity is $\rho = \lambda/\mu = .8$. The functional equation is $u = e^{-1.25(1-u)}$, with the solution given by $\alpha = .6286$. Using Equation 8.42 we get $L = 2.1540$.

8.38 A demand is lost if the warehouse is empty. The probability that the warehouse is empty is $1 - \rho = .2$. Since the demands occur according to a PP, this also the fraction of the demands lost. The expected amount of time an item stays in the warehouse is given by $W = 2.1540$ hours.

8.39 We are given $\mu = 10$ per hour and arrival rate $\lambda = 60/8 = 7.5$ per hour, i.e., mean inter-arrival time $8/60$ hours.

1. Exp(7.5): This is an $M/M/1$ queue. Hence $W_q = .3$ hours.
2. Deterministic with mean $8/60$. Hence the solution to the functional equation is $\alpha = .5456$. Hence $W_q = .1201$ hours.

Thus the deterministic distribution produces the smallest W_q and the exponential distribution produces the largest.

8.40 This is $G/M/1$ queue with arrival rate $\lambda = 12$ per hour (constant inter arrival times of 5 minutes), and service rate $\mu = 15$ per hour. The solution to the functional equation is $\alpha = .6286$. Hence $L = 2.1540$, and the fraction of the time the server is busy is given by $\lambda/\mu = 12/15 = .8$. Hence the cost rate is $.8 * 40 + 2.1540 * 2 = 36.3080$ dollars/hour. If each customer is charged c , the revenue rate is $12c$ per hour. Hence we must have $c \geq 36.3080/12 = 3.0257$. Thus each customer should be charged at least 3.03 dollars in order for the system to break even.

8.41 This is a tandem queue with $N = 2$, $\mu_1 = 15$, $\mu_2 = 20$, $\lambda_1 = 24$, $\lambda_2 = 0$. The solution to traffic equations is $a_1 = a_2 = 24$. Hence, for stability, we must have $s_1 \geq 2$, $s_2 \geq 2$. Since a total of 6 servers is available, we have the following three options for server allocations (s_1, s_2) : 1.(2,4), 2.(3,3), 3.(4,2). Let L_i be the expected number of customers in the entire network in steady state for option i . We get

$$L_1 = 5.6603, \quad L_2 = 3.2070, \quad L_3 = 3.5355.$$

Hence the optimal server allocation is three servers at each station.

8.42 This is a tandem queue with $N = 2$, $\mu_1 = 10$, $\mu_2 = 30$, $\lambda_1 = 24$, $\lambda_2 = 0$. The solution to traffic equations is $a_1 = a_2 = 24$. Hence, for stability, we must have $s_1 \geq 3$, $s_2 \geq 1$. Since a total of 6 servers is available, we have the following three options for server allocations (s_1, s_2) : 1.(3,3), 2.(4,2), 3.(5,1). Let L_i be the expected number of customers in the entire network in steady state for option i . We get

$$L_1 = 5.8077, \quad L_2 = 3.7829, \quad L_3 = 6.5048.$$

Hence the optimal server allocation is four servers at station 1 and 2 servers at station 2.

8.43 Suppose the arrival rate to the park is λ . Due to the symmetry of the routing matrix, the total arrival rate to each ride is also λ . Hence, for stability, each of the inequalities in Example 8.21 must be satisfied with $a_i = \lambda$ for $i = 1, 2, \dots, 6$. Hence, we must have $\lambda < 640$. Thus the maximum arrival rate has to be less than 640 for the park to be stable!

8.44 This is a Jackson Network with

$$N = 3, \quad \lambda_1 = 25, \quad \lambda_2 = \lambda_3 = 0,$$

$$\mu_1 = 20, \quad \mu_2 = 30, \quad \mu_3 = 24, \quad s_1 = s_2 = s_3 = 2,$$

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.3 & 0 & 0 \end{bmatrix},$$

$$r_1 = 0, \quad r_2 = 0, \quad r_3 = 0.7.$$

The solution to the traffic equation is

$$a = [35.7143 \quad 35.7143 \quad 35.7143].$$

1. We have

$$s_1\mu_1 = 40 > a_1, \quad s_2\mu_2 = 60 > a_2, \quad s_3\mu_3 = 48 > a_3.$$

Hence the network is stable.

2. Let $L - i$ be the expected number of customers at station i . Then we get

$$[L_1 \ L_2 \ L_3] = [8.8050 \ 1.8437 \ 3.3336].$$

Hence the total number of customers in the network is 13.9824.

3. The expected wait at station 3 is

$$W_3 = L_3/a_3 = .0933 \text{ hours} = 5.6 \text{ minutes.}$$

8.45 This is a Jackson Network with

$$N = 3, \ \lambda_1 = 20, \ \lambda_2 = \lambda_3 = 0,$$

$$\mu_1 = 12, \ \mu_2 = 20, \ \mu_3 = 15, \ s_1 = 4, \ s_2 = s_3 = 2,$$

$$P = \begin{bmatrix} 0 & .4 & .6 \\ 0.3 & 0 & 0 \\ 0.2 & 0 & 0 \end{bmatrix},$$

$$r_1 = 0, \ r_2 = 0.7, \ r_3 = 0.8.$$

The solution to the traffic equation is

$$a = [26.3158 \ 10.5263 \ 15.7895].$$

1. We have

$$s_1\mu_1 = 48 > a_1, \ s_2\mu_2 = 40 > a_2, \ s_3\mu_3 = 30 > a_3.$$

Hence the network is stable.

2. Let $L - i$ be the expected number of customers at station i . Then we get

$$[L_1 \ L_2 \ L_3] = [2.4658 \ 0.5655 \ 1.4559].$$

Hence the total number of customers in the network is 4.4872.

3. The expected wait at station 3 is

$$W_3 = L_3/a_3 = .0937 \text{ hours} = 5.62 \text{ minutes.}$$

- (iii) It is only necessary, therefore, to know λ plus the mean and variance of the service time distribution to obtain the expected value measures of effectiveness of an M/G/1 queue. These are indeed, powerful results, since this information concerning the service mechanism is usually readily available or can be estimated without great difficulty.

7.1.3 Various Formulae for (M/G/1): (∞ /GD)

These can be summarized as follows.

The other characteristic L_q , W_s and W_q of this model can be obtained by using Little's formula.

1. The average number of customers in the system,

$$L_s = \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)}, \quad \sigma^2 = \text{Var}(T), \quad \rho = \frac{\lambda}{\mu}$$

2. The average queue length,

$$L_q = \left[\frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} + \rho \right] - \rho = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)}$$

3. The average waiting time of a customer in the queue,

$$W_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2\lambda(1 - \rho)}$$

4. The average waiting time of a customer spends in the system,

$$W_s = \frac{\lambda^2 \sigma^2 + \rho^2}{2\lambda(1 - \rho)} + \frac{1}{\mu}$$

EXAMPLE 7.1 A one-man barbershop takes exactly 25 minutes to complete one haircut. If customers arrive at the barbershop in a Poisson fashion at an average rate of one every 40 minutes, how long on the average a customer spends in the shop? Also find the average time a customer must wait for service.

Solution As arrivals follow Poisson fashion and nothing is given about service time distribution, it is an M/G/1 model problem.

Service time T is constant = 25 minutes

$$\text{Service time} = \frac{1}{\mu} = 25 \text{ minutes} \Rightarrow \mu = \frac{1}{25} \text{ minute}$$

$$\therefore \text{Var}(T) = \sigma^2 = 0 \quad (\because T \text{ is constant})$$

628 ◇ Probability and Queueing Theory

Given: $\lambda = \frac{1}{40}$ (one every 40 minutes)

$$\text{and } \rho = \frac{\lambda}{\mu} = \frac{25}{40}$$

Performance measures:

The average number of customers in the shop (system). By Pollaczek–Khintchine formula,

$$\begin{aligned} L_s &= \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \\ &= \frac{25}{40} + \frac{0 + \left(\frac{25}{40}\right)^2}{2\left(1 - \frac{25}{40}\right)} = \frac{55}{48} \end{aligned}$$

Therefore, the average time a customer spends in the shop (system) by Little's formula,

$$W_s = \frac{L_s}{\lambda} \Rightarrow W_s = 40 \times \frac{55}{48} = 45.8 \text{ minutes}$$

The average time a customer must wait for service (in the queue),

$$W_q = W_s - \frac{1}{\mu} = 45.8 - 25 = 20.8 \text{ minutes}$$

Therefore, a customer has to spend 45.8 minutes in the shop and has to wait for service for 20.8 minutes on the average.

EXAMPLE 7.2 In a heavy machine shop, the overhead crane is 75% utilized. Time study observations gave the average slinging time as 10.5 minutes with a standard deviation of 8.8 minutes. (i) What is the average calling rate of the services of the crane, and (ii) What is the delay in getting service? (iii) If the average service time is cut to 8 minutes with a standard deviation of 6 minutes, how much reduction will occur on average in the delay of getting served?

[AU May '03]

Solution This is an (M/G/1): (∞ /GD) model.

$$\text{Given: } \rho = 0.75; \rho = \frac{\lambda}{\mu} \Rightarrow \lambda = \rho\mu$$

$$\begin{aligned} \text{and } \text{service time} &= \frac{1}{\mu} = 10.5 \text{ minutes} \Rightarrow \mu = \frac{1}{10.5} \text{ minute} \\ &\Rightarrow \mu = \frac{60}{10.5} = 5.71/\text{hour} \end{aligned}$$

Performance measures:

- (i) The average calling rate of the services of the crane,

$$\lambda = \rho \times \mu = 0.75 \times 5.71 = 4.29/\text{hour}$$

$$\sigma = 8.8 \text{ minutes} = \frac{8.8}{60} = 0.1467 \text{ hour}$$

- (ii) The average delay in getting service,

$$\begin{aligned} W_q &= \frac{\lambda^2 \sigma^2 + \rho^2}{2\lambda(1-\rho)} = \frac{(4.29)^2 \times (0.1467)^2 + (0.75)^2}{2 \times 4.29 \times (1 - 0.75)} \\ &= 0.449 \text{ hour} = 26.8 \text{ minutes} \end{aligned}$$

- (iii) If the service time is cut to 8 minutes with $\sigma = 6 \text{ minutes} = 6/60 = 0.1 \text{ hour}$ and $\mu = 60/8 = 7.5/\text{hour}$, then

$$\rho = \frac{4.29}{7.5} = 0.571$$

Utilization of the crane is reduced to 57.1%.

Now, the average delay in getting service is

$$\begin{aligned} W_q &= \frac{\lambda^2 \sigma^2 + \rho^2}{2\lambda(1-\rho)} \\ &= \frac{(4.29)^2 (0.1)^2 + (0.571)^2}{2 \times 4.29(1 - 0.571)} \\ &= 0.1386 \text{ hour} = 8.32 \text{ minutes} \\ \text{Reduction} &= 26.8 - 8.32 = 18.5 \text{ minutes} \end{aligned}$$

EXAMPLE 7.3 A patient who goes to a single-doctor clinic for a general check-up has to go through 4 phases. The doctor takes on the average 4 minutes for each phase of the check-up and the time taken for each phase is exponentially distributed. If the arrivals of the patients at the clinic are approximately Poisson at the average rate of 3 per hour, what is (i) the average time spent by a patient in the examination, (ii) the average time spent by a patient waiting in the clinic?

Solution Let X_1, X_2, X_3, X_4 denote the time required for the 4 phases of the check-up. X_i is exponential with mean 4 minutes. That is, $1/\lambda = 4$ or $\lambda = 1/4$.

Since the X_i is independent, $X_1 + X_2 + X_3 + X_4$ follows an Erlang's distribution with parameters λ and k .

$$\text{Given: } \lambda_1 = \frac{1}{4} \text{ and } k = 4$$

The mean and variance of Erlang's distribution are k/λ and k/λ^2 . Thus, if T represents the service time for a patient, then

$$\text{Mean} = \frac{k}{\lambda_1} = \frac{4}{\frac{1}{4}} = 16$$

$$\text{Var}(T) = \sigma^2 = \frac{k}{\lambda_1^2} = \frac{4}{\frac{1}{16}} = 64$$

Performance measures:

- (i) The average time for examination of each patient = 16 minutes
That is,

$$\text{Service time} = \frac{1}{\mu} = 16 \Rightarrow \mu = \frac{1}{16} \text{ minute}$$

If λ represents the arrival rate in the clinic, $\lambda = 3/\text{hour} = 1/20 \text{ minute}$,
then $\rho = \lambda/\mu = 16/20 = 0.8$.

- (ii) The average time spent by a patient waiting in the clinic (W_q). To find W_q we first find W_s and L_s .

$$\begin{aligned} L_s &= \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \\ &= 0.8 + \frac{(0.05)^2 \times 64 + (0.8)^2}{2 \times (1 - 0.8)} = \frac{14}{5} = 2.8 \text{ patients} \end{aligned}$$

∴ By Little's formula,

$$W_s = \frac{L_s}{\lambda_1} = 20 \times \frac{14}{5} = 56 \text{ minutes}$$

$$\text{Therefore, } W_q = W_s - \frac{1}{\mu} = 56 - 16 = 40 \text{ minutes}$$

i.e. a patient has to wait 40 minutes for check-up in the clinic.

EXAMPLE 7.4 Automatic car wash facility operates with only one bay. Cars arrive according to a Poisson distribution with a mean 4 cars per hour and may wait in the facility's parking lot if the bay is busy. The parking lot is large enough to accommodate any number of cars. If the service time for all cars is constant and equal to 10 minutes, determine (i) the mean number of customers in the system L_s , (ii) the mean number of customers in the queue L_q , (iii) the mean waiting time of a customer in the system W_s , and (iv) the mean waiting time of a customer in the queue W_q . [AU May '04; '08]

Solution This is an (M/G/1): (∞ /GD) model.

Given: $\lambda = 4 \text{ cars/hour}$

T is the service time and is constant equal to 10 minutes. Then

$$\frac{1}{\mu} = 10 \text{ minutes}$$

and

$$\text{Var}(T) = \sigma^2 = 0$$

$$\frac{1}{\mu} = 10 \Rightarrow \mu = \frac{1}{10}/\text{minute} \Rightarrow \mu = \frac{60}{10} = 6/\text{hour}$$

$$\text{Therefore, } \lambda = 4/\text{hour}, \mu = 6/\text{hour}, \sigma^2 = 0 \text{ and } \rho = \frac{\lambda}{\mu} = \frac{4}{6} = \frac{2}{3}$$

Performance measures:

- (i) The mean number of customers in the system L_s ,

$$L_s = \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} = \frac{2}{3} + \frac{0 + \left(\frac{2}{3}\right)^2}{2 \times \left[1 - \left(\frac{2}{3}\right)\right]} = 1.333 \text{ customers}$$

- (ii) The mean number of customers in the queue L_q ,

$$L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} = \frac{0 + \left(\frac{2}{3}\right)^2}{2 \times \left[1 - \left(\frac{2}{3}\right)\right]} = 0.667 \text{ customer}$$

- (iii) The mean waiting time of a customer in the system W_s ,

$$W_s = \frac{L_s}{\lambda} = \frac{1.333}{4} = 0.333 \text{ hour}$$

- (iv) The mean waiting time of a customer in the queue W_q ,

$$W_q = \frac{L_q}{\lambda} = \frac{0.667}{4} = 0.167 \text{ hour} = 10.02 \text{ minutes}$$

EXAMPLE 7.5 In a car-manufacturing plant, a loading crane takes exactly 10 minutes to load a car into a wagon and again come back to position to load another car. If the arrival of cars is a Poisson stream at an average of 1 every 20 minutes, calculate (i) the average number of cars in the system, (ii) the average number of cars in the queue, (iii) the average waiting time of a car in the system, and (iv) the average waiting time of a car in the queue.

Solution Service time general. Therefore, this is an (M/G/1): (∞ /GD) model.

Given: arrival rate $\lambda = \frac{1}{20}$ minute (1 in every 20 minutes)

Service time $T = 10$ minutes = constant

632 ◇ Probability and Queueing Theory

$$\begin{aligned} \text{Service time} &= \frac{1}{\mu} = 10 \text{ minutes} \Rightarrow \mu = \frac{1}{10} / \text{minute} \\ \therefore \rho &= \frac{\lambda}{\mu} = \frac{10}{20} = \frac{1}{2} \\ \text{and} \quad \text{Var}(T) &= \sigma^2 = 0 \end{aligned}$$

Performance measures:

- (i) The average number of cars in the system,

$$L_s = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1-\rho)} + \rho = \frac{0 + \left(\frac{1}{2}\right)^2}{2 \times \left[1 - \left(\frac{1}{2}\right)\right]} + \frac{1}{2} = \frac{3}{4} = 0.75 \text{ car} \approx 1 \text{ car}$$

- (ii) The average number of cars in the queue,

$$L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1-\rho)} = \frac{1}{4} \text{ car}$$

- (iii) The average waiting time of a car in the system,

$$W_s = \frac{L_s}{\lambda} = \frac{3}{4} \times 20 = 15 \text{ minutes}$$

- (iv) The average waiting time of a car in the queue,

$$W_q = \frac{L_q}{\lambda} = \frac{20}{4} = 5 \text{ minutes}$$

EXAMPLE 7.6 A car wash facility operates with only one bay. Cars arrive according to a Poisson distribution with a mean of 4 cars per hour and may wait in the facility's parking lot if the bay is busy. The parking lot is large enough to accommodate any number of cars. If the service time for a car has uniform distribution between 8 minutes and 12 minutes, find (i) the average number of cars waiting in the parking lot, and (ii) the average waiting time of a car in the parking lot. [AU May '04, December '07]

Solution This is an (M/G/1): (∞ /GD) model.

$$\text{Given: } \lambda = 4/\text{hour} = \frac{4}{60} / \text{minute} = \frac{1}{15} / \text{minute}$$

The service time follows uniform distribution between 8 minutes and 12 minutes.

$$\text{Service time} = \frac{1}{\mu} = \text{mean of the uniform distribution in (8, 12)}$$

Therefore, $\frac{1}{\mu} = \frac{8+12}{2} = 10 \text{ minutes} \Rightarrow \mu = \frac{1}{10} \text{/minute}$

and $\text{Var} = \sigma^2 = \frac{1}{12}(b-a)^2 = \frac{1}{12}(12-8)^2 = \frac{4}{3}$

Then, $\rho = \frac{\lambda}{\mu} = \frac{2}{3}$

Performance measures:

- (i) The average number of cars waiting in the parking lot. By P-K formula,

$$\begin{aligned} L_q &= \frac{\lambda^2 \sigma^2 + \rho^2}{2(1-\rho)} \\ &= \frac{\left(\frac{1}{225}\right) \times \left(\frac{4}{3}\right) + \left(\frac{2}{3}\right)^2}{2 \times \left[1 - \left(\frac{2}{3}\right)\right]} = 0.675 \text{ car} \end{aligned}$$

∴ The average number of cars waiting in the parking lot = 0.675 car

- (ii) The average waiting time of a car in the parking lot,

$$W_q = \frac{L_q}{\lambda} = 0.675 \times 15 = 10.125 \text{ minutes}$$

EXAMPLE 7.7 Automatic car wash facility operates with only one bay. Cars arrive according to a Poisson process with a mean rate of 4 cars per hour and may wait in the facility's parking lot if the bay is busy. If the service time follows normal distribution with mean 12 minutes and standard deviation 3 minutes, find the average number of cars waiting in the parking lot. Also find the mean waiting time of cars in the parking lot. [AU December '05]

Solution This is an (M/G/1): (∞ /GD) model.

$$\text{Given: } \lambda = 4/\text{hour} = \frac{4}{60} \text{/minute} = \frac{1}{15} \text{/minute}$$

The service time follows normal distribution with mean 12 minutes and standard deviation 3 minutes

$$\text{Service time} = \frac{1}{\mu} = 12 \text{ minutes} \Rightarrow \mu = \frac{1}{12} \text{ minute}$$

and $\text{Var}(T) = \sigma^2 = 9$ (standard deviation = 3)

634 ◇ Probability and Queueing Theory

Performance measures:

The average number of cars waiting in the parking lot.

By P-K formula,

$$\begin{aligned} L_s &= \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \\ &= \frac{12}{15} + \frac{\frac{1}{225}(9 + 144)}{2\left(1 - \frac{12}{15}\right)} = 2.5 \text{ minutes} \end{aligned}$$

The mean waiting time of cars in the parking lot.

By Little's formula,

$$L_q = L_s - \frac{\lambda}{\mu} \Rightarrow L_q = 2.5 - \frac{12}{15} = 1.7 \text{ cars}$$

The mean waiting time of cars in the parking lot,

$$W_q = \frac{L_q}{\lambda} = \frac{1.7}{\frac{1}{15}} = 25.5 \text{ minutes}$$

EXAMPLE 7.8 A car-manufacturing plant uses one big crane for loading cars into a truck. Cars arrive for loading by the crane according to a Poisson distribution with a mean of 5 cars per hour. Given that the service time for all cars is constant and equal to 6 minutes. Determine (i) the average number of customers in the system, (ii) the average number of customers in the queue, (iii) the average waiting time of a customer in the system, and (iv) the average waiting time of a customer in the queue.

Solution This is an (M/G/1): (∞ /GD) model.

Given: $\lambda = 5$ cars/hour

Service time T is constant.

$$\begin{aligned} \text{Service time } = \frac{1}{\mu} &= 6 \text{ minutes} \Rightarrow \mu = \frac{1}{6} \text{ minute} \Rightarrow \mu = \frac{60}{6} = 10/\text{hour} \\ (\because \lambda \text{ is given in hour}) \end{aligned}$$

Since the service time is constant, $\text{Var}(T) = \sigma^2 = 0$.

Therefore, $\lambda = 5/\text{hour}$, $\mu = 10/\text{hour}$, $\sigma^2 = 0$.

and

$$\rho = \frac{\lambda}{\mu} = \frac{5}{10} = \frac{1}{2}/\text{hour}$$

Performance measures:

- (i) The average number of customers in the system,

$$L_s = \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1-\rho)} = \frac{1}{2} + \frac{0 + \left(\frac{1}{2}\right)^2}{2 \times \left[1 - \left(\frac{1}{2}\right)\right]} = 0.75 \text{ customer}$$

- (ii) The average number of customers in the queue,

$$L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1-\rho)} = \frac{0 + \left(\frac{1}{2}\right)^2}{2 \times \left[1 - \left(\frac{1}{2}\right)\right]} = 0.25 \text{ customer}$$

- (iii) The average waiting time of a customer in the system,

$$W_s = \frac{L_s}{\lambda} = \frac{\frac{3}{4}}{5} = \frac{3}{20} \text{ hour} = \frac{3}{20} \times 60 = 9 \text{ minutes}$$

- (iv) The average waiting time of a customer in the queue,

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{1}{4}}{5} = \frac{1}{20} \text{ hour} = \frac{1}{20} \times 60 = 3 \text{ minutes}$$

7.2 QUEUE NETWORKS

7.2.1 Series Queues with Blocking

We consider a simple series queues with two stations S_1 and S_2 with single server at each station. No queue is allowed to form at both stations S_1 and S_2 . As the model is a sequential model, all the customers require service at S_1 and S_2 . A customer can enter the system only when S_1 is empty irrespective of whether S_2 is empty or not. After completing the service at S_1 , the customer will go to S_2 . The customer will leave from S_2 only after completing the service at S_2 .

A customer completing service at S_1 , will go to S_2 if it is empty or will wait in S_1 until S_2 becomes empty, i.e. the station S_1 is blocked for a new customer. If a customer is in process at station S_1 or if S_2 is blocked, then the arrivals (customers) are turned away.

Steady-state Probabilities

To find the steady-state probabilities $P(m, n)$ that there is m customer ($m = 0$

546 ◇ Probability and Queueing Theory

EXAMPLE 6.1 If λ, μ are the rates of arrival and departure in an M/M/1 queue respectively, give the formula for the probability that there are n customers in the queue at any time in the steady-state.

Solution The probability that there are n customers in the queue at any time in the steady-state,

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right)$$

EXAMPLE 6.2 If λ, μ are the rates of arrival and departure respectively in an M/M/1 queue, write the formulas for the average waiting time of a customer in the queue and the average number of customers in the queue in the steady-state.

Solution $L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$

and $L_s = \frac{\lambda}{(\mu - \lambda)}$

EXAMPLE 6.3 If the arrival and departure rates in a public telephone booth with a single phone are $1/12$ and $1/4$ respectively, find the probability that the phone is busy.

Solution There is only one phone.

∴ It is an M/M/1/∞ model with

$$\lambda = \frac{1}{12}$$

and $\mu = \frac{1}{4}$

$$P(\text{phone is busy}) = \rho = \frac{\lambda}{\mu} = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{1}{3}$$

Aliter

$$P[\text{phone is busy}] = 1 - P[\text{no customer in the booth}]$$

$$= 1 - \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu} = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{1}{3}$$

EXAMPLE 6.4 If the inter-arrival time and service time in a public telephone booth with a single phone follow exponential distributions with means of 10

and 8 minutes respectively, find the average number of callers in the booth at any time.

Solution Single phone is a server.

∴ It is an M/M/1/∞ model.

We know that the inter-arrival time follows exponential distribution with mean $= 1/\lambda = 10$ and service time $= 1/\mu = 8$.

$$\therefore \lambda = \frac{1}{10}$$

$$\text{and } \mu = \frac{1}{8}$$

∴ The average number of callers in the booth,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{10}}{\frac{1}{8} - \frac{1}{10}} = \frac{1}{10} \times \frac{40}{1} = 4 \text{ callers}$$

EXAMPLE 6.5 If the arrival and departure rates in an M/M/1 queue are $1/2$ per minute and $2/3$ per minute respectively, find the average waiting time of a customer in the queue.

$$\text{Solution Given: Arrival rate } \lambda = \frac{1}{2} \text{ /minute}$$

$$\text{and Departure rate } \mu = \frac{2}{3} \text{ /minute}$$

The average waiting time of a customer in the queue,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\frac{1}{2}}{\frac{2}{3}\left(\frac{2}{3} - \frac{1}{2}\right)} = 4.5 \text{ minutes}$$

EXAMPLE 6.6 Customers arrive at a railway ticket counter at the rate of 30 per hour. Assuming Poisson arrivals, exponential service time distribution and a single-server queue (M/M/1) model, find the average waiting time (before being served) if the average service time is 100 seconds.

Solution Poisson arrivals with mean $= \lambda = 30/\text{hour}$.

Exponential service time with mean $= 1/\mu = 100 \text{ seconds}$

$$\Rightarrow \mu = \frac{1}{100} \text{ second} \Rightarrow \mu = \frac{1}{100} \times 60 \times 60/\text{hour} = 36/\text{hour}$$

[since λ is given in hour, we convert μ also in hour]

548 ◇ Probability and Queueing Theory

The average waiting time in the queue,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{30}{36 \times 6} = \frac{5}{36} \text{ hour} = 8.33 \text{ minutes}$$

EXAMPLE 6.7 What is the probability that a customer has to wait more than 15 minutes to get his service completed in an M/M/1 queueing system, if $\lambda = 6$ per hour and $\mu = 10$ per hour?

Solution Given: $\lambda = 6/\text{hour}$
and $\mu = 10/\text{hour}$

The probability that the waiting time in the system exceeds 15 minutes or $15/60$ hour = $1/4$ hour (since λ and μ is given in hours),

$$P\left(W_s > \frac{1}{4}\right) = e^{-(10-6)\frac{1}{4}} = e^{-1} = 0.3679$$

EXAMPLE 6.8 Consider an M/M/1 queueing system. If $\lambda = 6$ per hour and $\mu = 8$ per hour find the probability of at least 10 customers in the system.

Solution Given: $\lambda = 6/\text{hour}$
and $\mu = 8/\text{hour}$

The probability that the number of customers in the system exceeds k ,

$$\begin{aligned} P(n \geq k) &= \left(\frac{\lambda}{\mu}\right)^k \\ P(n \geq 10) &= \left(\frac{6}{8}\right)^{10} = \left(\frac{3}{4}\right)^{10} \end{aligned}$$

EXAMPLE 6.9 Consider an M/M/1 queueing system. Find the probability of finding at least n customers in the system.

Solution The probability of at least n customers in the system,

$$\begin{aligned} P(N \geq n) &= \sum_{k=n}^{\infty} P_k = \sum_{k=n}^{\infty} P_k \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) \\ &= \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n \sum_{k=n}^{\infty} \left(\frac{\lambda}{\mu}\right)^{k-n} \\ &= \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \left(1 - \frac{\lambda}{\mu}\right)^{-1} = \left(\frac{\lambda}{\mu}\right)^n \end{aligned}$$

EXAMPLE 6.10 A duplicating machine maintained for office use is operated by office assistant. The time to complete each job varies according to an exponential distribution with mean 6 minutes. Assume a Poisson input with an average arrival rate of 5 jobs per hour. If an 8-hour day is used as a base, determine (i) the percentage of idle time of the machine, and (ii) the average time a job is in the system.

Solution Only one server.
 \therefore It is an M/M/1/ ∞ model.

Given: $\lambda = 5/\text{hour}$

and $\text{service time} = \frac{1}{\mu} = 6 \text{ minutes}$

$\Rightarrow \mu = \frac{60}{6} = 10/\text{hour}$

(i) $P[\text{the machine is idle}]$,

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{5}{10} = \frac{1}{2} = 0.5$$

Therefore, the percentage of idle time of the machine is 50%.

(ii) The average time a job is in the system,

$$L_s = \frac{1}{\mu - \lambda} = \frac{1}{10 - 5} = \frac{1}{5} \text{ hour or } 12 \text{ minutes}$$

EXAMPLE 6.11 In an (M/M/1): (∞ /FIFO) queueing model, the arrival and service rates are $\lambda = 12/\text{hour}$ and $\mu = 24/\text{hour}$. Find the average number of customers in the system and in the queue.

Solution Given: $\lambda = 12/\text{hour}$
 and $\mu = 24/\text{hour}$

The average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{12}{24 - 12} = 1 \text{ customer}$$

The average number of customers in the queue,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{144}{24 \times 12} = \frac{1}{2} \text{ customer} \approx 1$$

EXAMPLE 6.12 Customers arrive at a one-man barbershop according to a Poisson process with a mean inter-arrival time of 12 minutes. Customers spend an average of 10 minutes in the barber's chair, what is the probability that more than 3 customers are in the system?

550 ◇ Probability and Queueing Theory

Solution Only one server.

∴ It is an M/M/1/∞ model.

Inter-arrival time is exponential with mean $1/\lambda$

$$\text{Given: } \frac{1}{\lambda} = 12 \text{ minutes} \Rightarrow \lambda = \frac{1}{12} \text{ minute} = \frac{1}{12} \times 60/\text{hour} = 5/\text{hour}$$

$$\text{Service time} = \frac{1}{\mu} = 10 \text{ minutes} \Rightarrow \mu = \frac{1}{10} \text{ minute} \Rightarrow \mu = \frac{1}{10} \times 60 = 6/\text{hour}$$

The probability that more than 3 customers are in the system,

$$P(n > 3) = \left(\frac{\lambda}{\mu}\right)^{3+1} = \left(\frac{5}{6}\right)^4 = 0.4823$$

EXAMPLE 6.13 If a customer has to wait in an (M/M/1): (∞/FIFO) queue system, what is his average waiting time in the queue, if $\lambda = 8$ per hour and $\mu = 12$ per hour?

Solution Given: $\lambda = 8/\text{hour}$
and $\mu = 12/\text{hour}$

The average waiting time of a customer in the queue, if he has to wait,

$$W_q = \frac{1}{\mu - \lambda} = \frac{1}{12 - 8} = \frac{1}{4} \text{ hour} = 15 \text{ minutes}$$

EXAMPLE 6.14 Derive the average number of customers in the system for (M/M/1): (∞/FIFO) model.

Solution By definition,

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) = \left(1 - \frac{\lambda}{\mu}\right) \sum_{n=0}^{\infty} n \left(\frac{\lambda}{\mu}\right)^n \\ &= \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu}\right) \left[1 + 2\left(\frac{\lambda}{\mu}\right) + 3\left(\frac{\lambda}{\mu}\right)^2 + \dots\right] = \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu}\right) \left(1 - \frac{\lambda}{\mu}\right)^{-2} \\ &= \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1 - \rho} \end{aligned}$$

where

$$\rho = \frac{\lambda}{\mu}$$

EXAMPLE 6.15 Ms. Rose runs a one-person beauty parlour. She does not make appointments, but runs the parlour on a first come first served basis.

It is assumed that customers arrive according to a Poisson process with a mean arrival rate of 5 per hour. Because of her excellent reputation, customers were always willing to wait. The customer processing time was exponentially distributed with an average of 10 minutes. Calculate (i) the average number of customers in the system, (ii) the average number of customers in the queue, (iii) the idle time of Rose, and (iv) the average number of customers waiting when there is at least one person waiting.

Solution Given: there is only one server Ms. Rose and also no restriction on the capacity of the system. Therefore, it is an M/M/1 infinite capacity model problem.

Given: arrivals follow Poisson process with a mean rate of 5/hour.

The average or mean of Poisson distribution = λ

∴

$$\lambda = 5/\text{hour}$$

The processing time or service time follows exponential distribution with an average of 10 minutes.

The average or mean of exponential distribution = $1/\mu$

$$\therefore \frac{1}{\mu} = 10 \text{ minutes} \Rightarrow \mu = \frac{1}{10} \text{ minute}$$

$$\text{i.e. } \mu = \frac{1}{10} \times 60/\text{hour} \Rightarrow \mu = 6/\text{hour}$$

(since units should be the same for both λ and μ)

(i) The average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{5}{6 - 5} = 5 \text{ customers}$$

(ii) The average number of customers in the queue,

$$\begin{aligned} \text{Busy time} &= \frac{\lambda}{\mu} = P_0 \\ L_q &= \frac{\lambda}{\mu} \frac{\lambda}{\mu - \lambda} = \frac{5}{6} \frac{5}{6 - 5} = \frac{25}{6} = 4\frac{1}{6} \text{ customers} \end{aligned}$$

(iii) The idle time of Rose,

$$1 - P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{5}{6} = \frac{1}{6} = 0.167$$

or 16.7% of the time

(iv) The average number of customers waiting when there is at least one person waiting (non-empty queue),

$$L_q = \frac{\mu}{\mu - \lambda} = \frac{6}{6 - 5} = 6 \text{ customers}$$

EXAMPLE 6.16 A self-service store employs one cashier at its counter. Nine customers arrive on an average 5 minutes while the cashier can serve 10 customers in 5 minutes. Assuming Poisson distribution for arrival rate and exponential distribution for service rate, find (i) the average number of customers in the system, (ii) the average number of customers in queue or average queue length, (iii) the average time a customer waits before being served, and (iv) the average time a customer spends in the system.

Solution Assuming Poisson distribution for arrival rate and exponential distribution for service rate implies that it is an M/M/1 infinite capacity model.

Given: arrival rate = $\lambda = 9$ customers arrive per 5 minutes

$$= \frac{9}{5} = 1.8 \text{ customers/minute}$$

$$\text{Service rate} = \mu = \frac{10}{5} = 2 \text{ customers/minute}$$

Performance measures:

(i) The average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{1.8}{2 - 1.8} = 9 \text{ customers}$$

(ii) The average number of customers in the queue,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu} \frac{\lambda}{\mu - \lambda} = \frac{1.8}{2} \frac{1.8}{2 - 1.8} = 8.1 = 8 \text{ customers}$$

(iii) The average time a customer waits before being served (spends in the queue)

$$W_q = \frac{\lambda}{\mu} \frac{1}{\mu - \lambda} = \frac{1.8}{2} \frac{1}{2 - 1.8} = 4.5 \text{ minutes}$$

(iv) The average time a customer spends in the system,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{2 - 1.8} = \frac{1}{0.2} = 5 \text{ minutes}$$

EXAMPLE 6.17 Customers arrive at a watch repair shop according to a Poisson process at a rate of one per every 10 minutes and the service time is an exponential random variable with mean 8 minutes. Find (i) the average number of customers in the shop, (ii) the average waiting time a customer spends in the shop, and (iii) the average time a customer spends in waiting for service.

Solution One watch repair shop. The arrival follows Poisson process and the service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

For a Poisson process, the mean is λ

$$\therefore \text{Arrival rate} = \lambda = \frac{1}{10}/\text{minute} \text{ (one per every 10 minutes)}$$

For an exponential distribution, the mean is $1/\mu$.

$$\frac{1}{\mu} = 8 \text{ minutes (given)}$$

$$\therefore \mu = \frac{1}{8}/\text{minute}$$

Performance measures:

(i) The average number of customers in the shop,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{10}}{\frac{1}{8} - \frac{1}{10}} = \frac{1}{10} \times 40 = 4 \text{ customers}$$

(ii) The average waiting time a customer spends in the shop,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{8} - \frac{1}{10}} = \frac{80}{2} = 40 \text{ minutes}$$

(iii) The average time a customer spends in waiting for service,

$$\begin{aligned} W_q &= \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\frac{1}{10}}{\frac{1}{8}\left(\frac{1}{8} - \frac{1}{10}\right)} = \frac{\frac{1}{10}}{\frac{1}{8} \cdot \frac{2}{80}} \\ &= \frac{1}{10} \times \frac{8 \times 80}{2} = 32 \text{ minutes} \end{aligned}$$

EXAMPLE 6.18 A repairman is to be hired to repair machines which breakdown at an average rate of 3 per hour. The breakdowns follow Poisson distribution. Non-productive time of a machine is considered to cost ₹ 16 per hour. Two repairmen have been interviewed. One is slow but cheap, while the other is fast but expensive. The slow repairman charges ₹ 8 per hour and he services breakdown machines at the rate of 4 per hour. The fast repairman demands ₹ 10 per hour and he services at an average rate of 6 per hour. Which repairman should be hired? [AU December '04]

554 ◇ Probability and Queueing Theory

Solution One repairman is to be hired. The arrivals (breakdowns) follow Poisson distribution. Therefore, it is an M/M/1 infinite capacity model.

Given: arrival rate = $\lambda = 3/\text{hour}$
 Idle time cost of the machine = ₹ 16/hour
 For slow repairman, service rate $\mu = 4/\text{hour}$

Performance Measures:

The average downtime of the machine = the average time spent by the machine in the system for repair (non-productive time of the machine)

$$\therefore W_s = \frac{1}{\mu - \lambda} = \frac{1}{4 - 3} = 1 \text{ hour}$$

$$\therefore \text{Total cost/hour} = 1 \times 3 \times 16 + 1 \times 8 = 48 + 8 = ₹ 56$$

For fast repairman, service rate $\mu = 6/\text{hour}$

The average downtime of a machine (non-productive time of the machine),

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{6 - 3} = \frac{1}{3} \text{ hour}$$

$$\therefore \text{Total cost/hour} = \frac{1}{3} \times 3 \times 16 + 1 \times 10 = 16 + 10 = ₹ 26$$

∴ Fast repairman should be hired.

EXAMPLE 6.19 At a beauty parlour shop, with one beautician, ladies arrive according to Poisson distribution with mean arrival rate of 5 ladies per hour and hair design was exponentially distributed with an average design taking 10 minutes. As it is a very good parlour, customers do have patience to wait. Find (i) the average number of ladies in the shop, and the average number of ladies waiting to do the hair design, (ii) the probability that queue size can be greater than or equal to 1, (iii) the percentage of ladies who have to wait prior to getting into the chair for hair design, and (iv) the percentage of ladies who can walk inside the parlour without having to wait.

Solution One beauty parlour shop. The arrivals follow Poisson distribution and the service time is exponentially distributed. Therefore, it is an M/M/1 infinite capacity model.

$$\begin{aligned} \text{Arrival rate} &= \lambda = 5 \text{ ladies per hour} \\ &= \frac{1}{12} / \text{minute} \end{aligned}$$

Service time follows exponential distribution with mean $1/\mu$.

$$\therefore \frac{1}{\mu} = 10 \text{ minutes (given)}$$

$$\therefore \text{Service rate} = \mu = \frac{1}{10}/\text{minute}$$

Performance measures:

- (i) The average number of ladies in the shop (system)

$$= \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{12}}{\frac{1}{10} - \frac{1}{12}} = \frac{1}{12} \times \frac{120}{2} = 5 \text{ ladies}$$

The average number of ladies waiting to do the hair design is the average number in the queue,

$$\begin{aligned} L_q &= \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu - \lambda} \right) \\ &= 0.83 \times 5 = 4.15 \\ &= 4 \text{ ladies (approx.)} \end{aligned}$$

- (ii) The probability that queue size can be greater than or equal to 1,

$$P(n \geq 1) = 1 - P(n < 1) = 1 - P(n = 0) = 1 - \left(1 - \frac{\lambda}{\mu} \right) = \frac{\lambda}{\mu} = 0.83$$

- (iii) When there is one or more ladies in the system then the ladies entering into the system has to wait. The percentage of ladies who have to wait = 83.3%
- (iv) The percentage of ladies who can walk without waiting = $100 - 83.3 = 16.7\%$

EXAMPLE 6.20 Customers arrive at a one window drive in bank according to Poisson distribution with mean 10 per hour. Service time per customer is exponential with mean 5 minutes. The space in front of the window including that for the serviced car can accommodate a maximum of 3 cars. Other can wait outside this space. (i) What is the probability that an arriving customer can drive directly to the space in front of the window? (ii) What is the probability that an arriving customer will have to wait outside the indicated space? (iii) How long the arriving customer is expected to wait before starting service?

Solution One window drive in bank. The arrivals follow Poisson distribution and the service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

The arrivals follow Poisson distribution with mean 10/hour. Therefore, $\lambda = 10/\text{hour}$.

Service time per customer is exponential with mean $1/\mu = 5$ minutes
 $\Rightarrow \mu = 1/5$ minute

556 ◇ Probability and Queueing Theory

But λ is given per hour.

$$\therefore \mu = \frac{1}{5} \times 60 = 12 \text{ /hour}$$

Performance measures:

- (i) The probability that an arriving customer can drive directly to the space in front of the window.

It can happen only when there is no, 1 or 2 cars.

The required probability,

$$\begin{aligned} P_0 + P_1 + P_2 &= \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu} \right)^2 \right] P_0 \\ &= \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu} \right)^2 \right] \left(1 - \frac{\lambda}{\mu} \right) \\ &= \left(1 + \frac{10}{12} + \frac{100}{144} \right) \left(1 - \frac{10}{12} \right) = 0.42 \end{aligned}$$

- (ii) The probability that an arriving customer will have to wait outside the indicated space.

It can happen only when there are 3 or more cars in the indicated space.

The required probability,

$$\begin{aligned} P_3 + P_4 + P_5 + \dots &= 1 - (P_0 + P_1 + P_2) \\ &= 1 - 0.42 = 0.58 \end{aligned}$$

- (iii) The average time a customer waits before starting service (spends in the queue),

$$W_q = \frac{\lambda}{\mu} \frac{1}{\mu - \lambda} = \frac{10}{12} \frac{1}{(12 - 10)} = 0.417 \text{ hour}$$

EXAMPLE 6.21 In a supermarket, the average arrival rate of costomer is 10 in every 30 minutes following Poisson process. The average time taken by the cashier to list and calculate the customer's purchases is 2.5 minutes following exponential distribution. (i) What is the probability that the queue length exceeds 6? (ii) What is the expected time spent by a customer in the system?

Solution One supermarket. The arrivals follow Poisson distribution and the service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

$$\text{The average arrival rate } \lambda = \frac{10}{30} = \frac{1}{3} \text{ /minute}$$

Service time per customer is exponential with mean $1/\mu = 2.5$ minutes
 $\Rightarrow \mu = 1/2.5$ minute (given)

- (i) The probability that the queue length exceeds 6,

$$P(n > 6) = \left(\frac{\lambda}{\mu}\right)^6 = \left(\frac{2.5}{3}\right)^6 = 0.3348$$

- (ii) The expected time spent by a customer in the system,

$$W_q = \frac{1}{\mu - \lambda} = \frac{1}{24 - 20} = \frac{1}{4} = 15 \text{ minutes}$$

EXAMPLE 6.22 At a public telephone booth in a Post Office, arrivals are considered to be Poisson with an average inter-arrival time of 12 minutes. The length of the phone call may be assumed to be distributed exponentially with an average of 4 minutes.

- (i) What is the probability that a fresh arrival will not have to wait for the phone?
- (ii) What is the probability that an arrival will have to wait more than 10 minutes before the phone is free?
- (iii) What is the average length of queues formed from time to time?

Solution One public telephone booth. The arrivals follow Poisson distribution and the service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

Inter-arrival time of Poisson process follows exponential distribution with mean $1/\lambda$

$$\text{Inter-arrival time} = \frac{1}{\lambda} = 12 \text{ minutes} \Rightarrow \lambda = \frac{1}{12} / \text{minute}$$

The length of the phone call is the service time which follows exponential distribution with mean $1/\mu$

$$\therefore \frac{1}{\mu} = 4 \text{ minutes} \Rightarrow \mu = \frac{1}{4} / \text{minute} \quad \frac{\lambda}{\mu} = \frac{4}{12} = 0.333$$

Performance measures:

- (i) The probability that a fresh arrival will not have to wait for the phone (only when the server is idle),

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - 0.333 = 0.667$$

- (ii) The probability that an arrival will have to wait more than 10 minutes before the phone is free, i.e. the arrival has to wait for at least 10 minutes,

558 ◇ Probability and Queueing Theory

$$P(W_q > t) = \frac{\lambda}{\mu} e^{-(\mu - \lambda)t} = 0.333e^{-0.167 \times 10} = 0.0621$$

(iii) The average length of queues formed from time to time,

$$L_w = \frac{\mu}{\mu - \lambda} = \frac{0.25}{0.25 - 0.083} = 1.5$$

EXAMPLE 6.23 People arrive at a theatre ticket booth in Poisson distributed arrival rate of 25 per hour. Service time is constant at 2 minutes. Calculate (i) the mean number in the waiting line, (ii) the mean waiting time, and (iii) the utilization factor.

Solution One theatre ticket booth. The arrivals follow Poisson distribution. Therefore, it is an M/M/1 infinite capacity queueing model.

$$\text{Arrival rate} = \lambda = 25/\text{hour}$$

$$\text{Service time} = \frac{1}{\mu} = 2 \text{ minute} \Rightarrow \mu = \frac{1}{2} / \text{minute}$$

But λ is given per hour.

$$\therefore \mu = \frac{1}{2} \times 60 = 30/\text{hour}$$

Performance measures:

(i) The mean number in the waiting line (queue length),

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{25^2}{30(30 - 25)} = 4 \text{ (approx.)}$$

(ii) The mean waiting time,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{25}{150} = \frac{1}{6} \text{ hour} = \frac{1}{6} \times 60 = 10 \text{ minutes}$$

(iii) The utilization factor,

$$\rho = \frac{\lambda}{\mu} = \frac{25}{30} = 0.833$$

EXAMPLE 6.24 On average, 96 patients per 24-hour day require the service of an emergency clinic. Also on average, a patient requires 10 minutes of active attention. Assume that the facility can handle only one emergency at a time. Suppose that it costs the clinic ₹ 100 per patient treated to obtain an average servicing time of 10 minutes, and that each minute of decrease in this average time would cost ₹ 10 per patient treated. How much would have to be budgeted by the clinic to decrease the average size of the queue from $1\frac{1}{3}$ patients to $\frac{1}{2}$ patient?

[AU December '04]

Solution One emergency clinic. Assuming the arrival follows Poisson distribution, it is an M/M/1 infinite capacity model.

96 patients require the service of the clinic in 24 hours.

$$\therefore \text{In 1 hour, arrival rate } \lambda = \frac{96}{24} = 4 \text{ patients/hour}$$

Service time for 1 patient is 10 minutes and service time follows exponential distribution with mean $1/\mu$.

$$\therefore \frac{1}{\mu} = 10 \text{ minutes} \Rightarrow \mu = \frac{1}{10} \text{ minute}$$

But λ is given per hour.

$$\therefore \text{Service rate } \mu = \frac{1}{10} \times 60 = 6 \text{ /hour}$$

Performance measures:

The average number of patients in the queue,

$$L_q = \frac{\lambda}{\mu} \frac{\lambda}{\mu - \lambda} = \frac{4}{6} \frac{4}{6-4} = 1\frac{1}{3}$$

But this L_q is to be reduced from $1\frac{1}{3}$ to $1/2$

$$\text{Now, } L_q = \frac{\lambda}{\mu} \frac{\lambda}{\mu - \lambda} \Rightarrow \frac{1}{2} = \frac{4}{\mu} \frac{4}{\mu - 4}$$

$$\text{or } \mu^2 - 4\mu - 32 = 0 \Rightarrow (\mu - 8)(\mu + 4) = 0$$

$$(\mu - 8) = 0 \Rightarrow \mu = 8$$

[$\because \mu = -4$ is illogical and, hence, neglected]

\therefore The average time required by each patient,

$$\frac{1}{8} \times 60 = \frac{15}{2} \text{ minutes} = 7.5 \text{ minutes}$$

\therefore Decrease in the time required by each patient = $10 - 7.5 = 2.5$ minutes.
But, decrease in cost equal to ₹ 10/minute

\therefore The budget required for each patient = ₹ $(100 + 2.5 \times 10)$ = ₹ 125

Therefore, to decrease the size of the queue, the budget per patient should be increased from ₹ 100 to ₹ 125.

EXAMPLE 6.25 Customers arrive at the first class ticket counter of a theatre at a rate of 12 per hour. There is 1 clerk servicing the customers at the rate of 30 per hour.

560 ◇ Probability and Queueing Theory

- (i) What is the probability that there is no customer at the counter?
- (ii) What is the probability that there are more than 2 customers at the counter?
- (iii) What is the probability that there is no customer waiting to be served?
- (iv) What is the probability that a customer is being served and nobody is waiting?

Solution One ticket counter and assuming the arrivals follow Poisson, it is an M/M/1 infinite capacity model.

Given: arrival rate $\lambda = 12/\text{hour}$
 Service rate $\mu = 30/\text{hour}$

Performance measures:

- (i) The probability that there is no customer at the counter,

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{12}{30} = \frac{18}{30} = 0.6$$

- (ii) The probability that there are more than 2 customers in the counter,

$$\begin{aligned} P(n > 2) &= 1 - [P(n \leq 2)] \\ &= 1 - [P(n = 0) + P(n = 1) + P(n = 2)] \\ &= 1 - \left[P_0 + \frac{\lambda}{\mu} P_0 + \left(\frac{\lambda}{\mu} \right)^2 P_0 \right] = 1 - \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu} \right)^2 \right] P_0 \\ &= 1 - \left[1 + \frac{2}{5} + \left(\frac{2}{5} \right)^2 \right] 0.6 = 1 - 0.936 = 0.064 \end{aligned}$$

- (iii) The probability that there is no customer waiting to be served = the probability that at most 1 customer at the counter who is getting the service or no one at the counter,

$$P_0 + P_1 = P_0 + \frac{\lambda}{\mu} P_0 = 0.6 + 0.24 = 0.84$$

- (iv) The probability that a customer is being served and nobody is waiting,

$$P_1 = \frac{\lambda}{\mu} P_0 = 0.4 \times 0.6 = 0.24$$

EXAMPLE 6.26 A T.V. repairman finds that the time spent on his job has an exponential distribution with mean 30 minutes. The repaired sets arrive on an average of 10 per 8-hour day with Poisson, (i) What is the repairman's idle time each day? (ii) What is the average queue length? (iii) Also find the average number of jobs in the system.

[AU December '07]

Solution One T.V. repairman. The arrivals follow Poisson and the service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

Given: arrival rate = $\lambda = 10/8$ hours

Service time follows exponential distribution with mean $1/\mu$.

$$\therefore \text{Service time} = \frac{1}{\mu} = 30 \text{ minutes/set } \mu = \frac{1}{30} \text{ minute}$$

But λ is given in hours.

$$\therefore \mu = \frac{1}{30} \times 60 = 2/\text{hour}$$

Hence for 8 hours, the number of sets which can be repaired

$$\begin{aligned} &= 2 \times 8 \\ &= 16 \text{ per 8 hour} \end{aligned}$$

Performance measures:

$$(i) \quad \text{Repairman's busy time} = \frac{\lambda}{\mu} = \frac{10}{16} / \text{hour}$$

$$\Rightarrow \text{idle time} = 1 - \frac{10}{16} = \frac{6}{16} / \text{hour}$$

$$\therefore \text{Expected idle time per day} = 8 \times \frac{6}{16} = 3 \text{ hours}$$

(ii) The average queue length,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{10^2}{16(16 - 10)} = 1.04$$

(iii) The average number of jobs in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{10}{16 - 10} = 1.667$$

EXAMPLE 6.27 At a telephone booth, the arrivals are on the average 15 per hour. A call on the average takes 3 minutes. There is just one phone (Poisson and exponential arrival), find (i) the expected number of customers at the booth, and (ii) the idle time of the booth.

Solution One telephone booth. The arrival follows Poisson process. Assuming the service time follows exponential distribution, it is an M/M/1 infinite capacity model.

Given: $\lambda = \text{arrival rate} = 15/\text{hour}$

Service time is given and it follows exponential distribution with mean $1/\mu = 3$ minutes.

562 ◇ Probability and Queueing Theory

$$\Rightarrow \mu = \frac{1}{3} \text{ minute}$$

But λ is given in hours.

$$\therefore \mu = \frac{1}{3} \times 60 = 20 \text{ /hour}$$

Performance measures:

- (i) The expected number of customers at the booth (system),

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{15}{20 - 15} = \frac{15}{5} = 3 \text{ customers}$$

- (ii) The idle time of the booth,

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{15}{20} = \frac{5}{20} = 0.25 \text{ hour}$$

EXAMPLE 6.28 The mean rate of arrival of planes at an airport during the peak period is 20 per hour, but the actual number of arrivals in any hour follows a Poisson distribution. Sixty planes can land in the airport per hour on an average in good weather or 30 planes per hour in bad weather, but the actual number landed in any hour follows a Poisson distribution with respective averages. When there is congestion, the planes are forced to fly over the field in the stack awaiting the landing of other planes that arrived earlier.

- (i) How many planes would be flying over the field in the stack on an average in good and in bad weathers?
- (ii) How long a plane would be in the stack and in the process of landing in good and bad weathers?
- (iii) How much stack and landing time to allow so that priority to land out of order will have to be requested only 1 in 20 times?

Solution One airport. The arrivals follow Poisson distribution and the service also follows Poisson distribution. Therefore, it is an M/M/1 infinite capacity model.

Arrivals follow Poisson distribution with mean λ .

$$\begin{aligned} \therefore \text{Mean arrival rate} &= \lambda = 20/\text{hour} \\ \mu &= 60/\text{hour in good weather} \\ &= 30/\text{hour in bad weather} \end{aligned}$$

Performance measures:

- (i) The average number of planes flying over the field,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

$$\begin{aligned}
 &= \frac{20^2}{60(60-20)} = \frac{1}{6} \text{ plane (in good weather)} \\
 &= \frac{20^2}{30(30-20)} = \frac{4}{3} \text{ planes (in bad weather)}
 \end{aligned}$$

(ii) The average time for flying in the stack and for landing,

$$\begin{aligned}
 W_s &= \frac{1}{\mu - \lambda} \\
 &= \frac{1}{60-20} = \frac{1}{40} \text{ hour} = 1.5 \text{ minutes (in good weather)} \\
 &= \frac{1}{30-20} = \frac{1}{10} \text{ hour} = 6 \text{ minutes (in bad weather)}
 \end{aligned}$$

(iii) Let t be the maximum stack and landing time to be allowed, beyond which priority of order is to be requested. Then

$$P(W_s > t) = \frac{1}{20} \Rightarrow e^{-(\mu - \lambda)t} = 0.05$$

In good weather,

$$P(W_s > t) = \frac{1}{20} \Rightarrow e^{-40t} = 0.05 \Rightarrow t = 0.075 \text{ hour} = 4.5 \text{ minutes}$$

So, 4.5 minutes can be allowed for stack and landing time in good weather.

In bad weather,

$$\begin{aligned}
 P(W_s > t) &= \frac{1}{20} \Rightarrow e^{-(30-20)t} = e^{-10t} = 0.05 \\
 &\Rightarrow t = 0.299 \text{ hour} = 18 \text{ minutes}
 \end{aligned}$$

So, 18 minutes can be allowed for stack and landing time in bad weather.

EXAMPLE 6.29 A duplicating machine maintained for office use is operated by an office assistant who earns ₹ 5 per hour. The time to complete each job varies according to an exponential distribution with mean 6 minutes. Assume a Poisson input with an average arrival rate of 5 jobs per hour. If an 8-hour day is used as a base, determine (i) the percentage idle time of the machine, (ii) the average time a job is in the system, and (iii) the average earning per day of the assistant.

Solution One duplicating machine. The arrival is Poisson and the service

564 ◇ Probability and Queueing Theory

time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

Given: the input (arrival) is Poisson with mean λ

$$\therefore \text{Arrival rate} = \lambda = 5/\text{hour}$$

Service time of a job follows exponential distribution with mean $1/\mu = 6$ minutes (given)

$$\Rightarrow \mu = \frac{1}{6} \text{ minute}$$

But λ is given in hour. Therefore,

$$\mu = \frac{1}{6} \text{ minute} = \frac{1}{6} \times 60 = 10/\text{hour}$$

Performance measures:

(i) The idle time of the machine,

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{5}{10} = \frac{1}{2} = 0.5 \text{ hour}$$

\therefore The percentage of idle time of the machine = 50

(ii) The average waiting time of a job in the system,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{10 - 5} = \frac{1}{5} \text{ hour or } 12 \text{ minutes}$$

(iii) The expected (average) number of jobs (queue length) in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{5}{10 - 5} = 1$$

\therefore The average earning per day = (8-hour day and ₹ 5/hour)

$$= 8 \times 1 \times 5 = ₹ 40$$

EXAMPLE 6.30 If people arrive to purchase cinema tickets at the average rate of 6 per minute, it takes an average of 7.5 seconds to purchase a ticket. If a person arrives 2 minutes before the picture starts and if it takes exactly 1.5 minutes to reach the correct seat after purchasing the ticket, (i) can he expect to be seated for the start of the picture? (ii) what is the probability that he will be seated for the start of the picture? (iii) how early must he arrive in order to be 99% sure of being seated for the start of the picture?

Solution Customers arrive at a cinema theatre. Assuming arrival is Poisson and service time follows exponential distribution, it is an M/M/1 infinite capacity model.

Mean arrival rate = $\lambda = 6/\text{minute}$

To purchase a ticket, he takes 7.5 seconds

∴ The service time is 7.5 seconds (given)

$$\frac{1}{\mu} = 7.5 \text{ seconds}$$

$$\mu = \frac{1}{7.5} \text{ second}$$

λ is given in minutes.

$$\therefore \mu = \frac{60}{7.5} = 8/\text{minute}$$

Performance measures:

(i) The expected waiting time of a customer in the system,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{8 - 6} = \frac{1}{2} \text{ minute}$$

It takes him 1.5 minutes to reach the seat.

∴ Total time required to purchase the ticket and to reach the seat
 $= 0.5 + 1.5 = 2$ minutes.

As he arrives 2 minutes before, he can just be seated for the start of the picture.

(ii) He can be seated before the start of the picture if the waiting time in the system is less than 0.5 minute (if it takes less than 0.5 minute to purchase the ticket)

$$\begin{aligned} P(W_s < 0.5) &= 1 - P(W_s \geq 0.5) = 1 - e^{-(\mu-\lambda)0.5} \\ &= 1 - e^{-(8-6)0.5} = 1 - e^{-1} = 0.63 \end{aligned}$$

$$\begin{aligned} (\text{iii}) \quad P(W_s < t) &= 0.99 \\ \Rightarrow P(W_s > t) &= 1 - P(W_s < t) = 0.01 \\ \text{i.e. } e^{-(\mu-\lambda)t} &= 0.01 \Rightarrow -2t = \log(0.01) = -4.6 \\ t &= 2.3 \text{ minutes} \end{aligned}$$

If it takes less than 2.3 minutes to purchase the ticket and to reach the seat it takes 1.5 minutes, i.e.

$$1.5 + 2.3 = 3.8 \text{ minutes}$$

Therefore, the person must arrive at least 3.8 minutes early so as to be 99% sure of being seated for the start of the picture.

EXAMPLE 6.31 At what average rate must a clerk in a supermarket work in order to ensure a probability of 0.90 that the customer will not wait longer than 12 minutes? It is assumed that there is only one counter at which customers arrive in a Poisson fashion at an average rate of 15 per hour and that the length of the service by the clerk has an exponential distribution.

Solution Only one counter. Arrival is Poisson and service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

566 ◇ Probability and Queueing Theory

Given: customer's arrival follows Poisson fashion with mean λ .

$$\therefore \lambda = 15/\text{hour}$$

\therefore To find μ/hour (λ is given in hour)

The customer will not wait longer than 12 minutes = $12/60 = 1/5/\text{hour}$
 $P(\text{a customer will not wait for longer than } 1/5 \text{ hour in the queue}),$

$$P\left(W_q \leq \frac{1}{5}\right) = 0.90$$

$$\text{i.e. } P\left(W_q > \frac{1}{5}\right) = 0.10$$

$$\text{i.e. } P(W_q > 0.2) = 0.10$$

$$\text{But, } P(W_q > t) = \frac{\lambda}{\mu} e^{-(\mu - \lambda)t}$$

$$\therefore P(W_q > 0.2) = \frac{15}{\mu} e^{-(\mu - 15)(0.2)} = 0.1$$

$$\text{i.e. } \frac{15}{\mu} e^{-(\mu - 15) \times 0.2} = 0.1$$

$$\therefore (15 - \mu) \times 0.2 = \log(0.1) - \log 15 + \log \mu$$

$$0.2\mu + \log \mu = 3 - \log(0.006) \quad (\text{i})$$

$$0.2\mu + \log \mu - 8 = 0$$

Using Newton-Raphson method or by trial and error method, we can find the approximate value of Eq. (i) as

$$\mu = 24$$

That is, the clerk must serve at the rate of 24 customers per hour.

EXAMPLE 6.32 Customers arrive at a one-man barbershop according to a Poisson process with a mean inter-arrival time of 12 minutes. Customers spend an average of 10 minutes in the barber's chair.

- (i) What is the expected number of customers in the barbershop and in the queue?
- (ii) Calculate the percentage of time an arrival can walk straight into the barber's chair without having to wait.
- (iii) How much time can customer expect to spend in the barbershop?
- (iv) Management will provide another chair and hire another barber, when a customer's waiting time in the shop exceeds 1.25 hours. How much must the average rate of arrivals increase to warrant a second barber?
- (v) What is the average time customers spend in the queue?
- (vi) What is the probability that the waiting time in the system is greater than 30 minutes?

- (vii) Calculate the percentage of customers who have to wait prior to getting into the barber's chair.
- (viii) What is the probability that more than 3 customers are in the system?

Solution One-man barbershop. The arrival follows Poisson process. Therefore, it is an M/M/1 infinite capacity model.

Inter-arrival time of a Poisson process follows exponential distribution with mean $1/\lambda = 12$ minutes (given)

$$\therefore \frac{1}{\lambda} = 12 \text{ minutes} \Rightarrow \lambda = \frac{1}{12} / \text{minute}$$

One customer spends 10 minutes in the barber's chair, i.e. the service time is 10 minutes and the service time follows exponential distribution with mean $1/\mu = 10$ minutes

$$\therefore \mu = \frac{1}{10} \text{ minute}$$

Performance measures:

- (i) The expected number of customers in the queue,

$$L_q = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{12}}{\frac{1}{10} - \frac{1}{12}} = 5 \text{ customers}$$

The expected number of customers in the barbershop (system),

$$L_s = \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu - \lambda} \right) = \frac{\left(\frac{1}{12} \right)^2}{\frac{1}{10} \left(\frac{1}{10} - \frac{1}{12} \right)} = 4.17 \text{ customers}$$

- (ii) The probability that a customer walks straight to the barber's chair without having to wait = $P(\text{no customer in the system})$

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{\frac{1}{12}}{\frac{1}{10}} = \frac{1}{6}$$

\therefore The percentage of time an arrival need not wait = 16.7

- (iii) The expected waiting time of a customer in the barbershop (system),

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{10} - \frac{1}{12}} = 60 \text{ minutes}$$

568 ◇ Probability and Queueing Theory

- (iv) Management will provide another chair and a barber if the waiting time of a customer in the system exceeds 1.25 hours = 75 minutes, i.e.

$$W_s > 75 \Rightarrow \frac{1}{\mu - \lambda} > 75$$

i.e. $1 > 75 (\mu - \lambda) \Rightarrow \frac{1}{75} > (\mu - \lambda)$ if $\lambda > \mu - \frac{1}{75}$

i.e. if $\lambda > \frac{1}{10} - \frac{1}{75}$
 if $\lambda > \frac{13}{150}$

Hence, to hire a second barber, the average arrival rate must be increased by

$$\frac{13}{150} - \frac{1}{12} = \frac{1}{300} \text{ /minute}$$

- (v) The average time customers spend in the queue,

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\frac{1}{12}}{\frac{1}{10} \left(\frac{1}{10} - \frac{1}{12} \right)} = 50 \text{ minutes}$$

- (vi) The probability that the waiting time of a customer exceeds $t = 30$ minutes is

$$P(W > t) = e^{-(\mu - \lambda)t}$$

$$P(W > 30) = e^{-\left(\frac{1}{10} - \frac{1}{12}\right) \times 30} = 0.6065$$

- (vii) The percentage of customers who have to wait prior to getting into the barber's chair,

$$\begin{aligned} P(\text{a customer has to wait}) &= P(\text{at least one customer in the system}) \\ &= P(n > 0) \\ &= 1 - P(n = 0) \\ &= 1 - P_0 \\ &= 1 - \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu} \\ &= \frac{\frac{1}{12}}{\frac{1}{10}} = \frac{5}{6} \end{aligned}$$

∴ The percentage of customers who have to wait
 $= 5/6 \times 100 = 83.33$

(viii) The probability that there are more than 3 customers in the system is

$$P(n > 3) = \left(\frac{\lambda}{\mu}\right)^{3+1} = \left(\frac{\lambda}{\mu}\right)^4 = \left(\frac{5}{6}\right)^4 = 0.4823$$

EXAMPLE 6.33 Arrivals at a telephone booth are considered to be Poisson with an average time of 12 minutes between one arrival and the next. The length of a phone call is assumed to be distributed exponentially with mean 4 minutes

- (i) Find the average number of persons waiting in the system.
- (ii) What is the probability that a person arriving at the booth will have to wait in the queue? $P(\text{customer has to wait in the system for more than 10 minutes})$.
- (iii) What is the probability that it will take him more than 10 minutes altogether to wait for the phone and complete his call?
- (iv) Estimate the fraction of the day when the phone will be in use.
- (v) The telephone department will install a second booth, when convinced that an arrival has to wait on the average for at least 3 minutes for phone. By how much the flow of arrivals should increase in order to justify a second booth?
- (vi) What is the average length of the queue that forms from time to time?

[AU December '03, May '06]

Solution The average time between one arrival and the next is 12 minutes, i.e. the interval between two successive arrival is given as 12 minutes. Since the arrivals follow Poisson, the inter-arrival time follows exponential distribution with mean $1/\lambda$.

$$\text{Mean of exponential distribution} = \frac{1}{\lambda} = 12 \text{ minutes} \Rightarrow \lambda = \frac{1}{12} \text{ minute}$$

Given: service follows exponential distribution, therefore,

$$\text{Mean} = \frac{1}{\mu} = 4 \text{ minutes}$$

$$\therefore \text{Mean service rate} = \mu = \frac{1}{4} \text{ minute}$$

Performance measures:

- (i) The average number of persons waiting in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{12}}{\frac{1}{4} - \frac{1}{12}} = 0.5 \text{ customer}$$

570 ◇ Probability and Queueing Theory

- (ii) The probability that a person arriving at the booth will have to wait in the queue (this means that the queue is non-empty),

$$\begin{aligned} P(L_q > 0) &= 1 - P(L_q = 0) \\ &= 1 - P(\text{no customer in the system}) \\ &= 1 - P_0 \\ &= 1 - \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu} = \frac{1}{12} = \frac{1}{3} \end{aligned}$$

- (iii) The probability that the waiting time in the system exceeds 10 minutes (more than 10 minutes),

$$\begin{aligned} t &= 10 \text{ minutes} \\ P(W_s > 10) &= e^{-(\mu-\lambda)t} \\ &= e^{-\left(\frac{1}{4} - \frac{1}{12}\right) \times 10} = 0.1889 \end{aligned}$$

- (iv) The fraction of the day the phone will be in use.
The phone will be idle when there is no arrival.

$$\begin{aligned} P(\text{the phone will be idle}) &= P_0 = 1 - \frac{\lambda}{\mu} = \frac{2}{3} \\ P(\text{the phone will be in use}) &= \frac{\lambda}{\mu} = \frac{1}{3} \end{aligned}$$

∴ The fraction of the day when the phone will be in use = 1/3.

- (v) The second phone will be installed, if $W_q > 3$. That is, the arrivals have to wait on the average for at least 3 minutes, i.e.

$$\begin{aligned} \text{if } W_q &> 3 \\ \text{and if } \frac{\lambda}{\mu(\mu-\lambda)} &> 3 \\ \text{i.e. if } \frac{\lambda}{\frac{1}{4}\left(\frac{1}{4} - \lambda\right)} &> 3 \end{aligned}$$

where λ is the required arrival rate

$$\begin{aligned} \Rightarrow \lambda &> \frac{3}{4}\left(\frac{1}{4} - \lambda\right) \Rightarrow \lambda + \frac{3}{4}\lambda > \frac{3}{16} \\ \Rightarrow \frac{7}{4}\lambda &> \frac{3}{16} \Rightarrow \lambda > \frac{3}{28} \end{aligned}$$

Hence, to justify a second phone, the arrival rate should increase by

$$\frac{3}{28} - \frac{1}{12} = \frac{1}{42} \text{ /minute}$$

(vi) The average length of the queue that forms from time to time,

$$E(L_q/\text{the queue is always available}) = \frac{\mu}{\mu - \lambda} = \frac{\frac{1}{4}}{\frac{1}{4} - \frac{1}{12}} = \frac{3}{2}$$

EXAMPLE 6.34 In a railway marshalling yard, goods trains arrive at a rate of 30 trains per day. Assuming that the inter-arrival time follows an exponential distribution and the service time distribution is also exponential with an average of 36 minutes, calculate: (i) The mean queue size. (ii) The probability that the queue size exceeds 10. (iii) If the input of trains increases to an average 33 per day, what will be the change in (i) and (ii)? [AU December '07]

Solution One railway marshalling yard. The arrival and service time follow exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

The inter-arrival time of a Poisson process follows exponential distribution. So, the arrivals follow Poisson distribution with mean $\lambda = 30$ trains/day.

The service time distribution is exponential with mean $1/\mu$.

$$\therefore \text{Mean} = \frac{1}{\mu} = 36 \text{ minutes/train}$$

But λ is given per day and μ is given per minute. So. we find μ per day (or we can find λ per minute)

$$\Rightarrow \mu = \frac{1}{36} \text{ /minute} \Rightarrow \mu = \frac{1}{36} \times 60 \times 24 = 40 \text{ trains/day}$$

$$\lambda = 30 \text{ trains/day}$$

and

$$\mu = 40 \text{ trains/day}$$

\therefore The traffic intensity,

$$\rho = \frac{\lambda}{\mu} = \frac{30}{40} = \frac{3}{4}$$

Performance measures:

(i) The mean queue size (length),

$$L_q = \frac{\lambda}{\mu} \cdot \frac{\lambda}{\mu - \lambda} = \frac{30}{40} \cdot \frac{30}{40 - 30} = \frac{3}{4} \cdot \frac{3}{1} = \frac{9}{4} = 2\frac{1}{4} = 2.25$$

(ii) The probability that the queue size exceeds 10,

$$P(L_q \geq 10) = \left(\frac{\lambda}{\mu}\right)^{10} = \left(\frac{3}{4}\right)^{10} = 0.056$$

(iii) If the input (arrival) of trains increases to an average 33 per day, then

$$\lambda = 33 \text{ trains/day}$$

$$\therefore L_q = \frac{33}{40} \times \frac{33}{40 - 33} = \frac{33}{40} \times \frac{33}{7} = \frac{1089}{280} = 3.89 \text{ train/hour}$$

$$\therefore \text{The change in queue size} = 3.89 - 2.25 = 1.64 \text{ trains}$$

$$P(L_q \geq 10) = \left(\frac{\lambda}{\mu}\right)^{10} = \left(\frac{33}{40}\right)^{10} = 0.146$$

$$\text{The change in probability} = 0.146 - 0.056 = 0.090$$

EXAMPLE 6.35 Customers arrive at a watch repair shop according to a Poisson process at a rate of one per every 10 minutes and the service time is an exponential random variable with mean 8 minutes. Find (i) the average number of customers in the system, (ii) the average number of customers in the queue, (iii) the average waiting time a customer spends in the shop, and (iv) the average waiting time a customer spends in waiting for service.

Solution One repair shop. The arrival follows Poisson process and service time follows exponential distribution. Therefore, it is an M/M/1 infinite capacity model.

Arrival rate = Mean of Poisson process = $\lambda = \frac{1}{10}$ per minute (given 1 per every 10 minutes)

Service time is exponential.

\therefore Mean of the exponential distribution = $\frac{1}{\mu} = 8/\text{minutes}$

$$\Rightarrow \mu = \frac{1}{8} / \text{minute}$$

(i) The average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{10}}{\frac{1}{8} - \frac{1}{10}} = 4$$

(ii) The average number of customers in the queue,

$$L_q = \frac{\lambda}{\mu} \cdot \frac{\lambda}{\mu - \lambda} = \frac{1}{\frac{1}{8}} \times 4 = \frac{8}{10} \times 4 = 3\frac{1}{5} \text{ customers}$$

(iii) The average waiting time a customer spend in the shop,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{8} - \frac{1}{10}} = 40 \text{ minutes}$$

(iv) The average waiting time a customer spends in waiting for service

$$W_q = \frac{\lambda}{\mu} \cdot \frac{1}{\mu - \lambda} = \frac{8}{10} \times 40 = 32 \text{ minutes}$$

6.5 MODEL II—(M/M/c):(∞/FIFO) MULTI-SERVER POISSON QUEUEING MODEL

This queueing system deals with queues which are being served by parallel service channels (like reservation counters) in which each server has an independently and identically distributed exponential service distribution. The arrival is considered to be Poisson. This model differs from the first model in the sense that, in the first model, we considered the queue with single-service channel, whereas here the number of service channels is c .

In the case of c servers working together, we have two cases:

1. If the number of customers in the system is less than c , i.e. $n < c$, then only n of the c servers will be busy and, hence, the mean service rate will be $n\mu$.
2. The number of customers in the system is more than c , i.e. $n > c$, then all the c servers will be busy and, hence, the mean service rate will be $c\mu$, i.e.

$$\begin{aligned} \mu_n &= n\mu, && \text{if } 0 \leq n < c \\ &= c\mu, && \text{if } n \geq c \end{aligned}$$

and also

$$\lambda_n = \lambda$$

The steady-state difference equations are

$$\begin{aligned} P_0(t + \Delta t) &= P_0(t) (1 - \lambda \Delta t) + P_1(t) \mu \Delta t, \text{ for } n = 0 \\ P_n(t + \Delta t) &= P_n(t) [1 - (\lambda + n\mu) \Delta t] + P_{n-1}(t) \lambda \Delta t + P_{n+1}(t) (n+1)\mu \Delta t, \\ &\quad \text{for } n = 1, 2, 3, \dots, c-1 \\ &= P_n(t) [1 - (\lambda + c\mu) \Delta t] + P_{n-1}(t) \lambda \Delta t + P_{n+1}(t) c\mu \Delta t, \\ &\quad \text{for } n = c, c+1, c+2, c+3, \dots \end{aligned}$$

574 ◇ Probability and Queueing Theory

Dividing the equations by Δt and taking the limit as $\Delta t \rightarrow 0$, the difference equations are simplified to

$$\begin{aligned} P'_0(t) &= -\lambda P_0(t) + \mu P_1(t), && \text{for } n = 0 \\ P'_n(t) &= -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), && \text{for } n = 1, 2, 3, \dots, c-1 \\ P'_n(t) &= -(\lambda + c\mu) P_n(t) + \lambda P_{n-1}(t) + c\mu P_{n+1}(t), && \text{for } n = c, c+1, c+2, c+3, \dots \end{aligned}$$

Considering the case of steady-state independent of t as, $P'_n(t) \rightarrow 0$ for all n , we get from the above equations,

$$0 = -\lambda P_0 + \mu P_1, \quad \text{for } n = 0 \quad (6.20)$$

$$0 = -(\lambda + n\mu) P_n + \lambda P_{n-1} + (n+1)\mu P_{n+1}, \quad \text{for } 1 \leq n < c \quad (6.21)$$

$$0 = -(\lambda + c\mu) P_n + \lambda P_{n-1} + c\mu P_{n+1}, \quad \text{for } n \geq c \quad (6.22)$$

From Eq. (6.20), we get

$$P_1 = \frac{\lambda}{\mu} P_0$$

Substituting $n = 1, 2, 3, \dots, c$ in Eq. (6.21) and simplifying, we get

$$P_2 = \frac{\lambda}{2\mu} P_1 = \frac{1}{2!} \left(\frac{\lambda}{\mu} \right)^2 P_0$$

$$P_3 = \frac{\lambda}{3\mu} P_2 = \frac{1}{3!} \left(\frac{\lambda}{\mu} \right)^3 P_0$$

$$P_4 = \frac{\lambda}{4\mu} P_3 = \frac{1}{4!} \left(\frac{\lambda}{\mu} \right)^4 P_0$$

In general,

$$P_n = \frac{\lambda}{n\mu} P_{n-1} = \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0, \quad 1 \leq n < c$$

Again from Eq. (6.22), we get

$$P_c = \frac{\lambda}{c\mu} P_{c-1} = \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^c P_0$$

$$P_{c+1} = \frac{\lambda}{c\mu} P_c = \frac{1}{c} \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^{c+1} P_0$$

$$P_{c+2} = \frac{\lambda}{c\mu} P_{c+1} = \frac{1}{c^2} \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^{c+2} P_0$$

In general,

$$P_n = P_{c+(n-c)} = \frac{1}{c^{n-c}} \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^n P_0, \quad n \geq c$$

We know that the total probability is always equal to 1.
Therefore,

$$\begin{aligned} & \sum_{n=0}^{\infty} P_n = 1 \\ \text{i.e. } & \sum_{n=0}^{c-1} P_n + \sum_{n=c}^{\infty} P_n = 1 \\ \text{or } & \sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0 + \sum_{n=c}^{\infty} \frac{1}{c! c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0 = 1 \\ \text{or } & \left[\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n + \frac{c^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\lambda}{c\mu} \right)^n \right] P_0 = 1 \\ \text{or } & \left[\sum_{n=0}^{c-1} \frac{c^n}{n!} \left(\frac{\lambda}{c\mu} \right)^n + \frac{1}{c!} \sum_{n=c}^{\infty} \frac{c^n}{c^{n-c}} \left(\frac{\lambda}{c\mu} \right)^n \right] P_0 = 1 \\ \text{or } & \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^{\infty} \rho^n \right] P_0 = 1, \quad \rho = \frac{\lambda}{c\mu} \\ & P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^{\infty} \rho^n \right]^{-1}, \quad \rho = \frac{\lambda}{c\mu} < 1 \\ \therefore & P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1}, \quad \rho = \frac{\lambda}{c\mu} < 1 = \frac{\lambda}{c\mu} < 1 \end{aligned}$$

Therefore, the steady-state probability distribution of n arrivals in the system are

$$\begin{aligned} P_n &= \frac{\lambda}{n\mu} P_{n-1} = \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0, \quad 0 \leq n < c \\ &= P_{c+(n-c)} = \frac{1}{c^{n-c}} \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^n P_0, \quad n \geq c \end{aligned}$$

Note: System busy = server busy = $\rho = \frac{\lambda}{c\mu}$ is also called *traffic intensity*
(M/M/C-Model).

The condition for the existence of steady-state solution for M/M/c infinite capacity model is

$$\rho = \frac{\lambda}{c\mu} < 1$$

6.5.1 Performance Measures (M/M/c-Model)

1. The average number (expected number) of customers in the queue (L_q):
Only when all c servers are busy, there will be a queue,

$$\begin{aligned} L_q &= \sum_{n=c+1}^{\infty} (n-c) P_n \\ &= \sum_{n=c}^{\infty} (n-c) \frac{1}{c! c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{c^c}{c!} P_0 \sum_{n=c}^{\infty} (n-c) \left(\frac{\lambda}{c\mu}\right)^n \\ &= \frac{c^c}{c!} \left(\frac{\lambda}{c\mu}\right)^{c+1} P_0 \left[1 + 2\frac{\lambda}{c\mu} + 3\left(\frac{\lambda}{c\mu}\right)^2 + \dots \right] \\ &= \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \frac{\lambda}{c\mu} P_0 \left(1 - \frac{\lambda}{c\mu}\right)^{-2} \\ &= \frac{1}{c! c} \left(\frac{\lambda}{\mu}\right)^{c+1} \frac{1}{\left(1 - \frac{\lambda}{c\mu}\right)^2} P_0 \end{aligned}$$

Taking

$$\rho = \frac{\lambda}{c\mu}$$

$$L_q = \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0$$

2. The average number (expected number) of customers in the system (L_s),

$$L_s = L_q + \frac{\lambda}{\mu} = \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 + \frac{\lambda}{\mu}$$

3. The average time (expected waiting time) a customer spends in the system (W_s),

$$W_s = \frac{L_s}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 + \frac{1}{\mu}$$

4. The average time (expected waiting time) a customer spends in the queue (W_q),

$$W_q = \frac{L_q}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0$$

5. The probability that an arrival has to wait:

As there are c servers, an arrival has to wait if there are more than c customers in the system. Therefore,

$$\begin{aligned} P(W_s > 0) &= P(n \geq c) = \sum_{n=c}^{\infty} P_n \\ &= \sum_{n=c}^{\infty} \frac{1}{c! c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0 \\ &= \frac{c^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\lambda}{c\mu} \right)^n P_0 \\ &= \frac{c^c}{c!} P_0 \left[\left(\frac{\lambda}{c\mu} \right)^c + \left(\frac{\lambda}{c\mu} \right)^{c+1} + \left(\frac{\lambda}{c\mu} \right)^{c+2} + \dots \right] \\ &= \frac{c^c}{c!} P_0 \left(\frac{\lambda}{c\mu} \right)^c \left[1 + \frac{\lambda}{c\mu} + \left(\frac{\lambda}{c\mu} \right)^2 + \dots \right] \\ &= \frac{c^c}{c!} P_0 \left(\frac{\lambda}{c\mu} \right)^c \left(1 - \frac{\lambda}{c\mu} \right)^{-1} \\ &= \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0 \end{aligned}$$

6. The probability that an arrival has to get the service without waiting,
 $P(\text{arrival getting the service without waiting}) = 1 - P(\text{arrival has to wait})$

$$= 1 - \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0$$

7. The probability that some one will be waiting,

$$\begin{aligned} P(n \geq c+1) &= \sum_{n=c+1}^{\infty} P_n = \sum_{n=c+1}^{\infty} \frac{1}{c! c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0 \\ &= \frac{c^c}{c!} P_0 \sum_{n=c+1}^{\infty} \left(\frac{\lambda}{c\mu} \right)^n \end{aligned}$$

$$\begin{aligned}
 &= \frac{c^c}{c!} \left[\left(\frac{\lambda}{c\mu} \right)^{c+1} + \left(\frac{\lambda}{c\mu} \right)^{c+2} + \left(\frac{\lambda}{c\mu} \right)^{c+3} + \dots \right] P_0 \\
 &= \frac{c^c}{c!} P_0 \left(\frac{\lambda}{c\mu} \right)^{c+1} \left[1 + \frac{\lambda}{c\mu} + \left(\frac{\lambda}{c\mu} \right)^2 + \dots \right] \\
 &= \frac{c^c}{c!} \left(\frac{\lambda}{c\mu} \right)^{c+1} \frac{1}{1 - \frac{\lambda}{c\mu}} P_0 \\
 &= \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)} P_0
 \end{aligned}$$

$$\text{Note: } P(n \geq c) = \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0$$

8. The mean waiting time in the non-empty queue (for those who actually wait),

$$\begin{aligned}
 E(W_q/W_s) &= \frac{W_q}{P(W_s > 0)} = \frac{\frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0}{\frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0} \\
 &= \frac{1}{\lambda} \frac{\rho}{(1-\rho)}
 \end{aligned}$$

9. The average number of customers (in non-empty queues) who have to actually wait,

$$L_w = \frac{L_q}{P(n \geq c)} = \frac{\frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0}{\frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0} = \frac{\rho}{1-\rho}$$

EXAMPLE 6.36 What is the probability that an arrival to an infinite capacity 3 server Poisson queueing system with $\lambda/\mu = 2$ and $P_0 = 1/9$ enters the service without waiting?

Solution There are 3 servers.

\therefore It is an M/M/c/ ∞ model.

$$\text{Given: } \frac{\lambda}{\mu} = 2$$

$$\text{and } P_0 = \frac{1}{9}$$

An arrival enters the service without waiting when the number of customers getting service in the system are less than 3,

$$P[\text{without waiting}] = P[n < 3] = P_0 + P_1 + P_2$$

$$P_n = \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0, \quad \text{when } n \leq c, c = 3$$

$$\therefore P(n < 3) = \frac{1}{9} + \frac{2}{9} + \frac{1}{2} \times \frac{4}{9} = \frac{5}{9}$$

EXAMPLE 6.37 If there are 2 servers in an infinite capacity Poisson queue system with $\lambda = 10$ hour and $\mu = 15$ per hour, what is the percentage of idle time for each server?

Solution There are 2 servers.

It is an M/M/c/∞ model.

Given: $\lambda = 10/\text{hour}$, $\mu = 15/\text{hour}$ and $c = 2$

$$\rho = \frac{\lambda}{c\mu} = \frac{10}{30} = \frac{1}{3}$$

$$c\rho = \frac{2}{3}$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[\sum_{n=0}^{c-1} \frac{\left(\frac{2}{3}\right)^n}{n!} + \frac{\left(\frac{2}{3}\right)^2}{2!\left(1-\frac{1}{3}\right)} \right]^{-1} = \frac{1}{1+\frac{2}{3}+\frac{1}{3}} = \frac{1}{2} \end{aligned}$$

\therefore The percentage of idle time for each server = 50%

EXAMPLE 6.38 Consider an M/M/c queueing system. Find the probability that an arriving customer is forced to join queue.

Solution An arriving customer is forced to join the queue if there are c or more customers.

$$\begin{aligned} P(n \geq c) &= \sum_{n=c}^{\infty} \frac{1}{c!c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0 \\ &= \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^c P_0 \sum_{n=c}^{\infty} \left(\frac{\lambda}{c\mu} \right)^{n-c} = \frac{\left(\frac{\lambda}{\mu} \right)^c P_0}{c! \left(1 - \frac{\lambda}{c\mu} \right)} = \frac{c^c P^c}{c!(1-P)} P_0 \end{aligned}$$

580 ◇ Probability and Queueing Theory

EXAMPLE 6.39 If $\lambda/c\mu = 4/5$ in an (M/M/c): (∞ /FCFS) queueing system, find the average number of customers in the non-empty queue.

Solution Given: $\rho = \frac{\lambda}{c\mu} = \frac{4}{5}$

The average number of customers in the non-empty queue for an (M/M/c): (∞ /FCFS) queueing system is given by

$$L_w = \frac{\rho}{1 - \rho} = \frac{\frac{4}{5}}{1 - \left(\frac{4}{5}\right)} = 4$$

EXAMPLE 6.40 Given $c = 6$ and $\rho = \lambda/c\mu = 1/3$, find P_0 .

Solution Given $c = 6$.

Therefore, it is an (M/M/c): (∞ /FCFS) model.

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1}, \quad \rho = \frac{\lambda}{c\mu} = \frac{1}{3} < 1 \\ &= \left\{ \sum_{n=0}^5 \frac{(2)^n}{n!} + \frac{(2)^6}{6![1-(1/3)]} \right\}^{-1} = 0.1351 \quad \left[\because c\rho = \frac{6}{3} = 2 \right] \\ \therefore P_0 &= 0.1351 \end{aligned}$$

EXAMPLE 6.41 Find queue length of the system if $\lambda = 12$ per hour, $\mu = 4$ per hour, $P_0 = 0.0313$, and $c = 4$.

Solution Given: $\lambda = 12/\text{hour}$, $\mu = 4/\text{hour}$, $P_0 = 0.0313$ and $c = 4$

\therefore It is an M/M/c infinite capacity model.

The queue length of the system,

$$\begin{aligned} L_q &= \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 \\ &= \frac{4^4}{4!} \frac{\left(\frac{3}{4}\right)^5}{\left[1 - \left(\frac{3}{4}\right)\right]^2} \times 0.0313 \\ &= 1.26765 \end{aligned}$$

EXAMPLE 6.42 There are 2 clerks in a college to receive dues from the students. If the service time for each student is exponential with mean 4 minutes,

and the boys arrive in a Poisson fashion at the counter at the rate of 10/hour, what is the percentage of idle time for each clerk?.

Solution There are 2 clerks in a college

$$c = 2$$

Service time for each student is exponential with mean 4 minutes

$$\Rightarrow \frac{1}{\mu} = 4 \text{ minutes} \Rightarrow \mu = \frac{1}{4} / \text{minute} \Rightarrow \mu = \frac{1}{4} \times 60 = 15 / \text{hour}$$

The boys arrive in a Poisson fashion at the counter at the rate of 10/hour, i.e.

$$\begin{aligned} & \lambda = 10 / \text{hour} \\ \Rightarrow & \rho = \frac{\lambda}{c\mu} = \frac{10}{2 \times 15} = \frac{1}{3} \\ & c\rho = \frac{2}{3} \end{aligned}$$

The probability that the servers are busy

$$\rho = \frac{\lambda}{c\mu} = \frac{1}{3}$$

The probability that the servers are idle

$$= 1 - \rho = 1 - \frac{1}{3} = \frac{2}{3} = 0.667$$

The percentage of idle time for each clerk is 66.7%

EXAMPLE 6.43 If there are 3 cashiers in a bank to receive cash in the cash counter and the service time for each customer is 3 minutes, find the probability that there is no customer in the bank if the arrival rate is 15 per hour.

Solution There are 3 cashiers in a bank

$$c = 3$$

Service time is 3 minutes for each customer,

$$\Rightarrow \frac{1}{\mu} = 3 \text{ minutes} \Rightarrow \mu = \frac{1}{3} / \text{minute} \Rightarrow \mu = \frac{1}{3} \times 60 = 20 / \text{hour}$$

Arrival rate $\lambda = 15 / \text{hour}$

$$\Rightarrow \rho = \frac{l}{c\mu} = \frac{15}{3 \times 20} = \frac{1}{4}, \quad c\rho = \frac{3}{4}$$

The probability that there is no customer in the bank is

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[\sum_{n=0}^2 \frac{\left(\frac{3}{4}\right)^n}{n!} + \frac{\left(\frac{3}{4}\right)^3}{3! \left[1 - \left(\frac{1}{4}\right)\right]} \right]^{-1} = 0.4706 \end{aligned}$$

EXAMPLE 6.44 Four counters are being run on the frontier of a country to check the passports and necessary papers of the tourists. The tourist chooses a counter at random. If the arrival at the frontier is Poisson at the rate of λ , and the service time is exponential with parameter $\lambda/2$, what is the steady-state average queue at each counter?

Solution As there are 4 counters, the arrivals follow Poisson process and the service time follows exponential distribution, it is an M/M/c infinite capacity model.

$$\begin{array}{ll} \text{Given:} & \text{arrival rate} = \lambda \\ \text{and} & c = 4 \end{array}$$

$$\text{Service time is exponential with parameter } \mu = \frac{\lambda}{2}.$$

$$\therefore \rho = \frac{\lambda}{c\mu} = \frac{2}{4} = 0.5 \quad [\because c\rho = 4 \times 0.5 = 2.0]$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[\sum_{n=0}^3 \frac{(2)^n}{n!} + \frac{(2)^4}{4!(1-0.5)} \right]^{-1} = 0.1304 \end{aligned}$$

Performance measures:

The steady-state average queue at each counter,

$$\begin{aligned} L_q &= \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 \\ &= \frac{4^4}{4!} \frac{(0.5)^5}{(1-0.5)^2} \times 0.1304 = 0.1739 \end{aligned}$$

EXAMPLE 6.45 A petrol station has 2 pumps. The service time follows exponential distribution with mean 4 minutes and cars arrive for service in a Poisson process at the rate of 10 cars per hour. (i) Find the probability that a customer has to wait for service. (ii) What proportion of time the pump remains idle?

[AU December '05]

Solution As there are 2 pumps, the arrivals follow Poisson process and the service times follow exponential distribution, it is an M/M/c infinite capacity model.

Given: arrival rate = $\lambda = 10/\text{hour}$
Service time follows exponential distribution with mean

$$\frac{1}{\mu} = 4 \text{ minutes} \Rightarrow \mu = \frac{1}{4} \text{ minute}$$

But λ is given in hours.

$$\therefore \mu = \frac{1}{4} \times 60 = 15/\text{hour}$$

Given: $c = 2$

$$\rho = \frac{\lambda}{c\mu} = \frac{10}{2 \times 15} = 0.333$$

and $c\rho = 0.666$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[\sum_{n=0}^1 \frac{(0.666)^n}{n!} + \frac{(0.666)^2}{2!(1-0.333)} \right]^{-1} = 0.5 \end{aligned}$$

Performance measures:

- (i) The probability that a customer has to wait for service (customer has to wait only if there are 2 or more cars),

$$\begin{aligned} P(n \geq c) &= \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0 \\ &= \frac{2^2}{2!} \frac{(0.333)^2}{(1-0.333)} \times 0.5 = \frac{1}{6} = 0.167 \end{aligned}$$

- (ii) The proportion of time the pump remains idle

$$= 1 - \frac{\lambda}{c\mu} = 1 - \rho = 1 - 0.333 = 0.667$$

Therefore, 67% of time the pumps remain idle.

EXAMPLE 6.46 A supermarket has 2 girls attending to sales at the counters. If the service time for each customer is exponential with mean 4 minutes and if people arrive in Poisson fashion at the rate of 10 per hour, (i) what is the probability that a customer has to wait for service, and (ii) what is the expected percentage of idle time for each girl? (iii) If the customer has to wait in the queue, what is the expected length of his waiting time?

Solution As there are 2 girls for service, the people arrive in Poisson fashion and the service time follows exponential distribution, it is an M/M/c infinite capacity model.

$$\text{Arrival rate} = \lambda = 10/\text{hour}$$

$$\text{Service time} = \frac{1}{\mu}$$

$$\therefore \frac{1}{\mu} = 4 \text{ minutes}$$

$$\text{i.e. } \mu = \frac{1}{4}/\text{minute} = \frac{1}{4} \times 60 = 15/\text{hour} \text{ and } c = 2$$

$$\therefore \rho = \frac{\lambda}{c\mu} = \frac{10}{2 \times 15} = \frac{1}{3} \text{ and } \rho c = \frac{2}{3}$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(\rho c)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[1 + \frac{2}{3} + \frac{\left(\frac{2}{3}\right)^2}{2 \times \left(1 - \frac{1}{3}\right)} \right]^{-1} = \frac{1}{2} \end{aligned}$$

Performance measures:

- (i) A customer has to wait for service only when there are more than 2 customers in the system,

$$\begin{aligned} P(n \geq c) &= \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0 \\ \therefore P(n \geq c) &= \frac{2^2}{2!} \frac{\left(\frac{1}{3}\right)^2}{\left(1 - \frac{1}{3}\right)} \times \frac{1}{2} = \frac{1}{6} \end{aligned}$$

(ii) The fraction of time when the girls are busy,

$$\rho = \frac{\lambda}{c\mu} = \frac{10}{2 \times 15} = \frac{1}{3}$$

∴ The fraction of time when the girls are idle = $1 - \frac{1}{3} = \frac{2}{3}$

∴ The expected percentage of idle time for each girl = $\frac{2}{3} \times 100 = 67\%$

(iii) The customer has to wait in the queue, then the expected length of his waiting time,

$$W_q = \frac{1}{\lambda} \left(\frac{\rho}{1 - \rho} \right)$$

$$W_q = \frac{1}{10} \left(\frac{0.333}{1 - 0.333} \right) = 0.0499 \text{ hour} = 0.0499 \times 60 = 3 \text{ minutes}$$

EXAMPLE 6.47 Given an average arrival rate of 20 per hour. Is it better for a customer to get service at a single channel with mean service rate of 22 customers per hour or at one of the two channels in parallel with mean service rate of 11 customers per hour for each of the two channels? Assume both queues to be of Poisson type.

Solution To find whether a single-channel service is better or two channels in parallel are better for a customer, we have to find the waiting time of a customer in the system.

Given: the queues are of Poisson type.

For the single-channel service, it is an M/M/1 infinite capacity model.

Arrival rate = $\lambda = 20/\text{hour}$

Service rate = $\mu = 22/\text{hour}$

Performance measures:

The waiting time of a customer in the single-channel service,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{2} \text{ hour} = 0.5 \text{ hour}$$

For the two-channel service, it is an M/M/c infinite capacity model.

Arrival rate = $\lambda = 20/\text{hour}$

Service rate = $\mu = 11/\text{hour}$

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1}$$

$$= \left[\sum_{n=0}^1 \frac{(2\rho)^n}{n!} + \frac{(2\rho)^2}{2!(1-\rho)} \right]^{-1} = \left[1 + \frac{20}{11} + \frac{\left(\frac{20}{11}\right)^2}{2 \times \frac{1}{11}} \right]^{-1} = 0.0476$$

The waiting time of a customer in the two-channel service,

$$\begin{aligned} W_s &= \frac{L_s}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 + \frac{1}{\mu} \\ &= \frac{2^2}{20 \times 2} \frac{\left(\frac{20}{22}\right)^3}{\left(\frac{1}{11}\right)^2} \times 0.0476 + \frac{1}{11} \\ &= 0.0909 + 9.0909 \times 0.0476 \\ &= 0.5236 \text{ hour} \end{aligned}$$

As the average waiting time in the single-channel service is less than that in the two-channel service, the customer has to prefer the single-channel service.

EXAMPLE 6.48 A general insurance company has 3 claim adjusters in its branch office. People with claims against the company are found to arrive in Poisson fashion at an average rate of 20 per 8-hour day. The amount of time that an adjuster spends with a claimant is found to have negative exponential distribution with mean service time 40 minutes. Claimants are processed in the order of their appearance.

- (i) How many hours a week can an adjuster expect to spend with claimants?
- (ii) How much time, on the average, does claimant spend in the branch office?

Solution As there are 3 claim adjusters, the arrivals follow Poisson fashion and the service time follows negative exponential distribution, it is an M/M/c infinite capacity model.

$$\text{Arrival rate} = \lambda = \frac{20}{8} = 2.5/\text{hour}$$

Service time follows exponential distribution with mean

$$\frac{1}{\mu} = 40 \text{ minutes} \Rightarrow \mu = \frac{1}{40} \text{ minute}$$

But λ is given in hours.

$$\therefore \mu = \frac{1}{40} \times 60 = 1.5/\text{hour}$$

Given: $c = 3$

$$\rho = \frac{\lambda}{c\mu} = \frac{2.5}{3 \times 1.5} = 0.556$$

$$c\rho = 1.668$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[\sum_{n=0}^2 \frac{(1.667)^n}{n!} + \frac{(1.667)^3}{3!(1-0.556)} \right]^{-1} = 0.1727 \end{aligned}$$

Performance measures:

- (i) The number of hours a week an adjuster can be expected to spend with claimants,

$$\text{The number of hours busy per day} = \rho = \frac{\lambda}{c\mu} = \frac{2.5}{3 \times 1.5} = 0.556$$

$$\text{One week} = 5 \text{ working days} = 5 \times 8 = 40 \text{ hours}$$

$$\text{Number of hours busy per week} = 40 \times 0.556 = 22.24 \text{ hours}$$

- (ii) The average time a claimant spends in the branch office (system),

$$\begin{aligned} W_s &= \frac{L_s}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 + \frac{1}{\mu} \\ &= \frac{3^3}{2.5 \times 3!} \frac{(0.556)^4}{(1-0.556)^2} \times 0.1727 + \frac{1}{1.5} \\ &= 0.8167/\text{hour} = 49 \text{ minutes} \end{aligned}$$

EXAMPLE 6.49 A petrol pump station has 4 pumps. The service times follow the exponential distribution with a mean of 6 minutes and cars arrive for service in a Poisson process at the rate of 30 cars per hour.

- (i) What is the probability that an arrival would have to wait in line?
- (ii) Find the average waiting time, the average time spent in the system, and the average number of cars in the system.
- (iii) For what percentage of time would a pump be idle on an average?

Solution As there are 4 pumps, the arrivals (demand) follow Poisson process and the service times follow exponential distribution, it is an M/M/c infinite capacity model.

$$\text{Arrival rate} = \lambda = 30/\text{hour}$$

$$\text{Service time} = \frac{1}{\mu}$$

$$\therefore \frac{1}{\mu} = 6 \text{ minutes}$$

588 ◇ Probability and Queueing Theory

$$\mu = \frac{1}{6}/\text{minute} = \frac{1}{6} \times 60 = 10/\text{hour} (\lambda \text{ is given in hour})$$

Given: $c = 4$, $\lambda = 30/\text{hour}$, $\mu = 10/\text{hour}$ and $c\rho = 4 \times 0.75 = 3$

$$\rho = \frac{\lambda}{c\mu} = \frac{30}{4 \times 10} = 0.75$$

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(\rho c)^c}{c!(1-\rho)} \right]^{-1}$$

$$= \left[\sum_{n=0}^3 \frac{(3)^n}{n!} + \frac{(3)^4}{4!(1-0.75)} \right]^{-1} = 0.0377$$

Performance measures:

- (i) An arrival (car) has to wait in line only when there are 4 or more cars in the pump station.

The probability that an arrival would have to wait in line,

$$P(n \geq c) = \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0$$

$$P(n \geq 4) = \frac{4^4}{4!} \frac{(0.75)^4}{(1-0.75)} \times 0.0377 = 0.5090$$

- (ii) The average waiting time of a car in the queue,

$$W_q = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0$$

$$= \frac{4^4}{30 \times 4!} \frac{(0.75)^5}{(1-0.75)^2} \times 0.0377$$

$$= 0.0509 \text{ hour} = 3.05 \text{ minutes}$$

The average waiting time of a car in the system,

$$W_s = \frac{1}{\mu} + W_q = 6 + 3.05 = 9.05 \text{ minutes}$$

The average number of cars in the system,

$$L_s = \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 + \frac{\lambda}{\mu}$$

$$= \frac{4^4}{4!} \frac{(0.75)^5}{(1-0.75)^2} \times 0.0377 + 3 = 4.53 \text{ cars}$$

(iii) The fraction of time when the pumps are busy, Traffic intensity

$$\rho = \frac{\lambda}{c\mu} = \frac{3}{4}$$

∴ The fraction of time when the pumps are idle

$$= 1 - \frac{\lambda}{c\mu} = 1 - \frac{3}{4} = \frac{1}{4}$$

∴ The required percentage = 25%

EXAMPLE 6.50 There are 3 typists in an office. Each typist can type an average of 6 letters per hour. The letters arrive for being typed at the rate of 15 letters per hour.

- (i) What fraction of the time all the typists will be busy?
- (ii) What is the average number of letters waiting to be typed?
- (iii) What is the average time a letter has to spend for waiting and for being typed?
- (iv) What is the probability that a letter will take longer than 20 minutes waiting to be typed and being typed?

[AU December '03, May '06]

Solution There are 3 typists. Assume arrivals follow Poisson distribution. No system size is given.

∴ It is an M/M/c infinite capacity model.

Given: letters arrive for being typed at the rate of 15 letters per hour

∴ Arrival rate = $\lambda = 15/\text{hour}$

Typist can type an average of 6 letters per hour

∴ Service rate = $\mu = 6/\text{hour}$

The number of servers (typists) = $c = 3$

$$\rho = \frac{\lambda}{c\mu} = \frac{15}{3 \times 6} = \frac{5}{6} = 0.833$$

$$c\rho = 3 \times \frac{5}{6} = 2.5$$

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1}$$

$$= \left[\sum_{n=0}^2 \frac{(2.5)^n}{n!} + \frac{(2.5)^3}{3!(1-0.833)} \right]^{-1}$$

590 ◇ Probability and Queueing Theory

$$= \left[1 + 2.5 + \frac{(2.5)^2}{2} + \frac{(2.5)^3}{3!(1 - 0.833)} \right]^{-1} = 0.0449$$

$$\lambda = 15, \mu = 6 \text{ and } P_0 = 0.0449$$

Performance measures:

- (i) The fraction of the time all the typists will be busy = $P(\text{all the typists are busy}) = P(n \geq 3)$

$$P(n \geq c) = \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0$$

$$P(n \geq 3) = \frac{3^3}{3!} \frac{(0.833)^3}{(1-0.833)} \times 0.0449$$

$$= 0.7016$$

Hence the fraction of the time all the typists will be busy = 70%.

- (ii) The average number of letters waiting to be typed is the average number of letters waiting in the queue,

$$L_q = \frac{c^c}{c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0$$

$$= \frac{3^3}{3!} \frac{(0.833)^4}{(1-0.833)^2} \times 0.0449 = 3.5078$$

- (iii) The average time a letter has to spend for waiting and for being typed is the average time of the letter in the system,

$$W_s = \frac{L_q}{\lambda} \quad (\text{using Little's formula})$$

$$= \frac{1}{\lambda} \left(L_q + \frac{\lambda}{\mu} \right) = \frac{1}{15} \left(3.5078 + \frac{15}{6} \right) = 0.4005 \text{ hour}$$

$$= 24 \text{ minutes}$$

- (iv) The probability that a letter will take longer than 20 minutes waiting to be typed and being typed (waiting time of the letter in the queue is more than 20 minutes),

$$t = 20 \text{ minutes} = 20 \times \frac{1}{60} = \frac{1}{3} \text{ hour}$$

$$P(W > t) = e^{-\mu t} \left\{ 1 + \frac{\left(\frac{\lambda}{\mu} \right)^c \left[1 - e^{-\mu \left(c - 1 - \frac{\lambda}{\mu} \right)} \right]}{c! \left(1 - \frac{\lambda}{\mu c} \right) \left(c - 1 - \frac{\lambda}{\mu} \right)} P_0 \right\}$$

$$\therefore P\left(W > \frac{1}{3}\right) = e^{-\frac{6}{3}} \left\{ 1 + \frac{(2.5)^3 [1 - e^{(-2 \times -0.5)}] \times 0.0449}{6 \left(1 - \frac{2.5}{3}\right) (-0.5)} \right\}$$

$$= 0.4616$$

EXAMPLE 6.51 A telephone company is planning to install telephone booths in a new airport. It has established the policy that a person should not have to wait more than 10% of the times he tries to use a phone. The demand for use is estimated to be Poisson with an average of 30 per hour. The average phone call has an exponential distribution with a mean time of 5 minutes. How many phone booths should be installed?

Solution As there are more than one booth, the arrivals (demand) follow Poisson distribution and the service time follows exponential distribution, it is an M/M/c infinite capacity model.

$$\text{Arrival rate (demand)} = \lambda = 30/\text{hour}$$

The average phone call (service time) has exponential distribution with mean

$$= \frac{1}{\mu} = 5 \text{ minutes}$$

But λ is given in hour.

In order that infinite queue may not build up, the traffic intensity $\lambda/c\mu < 1$, for multi-server model, i.e.

$$c > \frac{\lambda}{\mu}$$

$$\text{or } c > \frac{30}{12} (= 2.5)$$

But a person should not have to wait more than 10% of the times he tries to use a phone.

Therefore, the telephone company must intall at least 3 booths.

Now we have to find the number of telephone booths, such that

$$\begin{aligned} P(W_s > 0) &\leq 0.10 \text{ or equivalently} \\ \text{i.e. } P(n \geq c) &\leq 0.10 \end{aligned}$$

We have to find c such that

$$P(n \geq c) = \frac{c^c}{c!} \frac{\rho^c}{(1-\rho)} P_0 \leq 0.10$$

This equation is not easily solvable. Hence we proceed by trials and find out the best value of c that it satisfies in this equation.

Let

$$c = 3.$$

$$\begin{aligned}
 P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(\rho c)^c}{c!(1-\rho)} \right]^{-1} \\
 &= \left\{ \left[1 + 2.5 + \frac{1}{2} \times (2.5)^2 \right] + \frac{(2.5)^3}{3! \left(1 - \frac{2.5}{3} \right)} \right\}^{-1} \\
 &= (22.25)^{-1} = 0.0449
 \end{aligned}$$

Then

$$\begin{aligned}
 P(W_s > 0) &= \frac{(2.5)^3 P_0}{6 \left(1 - \frac{2.5}{3} \right)} = 15.625 P_0 \\
 &= 15.625 \times 0.0449 \\
 &= 0.7022 > 0.10
 \end{aligned}$$

∴ Let $c = 4$.

$$\begin{aligned}
 P_0 &= \left[\sum_{n=0}^3 \frac{(2.5)^n}{n!} + \frac{(2.5)^4}{4!(1-0.833)} \right]^{-1} \\
 &= 0.0737
 \end{aligned}$$

Then

$$\begin{aligned}
 P(W_s > 0) &= \frac{(2.5)^4 P_0}{24 \left(1 - \frac{2.5}{4} \right)} = 4.3403 P_0 \\
 &= 4.3403 \times 0.0737 = 0.319 > 0.10
 \end{aligned}$$

Similarly, when $c = 5$, $P(W_s > 0) = 0.1304 > 0.10$
and when $c = 6$, $P(W_s > 0) = 0.047 < 0.10$

Hence, the number of booths to be installed = 6.

EXAMPLE 6.52 A bank has two tellers working on savings accounts. The first teller handles withdrawals only. The second teller handles deposits only. It has been found that the service time distributions for both deposits and withdrawals are exponential with mean service time of 3 minutes per customer. Depositors are found to arrive in a Poisson fashion with mean arrival rate of 16 per hour. Withdrawers also arrive in a Poisson fashion with mean arrival rate of 14 per hour. What would be the effect on the average waiting time for the customers, if each teller could handle both withdrawals and deposits? What would be the effect, if this could only be accomplished by increasing the service time to 3.5 minutes?

[AU December '03]

Solution When there is a separate channel for depositors and withdrawers, the arrivals follow Poisson distribution and the service time distributions are exponential, it is an M/M/1 infinite capacity model.

$$\text{Arrival rate (depositors)} = \lambda_1 = 16/\text{hour}$$

Service time follows exponential distribution with mean

$$\frac{1}{\mu} = 3 \text{ minutes/customer}$$

But λ_1 is given in hour.

$$\text{i.e. } \mu = \frac{1}{3} \text{ minute} = \frac{1}{3} \times 60 = 20/\text{hour}$$

When there is a separate channel for the withdrawers, arrival rate (withdrawers) $\lambda_2 = 14/\text{hour}$

Service time is the same for withdrawers also.

$$\therefore \mu = 20/\text{hour}$$

Performance measures:

The waiting time of a customer in the single-channel service,

$$\begin{aligned} W_q \text{ for depositors} &= \frac{\lambda_1}{\mu(\mu - \lambda_1)} \\ &= \frac{16}{20(20-16)} = \frac{1}{5} \text{ hour} = 12 \text{ minutes} \\ W_q \text{ for withdrawers} &= \frac{\lambda_2}{\mu(\mu - \lambda_2)} \\ &= \frac{14}{20(20-14)} = \frac{7}{60} \text{ hour} = 7 \text{ minutes} \end{aligned}$$

If both tellers do both service, then it will be an M/M/c model with

$$c = 2, \mu = 20/\text{hour}, \lambda = \lambda_1 + \lambda_2 = 30/\text{hour}.$$

Then

$$\begin{aligned} \rho &= \frac{\lambda}{c\mu} = \frac{30}{2 \times 20} = \frac{3}{4} = 0.75 \\ c\rho &= 1.5 \end{aligned}$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(\rho c)^c}{c!(1-\rho)} \right]^{-1} \\ &= \left[1 + 1.5 + \frac{(1.5)^2}{2 \times 0.25} \right]^{-1} = \frac{1}{7} \end{aligned}$$

594 ◇ Probability and Queueing Theory

The waiting time of a customer in the two-channel service,

$$\begin{aligned} W_q &= \frac{L_q}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 \\ &= \frac{2^2}{30 \times 2!} \frac{(0.75)^3}{(1-0.75)^2} \times \frac{1}{7} = 0.0643 \text{ hour} = 3.858 \text{ minutes} \end{aligned}$$

Hence, if both tellers do both types of service, the customers get benefited as their waiting time is considerably reduced.

Now, if both tellers do both types of service with increased service time,

$$\begin{aligned} c &= 2, \lambda = 30, \mu = \frac{60}{3.5} = \frac{120}{7} / \text{hour} \\ \therefore c\rho &= \frac{\lambda}{\mu} = \frac{30 \times 7}{120} = 1.75 \text{ and } \rho = 0.875 \\ P_0 &= \left[1 + 1.75 + \frac{(1.75)^2}{2 \times \frac{1}{8}} \right]^{-1} = \frac{1}{15} \end{aligned}$$

The waiting time of a customer in this case,

$$\begin{aligned} W_q &= \frac{L_q}{\lambda} = \frac{c^c}{\lambda c!} \frac{\rho^{c+1}}{(1-\rho)^2} P_0 \\ &= \frac{2^2}{30 \times 2!} \frac{(0.875)^3}{(1-0.875)^2} \times \frac{1}{15} = 0.1910 \text{ hour} = 11.46 \text{ minutes} \end{aligned}$$

As the waiting time of the depositors is 12 minutes in the previous case, they will get slightly benefited, whereas withdrawers stand to lose as their waiting time is now increased to 12 minutes (in the previous case only 7 minutes they have to wait).

6.6 MODEL III — (M/M/1): (N/FIFO) SINGLE SERVER WITH FINITE CAPACITY POISSON QUEUEING MODEL

This model represents the situation in which the system can accommodate only a finite number N of arrivals. If a customer arrives and the queue is full, the customer leaves without joining the queue. Therefore, for this model,

$$\begin{aligned} \mu_n &= \mu, n = 1, 2, 3, \dots \\ \text{and } \lambda_n &= \begin{cases} \lambda, & \text{for } n = 0, 1, 2, \dots, (N-1) \\ 0, & \text{for } n = N, N+1, \dots \end{cases} \end{aligned}$$

The difference equations for this model are:

$$P'_0(t) = -\lambda P_0(t) + \mu P_1(t), \quad \text{for } n = 0$$

$$\begin{aligned} P'_n(t) &= \lambda P_{n-1}(t) - (\lambda + \mu) P_n(t) + \mu P_{n+1}(t), \quad \text{for } 1 \leq n < N \\ &= \lambda P_{N-1}(t) - \mu P_N(t), \quad \text{for } n \geq N \end{aligned}$$

For the Poisson queue system, $P_n = P(N = n)$ in the steady-state is given by the difference equations,

$$\begin{aligned} -\lambda P_0 + \mu P_1 &= 0, \quad \text{for } n = 0 \\ \lambda P_{n-1} - (\lambda + \mu) P_n + \mu P_{n+1} &= 0, \quad \text{for } 1 \leq n < N \\ \lambda P_{N-1} - \mu P_N &= 0, \quad \text{for } n = N \end{aligned}$$

From the above equations, we have

$$\mu P_1 = \lambda P_0 \quad (6.23)$$

$$\mu P_{n+1} = (\lambda + \mu) P_n - \lambda P_{n-1}, \quad \text{for } 1 \leq n \leq N-1 \quad (6.24)$$

$$\mu P_N = \lambda P_{N-1}, \quad \text{for } n = N \quad (\because P_{N+1} \text{ has no meaning}) \quad (6.25)$$

From Eq. (6.23),

$$P_1 = \frac{\lambda}{\mu} P_0$$

From Eq. (6.24),

$$\mu P_2 = (\lambda + \mu) \frac{\lambda}{\mu} P_0 - \lambda P_0$$

Simplifying,

$$P_2 = \left(\frac{\lambda}{\mu} \right)^2 P_0$$

Substituting $n = 2, 3, 4, \dots$ in Eq. (6.24) and simplifying, we get

$$P_n = \left(\frac{\lambda}{\mu} \right)^n P_0, \quad 1 \leq n \leq N-1$$

From Eq. (6.25),

$$P_N = \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu} \right)^{N-1} = \left(\frac{\lambda}{\mu} \right)^N P_0$$

As total probability is always equal to 1, we have

$$\begin{aligned} \sum_{n=0}^N P_n &= 1 \Rightarrow P_0 \sum_{n=0}^N \left(\frac{\lambda}{\mu} \right)^n = 1 \\ \Rightarrow P_0 \frac{\left[1 - \left(\frac{\lambda}{\mu} \right)^{N+1} \right]}{1 - \frac{\lambda}{\mu}} &= 1 \quad \left[\because 1 + x + x^2 + \dots + x^n = \frac{1 - x^{n+1}}{1 - x} \right] \end{aligned}$$

$$P_0 = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}} = \frac{1 - \rho}{1 - \rho^{N+1}}, \text{ where } \rho = \frac{\lambda}{\mu}$$

which is valid even for $\lambda > \mu$, i.e. for $\rho > 1$ also.

When $\lambda = \mu$

$$\begin{aligned} \lim_{\frac{\lambda}{\mu} \rightarrow 1} \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}} &= \frac{1}{N+1} && \text{(using L'Hospital's rule)} \\ \therefore P_0 &= \frac{1 - \rho}{1 - \rho^{N+1}}, \quad \lambda \neq \mu \\ &= \frac{1}{N+1}, \quad \lambda = \mu \end{aligned} \quad (6.26)$$

and

$$\begin{aligned} P^n &= \rho^n P_0, \quad \lambda \neq \mu \\ &= \frac{1}{N+1}, \quad \lambda = \mu \end{aligned} \quad (6.27)$$

6.6.1 Performance Measures

1. The average number of customer in the system,

$$\begin{aligned} L_s &= \sum_{n=0}^N n P_n = \frac{1 - \rho}{1 - \rho^{N+1}} \sum_{n=0}^N n \rho^n = \frac{1 - \rho}{1 - \rho^{N+1}} \rho \sum_{n=0}^N \frac{d}{dx} (\rho^n) \\ &\quad \left[\because \frac{d}{dx} (\rho^n) = n \rho^{n-1} \right] \\ &= \frac{1 - \rho}{1 - \rho^{N+1}} \rho \frac{d}{dx} \left(\sum_{n=0}^N \rho^n \right) = \frac{1 - \rho}{1 - \rho^{N+1}} \rho \frac{d}{dx} \left(\frac{1 - \rho^{N+1}}{1 - \rho} \right) \\ &= \frac{(1 - \rho) \rho}{1 - \rho^{N+1}} \left\{ \frac{(1 - \rho) [-(N+1)\rho^N] + (1 - \rho^{N+1})}{(1 - \rho)^2} \right\} \\ &= \frac{\rho}{1 - \rho} - \frac{(N+1)\rho^{N+1}}{1 - \rho^{N+1}} \\ \therefore L_s &= \frac{\rho}{1 - \rho} - \frac{(N+1)\rho^{N+1}}{1 - \rho^{N+1}}, \quad \lambda \neq \mu \end{aligned}$$

and

$$L_s = \sum_{n=0}^N \frac{n}{N+1} = \frac{N}{2}, \text{ if } \lambda = \mu \quad (6.28)$$

2. The average number of customers in the queue,

$$\begin{aligned} L_q &= \sum_{n=1}^N (n-1) P_n \\ &= \sum_{n=0}^N nP_n - \sum_{n=1}^N P_n = L_s - (1 - P_0) \end{aligned} \quad (6.29)$$

3. The overall effective arrival rate:

As per Little's formula,

$$L_q = L_s - \frac{\lambda}{\mu} \quad (6.30a)$$

which is true when the average arrival rate is λ throughout. But, $1 - P_0 \neq \frac{\lambda}{\mu}$, because the average arrival rate is λ as long as there is a vacancy in the queue and it is zero when the system is full.

Hence we define the overall effective arrival rate, denoted by λ' or λ_{eff} by using Eq. (6.27) and Little's formula as

$$\frac{\lambda}{\mu} = 1 - P_0 \quad \text{or} \quad \lambda' = \mu(1 - P_0)$$

Equation (6.30a) can be rewritten as

$$L_q = L_s - \frac{\lambda'}{\mu} \quad (6.30b)$$

which is the modified Little's formula for this model.

4. The average waiting time in the system and in the queue:

By the modified Little's formula,

$$W_s = \frac{1}{\lambda'} L_s \quad (6.31)$$

$$W_q = \frac{1}{\lambda'} L_q \quad (6.32)$$

where λ' is the effective arrival rate.

EXAMPLE 6.53 If $\lambda = 4$ per hour and $\mu = 12$ per hour in an (M/M/1): (4/FIFO) queueing system, find the probability that there is no customer in the system.

Solution It is an M/M/1/N model with $N = 4$.

Given: $\lambda = 4/\text{hour}$

598 ◇ Probability and Queueing Theory

and $\mu = 12/\text{hour}$

$$\therefore \rho = \frac{\lambda}{\mu} = \frac{4}{12} = \frac{1}{3}$$

$$P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - \frac{1}{3}}{1 - \left(\frac{1}{3}\right)^5} = \frac{\frac{2}{3}}{1 - \frac{1}{243}} = \frac{2}{3} \times \frac{243}{242} = \frac{81}{121}$$

EXAMPLE 6.54 Using the Little's formula, obtain the average waiting time in the system for M/M/1/N model.

Solution By the modified Little's formula,

$$W_s = \frac{1}{\lambda'} L_s$$

where λ' is the effective arrival rate and $\lambda' = \mu(1 - P_0)$

EXAMPLE 6.55 Find the length of the queueing system in an (M/M/1): (4/FIFO) model, if $\lambda = 3$ per hour and $\mu = 3$ per hour.

Solution Given: $\lambda = \mu = 3$ per hour for (M/M/1): (4/FIFO) model with $N = 4$.

$$\therefore L_s = \frac{N}{2} = \frac{4}{2} = 2$$

EXAMPLE 6.56 If $\lambda = 2$ per hour and $\mu = 4$ per hour in a 2-capacity single-server queueing system, find the effective arrival rate.

Solution Given: $\lambda = 2/\text{hour}$

$\mu = 4/\text{hour}$ for (M/M/1):(N/FIFO) model with $N = 2$

$$\text{and } \rho = \frac{\lambda}{\mu} = \frac{2}{4} = \frac{1}{2}$$

The effective arrival rate $\lambda' = \mu(1 - P_0)$

$$\text{where } P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - \left(\frac{1}{2}\right)}{1 - \left(\frac{1}{2}\right)^3} = \frac{4}{7}$$

$$\text{The effective arrival rate } \lambda' = \mu(1 - P_0) = 4 \left[1 - \left(\frac{4}{7}\right) \right] = 1.7143$$

EXAMPLE 6.57 If $\lambda = 6$ per hour and $\mu = 6$ per hour in an (M/M/1): (5/FIFO) model, find the probability that there is no customer in the system.

Solution Given: $\lambda = \mu = 6$ per hour for (M/M/1): (N/FIFO) model with $N = 5$. Therefore,

$$P_0 = \frac{1}{N+1}, \lambda = \mu$$

The probability that there is no customer in the system,

$$P_0 = \frac{1}{5+1} = 0.1667$$

EXAMPLE 6.58 A one-person barbershop has 6 chairs to accommodate people waiting for haircut. Assume customers who arrive when all 6 chairs are full, leave without entering the barbershop. Customers arrive at the average rate of 3 per hour and spend an average of 15 minutes in the barbershop. Find (i) the probability that a customer can get directly into the barber chair upon arrival, (ii) the expected number of customers, waiting for haircut, (iii) the effective arrival rate, and (iv) the time a customer can expect to spend in the barbershop.

Solution Given: 1 chair in service and 6 chairs to accommodate waiting people = $6 + 1 = 7$, i.e. it is one-man barbershop with 7-seat capacity.
 \therefore It is an M/M/1 finite capacity model problem.

Given: average arrival rate = $\lambda = 3/\text{hour}$

$$\text{Service time for customer} = \frac{1}{\mu} = 15 \text{ minutes}$$

$$\therefore \text{Service rate} = \mu = \frac{1}{15}/\text{minute} = 4/\text{hour} \quad (\lambda \text{ is given in hour})$$

$$N = \text{capacity of the system} = 7$$

Performance measures:

- (i) The customer will directly go to barberchair when the system at the time of his arrival is empty. The probability in this situation is given by P_0 ,

$$\begin{aligned} P_0 &= \frac{1-\rho}{1-\rho^{N+1}}, \rho = \frac{3}{4} \quad \left(\because \rho = \frac{\lambda}{\mu} \right) \\ &= \frac{1-0.75}{1-(0.75)^7+1} = 0.2778 \end{aligned}$$

- (ii) The expected number of customers waiting for the haircut is the expected number of customers in the queue,

$$\begin{aligned} L_q &= \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} - (1-P_0) \\ &= \frac{0.75}{1-0.75} - \frac{8 \times 0.75^8}{1-0.75^8} - (1-0.2778) = 1.36 \end{aligned}$$

600 ◇ Probability and Queueing Theory

$$\begin{aligned}\text{(iii) The effective arrival rate} &= \mu(1 - P_0) \\ &= 4(1 - 0.2778) \\ &= 2.89/\text{hour}\end{aligned}$$

(iv) The time a customer can expect to spend in the barbershop (system),

$$\begin{aligned}&= \frac{L_s}{\mu'} = \frac{L_q + 1 - P_0}{\text{Effective arrival rate}} \\ &= \frac{1.36 + 1 - 0.2778}{2.89} \\ L_q &= 43.2 \text{ minutes}\end{aligned}$$

EXAMPLE 6.59 In a railway marshalling yard, goods trains arrive at the rate of 30 trains per day. Assume that the inter-arrival time follows an exponential distribution and the service time is also to be assumed as exponential with mean of 36 minutes. Calculate (i) the probability that the yard is empty, and (ii) the average queue length, assuming the line capacity of the yard is 9 trains.

Solution One yard and the maximum capacity is 9.

∴ It is an M/M/1/N model

$$\begin{aligned}\text{Given:} \quad \text{arrival rate} &= \lambda = 30/\text{day} \\ &= \frac{30}{24 \times 60} = \frac{1}{48}/\text{minute}\end{aligned}$$

$$\text{Service time for customer} = \frac{1}{\mu} = 36 \text{ minutes}$$

$$\text{i.e.} \quad \mu = \frac{1}{36}/\text{minute}$$

$$\text{Here} \quad N = 9$$

$$\text{Traffic intensity} \quad \rho = \frac{\lambda}{\mu} = \frac{36}{48} = \frac{3}{4} = 0.75$$

Performance measures:

(i) The probability that the yard is empty,

$$\begin{aligned}P_0 &= \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - 0.75}{1 - (0.75)^{10}} \\ &= \frac{0.25}{0.90} = 0.28\end{aligned}$$

(ii) The average queue length (system),

$$\begin{aligned} L_s &= \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} \\ &= \frac{0.75}{1-0.75} - \frac{10 \times (0.75)^{10}}{1-(0.75)^{10}} = 3 \text{ trains} \end{aligned}$$

Note: Since the arrival rate λ is given in hour and the service time μ is given in minutes, we can either convert to minutes or to hours.

EXAMPLE 6.60 A barbershop has space to accommodate only 10 customers. That can serve only one person at a time. If a customer comes to the shop and finds it full, he goes to the next shop. Customers randomly arrive at an average rate $\lambda = 10$ per hour and the barber service time is exponential with an average of 5 minutes per customer. Find (i) P_0 , and (ii) P_n .

Solution Given: one barbershop with maximum capacity 10.
 \therefore It is an M/M/1/N model problem.

$$\begin{aligned} N &= 10 \\ \text{Arrival rate} &= \lambda = 10/\text{hour} \\ \text{Service time} &= \frac{1}{\mu} = 5 \text{ minutes} \\ \therefore \text{Service rate} &= \mu = \frac{1}{5} / \text{minute} \\ \text{But } \lambda \text{ is given in hour.} \\ \therefore \mu &= \frac{1}{5} \times 60 = 12/\text{hour} \\ \text{Traffic intensity} &= \rho = \frac{\lambda}{\mu} = \frac{5}{6} = 0.8333 \end{aligned}$$

Performance measures:

$$(i) P_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{1-0.8333}{1-(0.8333)^{11}} = 0.1926$$

$$(ii) P_n = \rho^n P_0 = (0.8333)^n \times 0.1926, n = 0, 1, 2, 3, \dots, 10$$

EXAMPLE 6.61 In a single-server queueing system with Poisson input and exponential service times, if the mean arrival rate is 3 calling units per hour, the expected service time is 0.25 hour, and the maximum possible number of calling units in the system is 2, find (i) $P_n (n \geq 0)$, (ii) the average number of calling units in the system, (iii) the average number of calling units in the queue, (iv) the average waiting time in the system, and (v) the average waiting time in the queue.

602 ◇ Probability and Queueing Theory

Solution It is a single-server Poisson queue model with finite capacity.
 \therefore It is an M/M/1/N model.

Given: Arrival rate $\lambda = 3/\text{hour}$

$$\text{Service time} = \frac{1}{\mu} = 0.25/\text{hour}$$

$$\text{Service rate} = \mu = \frac{1}{0.25} = 4/\text{hour}$$

Maximum capacity $= N = 2$

$$P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - 0.75}{1 - (0.75)^3} = 0.4324$$

Performance measures:

- (i) $P_n = \rho^n P_0 = (0.75)^n \times 0.4324, n \geq 1$
- (ii) The average number of calling units in the system,

$$L_s = \frac{0.75}{1 - 0.75} - \frac{3 \times (0.75)^3}{1 - (0.75)^3} = 0.8 \text{ unit}$$

- (iii) The average number of calling units in the queue,

$$L_q = L_s - (1 - P_0) = 0.8 - (1 - 0.4324) = 0.24$$

- (iv) The average waiting time in the system,

$$\lambda' = \mu(1 - P_0) = 4(1 - 0.4324) = 2.2702$$

$$\begin{aligned} W_s &= \frac{1}{\lambda'} L_s = 0.4405 \times 0.8 \\ &= 0.3524 \text{ hour} = 21.144 \text{ minutes} \end{aligned}$$

- (v) The average waiting time in the queue,

$$\begin{aligned} W_q &= \frac{1}{\lambda'} L_q = 0.24 \times 0.4405 \\ &= 0.1057 \text{ hour} = 6.34 \text{ minutes} \end{aligned}$$

EXAMPLE 6.62 The local one-person barbershop can accommodate a maximum of 5 people at a time (4 waiting and 1 getting haircut). Customers arrive according to a Poisson distribution with mean 5 per hour. The barber cuts hair at an average rate of 4 per hour (exponential service time). (i) What percentage of time is the barber idle? (ii) What fraction of the potential customers are turned away? (iii) What is the expected number of customers waiting for a haircut? (iv) How much time can a customer expect to spend in the barbershop?

Solution One-person barbershop can accommodate a maximum of 5 people at a time.

∴ It is an M/M/1/N queueing model problem.

Arrivals follow Poisson distribution with

$$\begin{aligned}\text{mean} &= \lambda = 5/\text{hour} \\ \text{average service rate} &= \mu = 4/\text{hour} \\ \text{maximum capacity} &= N = 5\end{aligned}$$

$$\rho = \frac{\lambda}{\mu} = \frac{5}{4} = 1.25$$

Performance measures:

(i) The barber will be idle only when there is no customer in the system,

$$P(\text{the barber is idle}) = P(n = 0)$$

$$P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - 1.25}{1 - (1.25)^6} = 0.0888$$

∴ The percentage of time when the barber is idle

$$= 0.0888 \times 100 \approx 9\%.$$

(ii) A customer will turn away only when the system is full. That is, when $N \geq 5$,

$$\begin{aligned}\therefore P(\text{a customer is turned away}) P(N \geq 5) &= \rho^5 P_0 = (1.25)^5 \times 0.0888 \\ &= 0.2711\end{aligned}$$

(iii) The expected number of customers waiting for haircut is the expected number of customers in the queue,

$$\begin{aligned}L_q &= L_s - (1 - P_0) \\ &= \frac{\rho}{1 - \rho} - \frac{(N + 1)\rho^{N+1}}{1 - \rho^{N+1}} - (1 - P_0) \\ &= \frac{1.25}{1 - 1.25} - \frac{6(1.25)^6}{1 - (1.25)^6} - (1 - 0.0888) = 2.2205\end{aligned}$$

(iv) The expected time a customer spend in the barbershop (system),
Effective arrival rate $\lambda' = \mu(1 - P_0)$

$$L_s = L_q + (1 - P_0) = 2.2205 + (1 - 0.0888) = 3.1317$$

$$W_s = \frac{L_s}{\lambda'} = \frac{L_s}{\mu(1 - P_0)} = \frac{3.1317}{4 \times (1 - 0.0888)} = 0.8592 \text{ hour} = 51.5 \text{ minutes}$$

EXAMPLE 6.63 At a railway station, only one train is handled at a time. The railway yard is sufficient only for 2 trains to wait, while the other is given signal to leave the station. Trains arrive at the station at an average rate of 6 per hour and the railway station can handle them on an average of 6 per hour.

604 ◇ Probability and Queueing Theory

Assuming Poisson arrivals and exponential service distribution, find (i) the probability for the number of trains in the system, (ii) the average waiting time of a new train coming into the yard. If the handling rate is doubled, (iii) how will the above results get modified?

Solution One railway station with maximum capacity 3. Arrivals follow Poisson distribution and service time follows exponential distribution.

∴ It is an M/M/1/N queueing model.

$$\begin{aligned}\text{Given:} \quad & \text{Arrival rate} = \lambda = 6/\text{hour} \\ & \text{Service rate} = \mu = 6/\text{hour} \\ & \text{System capacity} = N = 2 + 1 = 3\end{aligned}$$

$$\text{Since } \lambda = \mu, P_0 = \frac{1}{N+1} = \frac{1}{4}$$

$$\text{The effective arrival rate} = \lambda' = \mu(1 - P_0)$$

Performance measures:

- (i) The probability for the number of trains in the system,
Since $\lambda = \mu$

$$P_n = \frac{1}{N+1} = \frac{1}{4} \text{ for } (n = 1, 2, 3)$$

- (ii) The average waiting time of a new train in the yard (system),

$$\begin{aligned}L_s &= \frac{N}{2} = 1.5 \text{ trains} \\ W_s &= \frac{1}{\lambda'} L_s \\ &= \frac{1.5}{\mu(1 - P_0)} = \frac{1.5}{6 \times \frac{3}{4}} = \frac{1}{3} \text{ hour} = 20 \text{ minutes}\end{aligned}$$

- (iii) When the handling is doubled, i.e. $\mu = 6 + 6 = 12/\text{hour}$, then $\lambda = 6$,

$$\mu = 12, N = 3, \rho = \frac{\lambda}{\mu} = \frac{6}{12} = \frac{1}{2} = 0.5$$

$$\text{Since } \lambda \neq \mu, P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} = \frac{1 - 0.5}{1 - (0.5)^4} = 0.5333$$

The probability for the number of trains in the system,

$$P_n = \rho^n P_0 = (0.5)^n \times 0.5333, n = 1, 2, 3$$

The average waiting time of a new train in the yard (system),

$$W_s = \frac{L_s}{\lambda'}$$

$$\begin{aligned}
 L_s &= \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} \\
 &= \frac{0.5}{1-0.5} - \frac{4 \times (0.5)^4}{1-(0.5)^4} = 0.73 \\
 W_s &= \frac{L_s}{\lambda'} = \frac{L_s}{\mu(1-P_0)} = \frac{0.73}{12 \times (1-0.5333)} \\
 &= 0.131 \text{ hour} = 7.9 \text{ minutes}
 \end{aligned}$$

EXAMPLE 6.64 Patients arrive at a clinic according to Poisson distribution at a rate of 30 patients per hour. The waiting room does not accommodate more than 14 patients. Examination time per patient is exponential with mean rate of 20 per hour. (i) Find the effective arrival rate at a clinic. (ii) What is the probability that an arriving patient will not wait? (iii) What is the expected waiting time until a patient is discharged from the clinic?

[AU December '09]

Solution One clinic with maximum capacity 15 (= 14 + 1 in service).

Patients arrive according to Poisson distribution and service (examination) time follows exponential distribution.

It is an M/M/1/N queueing model.

$$\text{Arrival rate} = \lambda = 30/\text{hour}$$

Examination (service) time follows exponential distribution with mean rate 20/hour.

∴

$$\text{Service rate} = \mu = 20/\text{hour}$$

$$\text{Maximum capacity } N = 14 + 1 = 15 (\text{14 accommodate} + 1 \text{ in service})$$

$$\rho = \frac{\lambda}{\mu} = \frac{30}{20} = \frac{3}{2} = 1.5$$

$$\text{Since } \lambda \neq \mu, P_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{1-1.5}{1-(1.5)^{16}} = 0.00076$$

Performance measures:

(i) The effective arrival rate,

$$\lambda' = \mu(1 - P_0) = 20 \times (1 - 0.00076) = 19.98/\text{hour}$$

(ii) An arriving patient will not wait only when the system is empty.

∴ The probability that an arriving patient will not wait is $P_0 = 0.00076$

(iii) The expected waiting time until a patient is discharged from the clinic (system),

$$W_s = \frac{1}{\lambda'} L_s$$

$$L_s = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} = \frac{1.5}{1-1.5} - \frac{16(1.5)^{16}}{1-(1.5)^{16}} = 13 \text{ nearly}$$

$$W_s = \frac{1}{\lambda'} L_s = \frac{13}{19.98} = 0.65 \text{ hour} = 39 \text{ minutes}$$

6.7 MODEL IV — (M/M/c): (N/FIFO) MULTIPLE SERVER WITH FINITE CAPACITY POISSON QUEUEING MODEL

This model differs from the previous model in the sense that, in that model we have considered the queue with single-service channel, whereas the number of service channels here is c .

This queueing system is the same as that of model II except that the maximum number of customers in the system is limited to N where $N > c$ (c is the number of servers). Therefore, for this (M/M/c): (N/FIFO) model,

$$\begin{aligned}\lambda_n &= \lambda, & \text{for } 0 \leq n \leq N \\ &= 0, & \text{for } n > N \\ \mu_n &= n\mu, & \text{for } 0 \leq n < c \\ &= c\mu, & \text{for } c \leq n \leq N\end{aligned}$$

Using these values of λ_n and μ_n and in Eqs. (6.21) and (6.22) noting that $1 < c < N$, we get

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1}, \quad \rho = \frac{\lambda}{c\mu} < 1$$

and

$$\begin{aligned}P_n &= \frac{(c\rho)^n}{n!} P_0, & 0 \leq n < c \\ &= \frac{c^c}{c!} \rho^n P_0, & c \leq n \leq N \\ &= 0, & n > N\end{aligned}$$

6.7.1 Performance Measures

1. The average queue length or the expected number of customers in the queue,

$$L_q = E(N - c) = \sum_{n=c}^N (n - c) P_n$$

608 ◇ Probability and Queueing Theory

3. The average waiting time in the system and in the queue:
By the modified Little's formulas,

$$W_s = \frac{L_s}{\lambda'}$$

and

$$W_q = \frac{1}{\lambda'} L_q$$

where λ' is the effective arrival rate.

EXAMPLE 6.65 There are 3 servers in a 4-capacity service centre with $\lambda = 3$ per hour and $\mu = 4$ per hour. Calculate the probability that there are 2 customers in the centre.

Solution Given: $\lambda = 3/\text{hour}$, $\mu = 4/\text{hour}$, $c = 3$ and $N = 4$

\therefore It is an M/M/c/N queueing model problem.

$$\rho = \frac{\lambda}{c\mu} = \frac{3}{12} = 0.25$$

$$c\rho = \frac{3}{4} = 0.75$$

$$\begin{aligned} \text{Now, } P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\ P_0 &= \left[\sum_{n=0}^2 \frac{(0.75)^n}{n!} + \frac{3^3}{3!} \sum_{n=3}^4 (0.25)^n \right]^{-1} = 0.4719 \end{aligned}$$

We know that,

$$P_n = \frac{(c\rho)^n}{n!} P_0, \quad 0 \leq n < c$$

$$P = \frac{(0.75)^2}{2!} \times 0.4719 = 0.1327$$

EXAMPLE 6.66 A two-person health clinic has 2 chairs for waiting customers. If $\lambda = 2$ per hour and $\mu = 3$ per hour, calculate the probability that the system is empty.

Solution Given: $\lambda = 2/\text{hour}$, $\mu = 3/\text{hour}$, $c = 2$ and $N = 4$ (2 in service and 2 waiting customers)

\therefore It is an M/M/c/N queueing model problem

$$\rho = \frac{\lambda}{c\mu} = \frac{2}{6} = \frac{1}{3}$$

$$\therefore c\rho = \frac{2}{3}$$

The probability that the system is empty is given by

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\ &= \left[\sum_{n=0}^1 \frac{\left(\frac{2}{3}\right)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^4 \left(\frac{1}{3}\right)^n \right]^{-1} = 0.5006 \end{aligned}$$

EXAMPLE 6.67 If $\lambda = 2$ per hour and $\mu = 3$ per hour in an M/M/4/N queueing system and there are 2 chairs for waiting customers, calculate the probability that there are 7 customers in the system.

Solution In the system, 4 in service and 2 chairs for waiting customers, i.e. $N = 6$.

The system capacity is only 6.

Therefore, the probability that there are 7 customers in the system is 0.

EXAMPLE 6.68 If $\lambda/\mu = 3/4$, $c = 2$ and $N = 4$ in an M/M/c/N queueing system, find the probability that there are 3 customers in the system.

Solution Given: $\frac{\lambda}{\mu} = \frac{3}{4}$, $c = 2$ and $N = 4$

$$\rho = \frac{\lambda}{c\mu} = \frac{3}{8} \quad \therefore c\rho = \frac{3}{4}$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\ &= \left[\sum_{n=0}^1 \frac{(0.75)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^4 \left(\frac{3}{8}\right)^n \right]^{-1} = 0.4595 \end{aligned}$$

To find the probability that there are 3 customers in the system,

$$P_n = \frac{c^c}{c!} \rho^n P_0, \quad c \leq n \leq N$$

$$P_3 = \frac{2^2}{2!} \left(\frac{3}{8}\right)^3 \times 0.4595 = 0.0485$$

610 ◇ Probability and Queueing Theory

EXAMPLE 6.69 A two-person barbershop has 5 chairs to accommodate waiting customers. Potential customers, who arrive when all 5 chairs are full, leave without entering barbershop. Customers arrive at the average rate of 4 per hour and spend an average of 12 minutes in the barbers chair. Compute P_0 , P_1 , P_7 , L_q , L_s and W_s . [AU December '03; '06]

Solution The situation in this problem is finite capacity, multiserver Poisson queue model.

∴ It is an M/M/c/N queueing model problem.

Given, Arrival rate = $\lambda = 4/\text{hour}$

Service rate = $\mu = 5/\text{hour}$

$c = \text{number of servers} = 2$

and $N = \text{maximum capacity} = 2 + 5 = 7$ (5 accommodate + 2 in service)

$$\rho = \frac{\lambda}{c\mu} = \frac{4}{2 \times 5} = 0.4$$

Performance measures:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1}$$

$$= \left[\sum_{n=0}^1 \frac{(2 \times 0.4)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^7 (0.4)^n \right]^{-1} = 0.4287$$

$$P_n = \frac{(c\rho)^n}{n!} P_0, 0 \leq n < c$$

$$P_1 = \frac{(2 \times 0.4)^1}{1!} P_0 = 0.8 \times 0.4287 = 0.34296$$

$$P_n = \frac{c^c}{c!} \rho^n P_0, \quad c \leq n \leq N$$

$$= 0, \quad n > N$$

$$P_7 = \frac{2^2}{2!} (0.4)^7 \times 0.4287 = 0.0014$$

The average queue length,

$$L_q = \frac{(c\rho)^c \rho}{c!(1-\rho)^2} [1 - \rho^{N-c+1} - (1-\rho)(N-c+1)\rho^{N-c}] P_0$$

$$= \frac{(0.8)^2 \times 0.4}{2!(1-0.4)^2} [1 - (0.4)^6 - (1-0.4) \times 6(0.4)^5] \times 0.4287 = 0.462$$

The average number of customers in the system,

$$\begin{aligned}
 L_s &= L_q + c - \sum_{n=0}^{c-1} (c-n) P_n \\
 &= 0.1462 + 2 - \sum_{n=0}^1 (2-n) P_n \\
 &= 2.1462 - (2 \times P_0 + 1 \times P_1) \\
 &= 2.1462 - (2 \times 0.4287 + 1 \times 0.3430) \\
 &= 0.9458 \text{ customer}
 \end{aligned}$$

The waiting time of a customer in the system,

$$\begin{aligned}
 W_s &= \frac{L_s}{\lambda'} \\
 \text{where } \lambda' &= \mu \left[c - \sum_{n=0}^{c-1} (c-n) P_n \right] \\
 &= 5[2 - (2 \times 0.4287 + 1 \times 0.3430)] \\
 &= 3.998 \\
 W_s &= \frac{0.9458}{3.998} = 0.2365 \text{ hour} \\
 &= 14.2 \text{ minutes}
 \end{aligned}$$

EXAMPLE 6.70 At a port, there are 6 unloading berths and 4 unloading (crews). When all the berths are full, arriving ships are diverted to an overflow facility 20 km down the river. Tankers arrive according to a Poisson process with a mean of 1 every 2 hours. It takes for an unloading crew, on the average, 10 hours to unload a tanker, the unloading time following an exponential distribution. Find: (i) How many tankers are at the port on the average? (ii) How long does a tanker spend at the port on the average? (iii) What is the average arrival rate at the overflow facility?

Solution This problem is a finite capacity, multiserver Poisson queue model.
 \therefore It is an M/M/c/N queueing model problem.

$$\text{Arrival rate} = \frac{1}{2} \text{ hour or } \lambda = \frac{1}{2} / \text{hour}$$

$$\mu = \frac{1}{10} / \text{hour}, c = 4 \text{ and } N = 6$$

$$\text{Unloading time} = \text{service time} = \frac{1}{\mu} = 10 \text{ hours}$$

612 ◇ Probability and Queueing Theory

$$\text{Service rate (unloading)} = \mu = \frac{1}{10} \text{ /hour}$$

$$\rho = \frac{\lambda}{c\mu} = \frac{10}{8} = 1.25$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\ &= \left[\sum_{n=0}^3 \frac{(4 \times 1.25)^n}{n!} + \frac{4^4}{4!} \sum_{n=4}^6 (1.25)^n \right]^{-1} = 0.0072 \end{aligned}$$

Performance measures:

- (i) The average number of tankers at the port,

$$\begin{aligned} L_q &= \frac{(c\rho)^c \rho}{c!(1-\rho)^2} [1 - \rho^{N-c+1} - (1-\rho)(N-c+1)\rho^{N-c}] P_0 \\ &= \frac{(5)^4 \times 1.25}{4!(1-1.25)^2} [1 - (1.25)^3 - (1-1.25) \times 3(1.25)^2] \times 0.0072 \\ &= 0.8203 \text{ tanker} \end{aligned}$$

$$\begin{aligned} L_s &= L_q + c - \sum_{n=0}^{c-1} (c-n) P_n \\ &= 0.8203 + 4 - \sum_{n=0}^3 (4-n) P_n \\ &= 4.8203 - \sum_{n=0}^3 (4-n) \frac{5^n}{n!} P_0 = 4.3535 \end{aligned}$$

- (ii) The average time spent by the tanker at the port (system),

$$W_s = \frac{L_s}{\lambda'}$$

$$\begin{aligned} \text{where } \lambda' &= \mu \left[c - \sum_{n=0}^{c-1} (c-n) P_n \right] \\ &= \frac{1}{10} \left\{ 4 - \left[\sum_{n=0}^3 (4-n) \frac{5^n}{n!} \right] \times 0.0072 \right\} = 0.3533 \end{aligned}$$

$$W_s = \frac{4.3535}{0.3533} = 12.32 \text{ hours}$$

(iii) When $n = 6 = N$, the number of tankers in the port is 6, overflow occurs.

$$P_n = \frac{c^c}{c!} \rho^n P_0, \quad n = N$$

$$P_6 = P(n=6) = \frac{4^4}{4!} (1.25)^6 \times 0.0072 = 0.2930$$

The average arrival rate at the overflow facility = (average arrival rate at the port) \times (probability that overflow occurs)

$$= \frac{1}{2} \times 0.2930 = 0.586/\text{hour}$$

EXAMPLE 6.71 A car-servicing station has 2 bays where service can be offered simultaneously. Because of space limitation, only 4 cars are accepted for servicing. The arrival pattern is Poisson with 12 cars per day. The service time in both the bays is exponentially distributed with $\mu = 8$ cars per day per bay. Find (i) the average number of cars in the service station, (ii) the average number of cars waiting for service, and (iii) the average time a car spends in the system. [AU December '03; '07]

Solution A car-servicing station has 2 bays, the arrival pattern is Poisson and the service time is exponentially distributed. The system can accommodate only 4 cars.

∴ It is an M/M/c/N queueing model problem.

The arrival pattern is Poisson with mean $\lambda = 12/\text{day}$.

Service time follows exponential distribution with $\mu = 8/\text{day}$.

Number of service station $c = 2$, space limitation $N = 4$.

$$\rho = \frac{\lambda}{c\mu} = \frac{12}{16} = 0.75$$

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1}$$

$$= \left[\sum_{n=0}^1 \frac{(1.5)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^4 (0.75)^n \right]^{-1} = 0.1960$$

Performance measures:

(i) The average number of cars waiting for service

$$L_q = \frac{(c\rho)^c \rho}{c!(1-\rho)^2} \left[1 - \rho^{N-c+1} - (1-\rho)(N-c+1)\rho^{N-c} \right] P_0$$

$$= \frac{(1.5)^2 (0.75)}{2!(1-0.75)^2} [1 - (0.75)^3 - (1-0.75)3(0.75)^2] \times 0.1960$$

$$= 0.4135 \text{ car}$$

614 ◇ Probability and Queueing Theory

(ii) The average number of cars in the service station (system)

$$\begin{aligned}
 L_s &= L_q + c - \sum_{n=0}^{c-1} (c-n) P_n \\
 &= 0.4135 + 2 - \sum_{n=0}^1 (2-n) \frac{(1.5)^n}{n!} P_0 \\
 &= 2.4135 - (2 \times 0.1960 + 1.5 \times 0.1960) \\
 &= 1.73 \approx 2 \text{ cars}
 \end{aligned}$$

(iii) The average time a car spends in the system,

$$\begin{aligned}
 W_s &= \frac{1}{\lambda'} L_s \\
 \text{where } \lambda' &= \mu \left[c - \sum_{n=0}^{c-1} (c-n) P_n \right] \\
 &= 8[2 - (2P_0 + P_1)] \\
 &= 10.512 \\
 \therefore W_s &= \frac{1.73}{10.512} = 0.1646 \text{ day}
 \end{aligned}$$

EXAMPLE 6.72 A group of engineers has 2 terminals available to aid in their calculations. The average computing job requires 20 minutes of terminal time and each engineer requires some computation about once every half an hour. Assume that these are distributed according to an exponential distribution. If there are 6 engineers in the group, find (i) the expected number of engineers waiting to use one of the terminals, (ii) the expected number of engineers waiting in the computing centre, (iii) the average time spent by the engineer in the centre, and (iv) the total time lost per day.

Solution There are 6 engineers in the group, the arrival pattern and the service time is exponentially distributed. The system has 2 terminals.

∴ It is an M/M/c/N queueing model problem.

Each engineer requires some computation about once every half an hour.

$$\text{Service time} = \frac{1}{\mu} = 20 \text{ minutes} \Rightarrow \mu = \frac{1}{20} \text{ minute} \Rightarrow \mu = \frac{1}{20} \times 60 = 3/\text{hour}$$

$$\text{Number of terminals} = c = 2$$

$$\text{Group of engineers} = N = 6$$

$$\begin{aligned}
 \lambda &= 2/\text{hour}, \mu = 3/\text{hour}, c = 2 \text{ and } N = 6 \\
 &\quad (\text{Given: Once every half an hour})
 \end{aligned}$$

$$\therefore \rho = \frac{\lambda}{c\mu} = \frac{2}{2 \times 3} = 0.333 \text{ and } c\rho = 0.667$$

$$\begin{aligned}
 P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\
 &= \left[\sum_{n=0}^1 \frac{(0.667)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^6 (0.333)^n \right]^{-1} = 0.5003
 \end{aligned}$$

Performance measures:

- (i) The expected number of engineers waiting to use one of the terminals,

$$\begin{aligned}
 L_q &= \frac{(c\rho)^c \rho}{c!(1-\rho)^2} [1 - \rho^{N-c+1} - (1-\rho)(N-c+1)\rho^{N-c}] P_0 \\
 &= \frac{(0.667)^2 (0.333)}{2!(1-0.333)^2} [1 - (0.333)^5 - 5(1-0.333)(0.333)^4] \times 0.5003 \\
 &= 0.0796 \text{ engineer}
 \end{aligned}$$

- (ii) The expected number of engineers waiting in the computing centre (system),

$$\begin{aligned}
 L_s &= L_q + c - \sum_{n=0}^{c-1} (c-n) P_n \\
 &= 2.0796 - (2P_0 + P_1) \\
 &= 2.0796 - \left(2 \times 0.5003 + \frac{2}{3} \times 0.5003 \right) \\
 &= 0.75 \approx 1 \text{ engineer}
 \end{aligned}$$

- (iii) The average time spent by the engineer in the centre (system),

$$W_s = \frac{1}{\lambda'} L_s$$

$$\begin{aligned}
 \text{where } \lambda' &= \mu \left[c - \sum_{n=0}^{c-1} (c-n) P_n \right] \\
 &= 3[2 - (2P_0 + P_1)] = 3[2 - (2 \times 0.5003 + 0.667 \times 0.5003)] \\
 &= 1.9976
 \end{aligned}$$

$$W_s = \frac{0.75}{1.9976} = 0.3754 \text{ hour}$$

- (iv) Every time an engineer approaches the computing centre, he has to lose 0.3754 hour by way of waiting (0.3754×60 minutes)

⇒ If the day consists of 8 working hours, he has to approach the centre 16 times ($8 \times 0.3754 \times 60 = 16$)

616 ◇ Probability and Queueing Theory

⇒ The time lost in waiting in a day per engineer

$$= 16 \times 0.0398 = 0.6368 \text{ hour}$$

⇒ The time lost in waiting in a day by all the 6 engineers

$$= 6 \times 0.63689 = 3.82 \text{ hours}$$

EXAMPLE 6.73 A barbershop has 2 barbers and 3 chairs for customers. Assume that the customers arrive in Poisson fashion at a rate of 5 per hour and that each barber services customers according to an exponential distribution with mean of 15 minutes. Further, if a customer arrives and there are no empty chairs in the shop, he will leave. (i) What is the probability that the shop is empty? (ii) What is the expected number of customers in the shop?

Solution Two barbers and 3 chairs for customers, the arrival pattern is Poisson and the service time is exponentially distributed.

∴ It is an M/M/c/N queueing model problem.

$$\text{Arrival rate} = \lambda = 5/\text{hour}, c = 2 \text{ and } N = 3$$

Service time follows an exponential distribution with mean of 15 minutes.

$$\therefore \frac{1}{\mu} = 15 \text{ minutes}$$

$$\Rightarrow \mu = \frac{1}{15} \text{ minutes}$$

$$\Rightarrow \mu = \frac{1}{15} \times 60 = 4/\text{hour}$$

$$\rho = \frac{\lambda}{c\mu} = \frac{5}{8} = 0.625$$

$$\therefore c\rho = 1.25$$

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^N \rho^n \right]^{-1} \\ &= \left[\sum_{n=0}^1 \frac{(1.25)^n}{n!} + \frac{2^2}{2!} \sum_{n=2}^3 (0.625)^n \right]^{-1} = 0.28 \end{aligned}$$

Performance measures:

(i) The probability that the shop is empty,

$$P_0 = 0.28$$

(ii) The expected number of customers in the shop,

$$L_s = L_q + c - \sum_{n=0}^{c-1} (c-n) P_n$$

$$\begin{aligned}
 &= \frac{(c\rho)^c \rho}{c!(1-\rho)^2} [1 - \rho^{N-c+1} - (1-\rho)(N-c+1)\rho^{N-c}] P_0 \\
 &\quad + c - \sum_{n=0}^1 (2-n) P_n \\
 &= \frac{(1.25)^2 (0.625)}{2!(1-0.625)^2} [1 - (0.625)^2 - 2(1-0.625)(0.625)] \times 0.28 \\
 &\quad + 2 - \sum_{n=0}^1 (2-n) P_n \\
 &= 1.226 \text{ customers}
 \end{aligned}$$

EXERCISES

1. Define a queue.
2. What are the basic characteristics of a queueing system?
3. What do you understand by Kendall's notation?
4. Give the formula for probability of n units in the system under single-server (M/M/1): (∞ /FIFO).
5. Write the formula for P_0 under (M/M/1):(∞ /FIFO).
6. State the formula for queue length under (M/M/c): (∞ /FIFO).
7. Give an example of first come first served queueing system.
8. Give an example of first come last served queueing system.
9. Write the formula for system queue length under the (M/M/1): (N/FIFO).
10. The inter-arrival time under Poisson queue follows which distribution?
11. Write the formula for waiting time in the queue under (M/M/c): (N/FIFO).
12. Write the formula for waiting time in the system and in the queue under (M/M/1): (∞ /FIFO).
13. Write the formula for waiting time in the queue and in the system under (M/M/c): (∞ /FIFO).
14. Under (M/M/1): (∞ /FCFS) system write the formula for probability of empty system.
15. What is the formula for the probability for a customer to wait in the queue under (M/M/1): (∞ /FCFS)?
16. Write Little's formula for (M/M/c): (∞ /FIFO) queueing model.
17. Write Little's formula for (M/M/1): (N/FIFO) queueing model.
18. Write Little's formula for (M/M/c): (N/FIFO) queueing model.
19. Define transient state and steady state.
20. What is the formula for traffic intensity under the queueing system (M/M/1): (∞ /FCFS)?

618 ◇ Probability and Queueing Theory

21. What is the formula for traffic intensity under the queueing system (M/M/c): (∞ /FCFS)?
22. What are the classifications of queueing models?
23. Give the probability that there is no customer in the (M/M/c): (∞ /FCFS) queueing system.
24. What is the probability that there is no customer in the (M/M/c): (N/FCFS) queueing system?
25. What is the probability that there is no customer in the (M/M/1): (N/FCFS) queueing system?
26. Define effective arrival rate with respect to (M/M/1): (N/FIFO) queueing model.
27. Write the effective arrival rate with respect to (M/M/c): (N/FIFO) queueing model.
28. What is the probability that there are n customers in the (M/M/c): (N/FCFS) queueing system?
29. Customers arrive at the first class ticket counter of a theatre at a rate of 12 per hour. There is one clerk serving the customer at the rate of 30 per hour. What is the probability that there are more than 2 customers in the counter? [Ans. $(2/5)^2$]
30. At a public telephone booth, the arrivals are on the average 15 per hour. A call on the average takes 3 minutes. If there is just one phone, what is the expected number of callers in the booth at any time? [Ans. 3]
31. Given $\lambda = 0.5$ and $\mu = 0.67$, find the waiting time in the queue of (M/M/1): (∞ /FCFS). [Ans. 4.39]
32. Given $\lambda = 10$ per hour, $\mu = 3$ per hour, $P_0 = 0.0213$ and $c = 4$, find the queue length of the system. [Ans. 6.61]
33. There are 2 clerks in a college to receive dues from the students. If the service time for each student is exponential with mean 4 minutes, and the boys arrive in a Poisson fashion at the counter at the rate of 10 per hour, what is the percentage of idle time for each clerk? [Ans. 67%]
34. Given $c = 4$ and $\rho = \lambda/c\mu = 1/2$, find P_0 . [Ans. 0.087]
35. Given $\lambda = 3$ per hour, $\mu = 4$ per hour and maximum capacity $N = 7$, find the queue length of the system. [Ans. 2.11]
36. If $\lambda/c\mu = 2/3$ in an (M/M/c): (∞ /FCFS) queueing system, find the average number of customers in the non-empty queue. [Ans. 2]
37. In a 3-server infinite capacity Poisson queue model if $\lambda/c\mu = 2/3$, find P_0 . [Ans. 0.111]
38. In the usual notation of an (M/M/1): (∞ /FCFS) queueing system if $\lambda = 12$ per hour and $\mu = 24$ per hour, find the average number of customers in the queue, and in the system. [Ans. 1, 0.5]

39. If a customer has to wait in an (M/M/1): (∞ /FCFS) queue system, what is his average waiting time in the queue if $\lambda = 8$ per hour and $\mu = 12$ per hour? [Ans. 5 minutes]
40. What is the probability that a customer has to wait more than 15 minutes to get his service completed in an (M/M/1): (∞ /FCFS) queueing system if $\lambda = 6$ per hour and $\mu = 10$ per hour? [Ans. 0.3679]
41. A branch of Punjab National Bank has only one typist. Since the typing work varies in length (number of pages to be typed), the typing rate is randomly distributed approximately a Poisson distribution with mean service rate of 8 letters per hour. The letters arrive at a rate of 5 per hour during the entire 8-hour workday. If the typewriter is valued at ₹ 150 per hour, determine (i) the equipment utilization, (ii) the percentage of time the arriving letter has to wait, (iii) the average system time, and (iv) the average idle time cost of the typewriter per day.
[Ans. (i) 0.625, (ii) 62.5%, (iii) 20 minutes, and (iv) ₹ 4.50]
42. A repair shop attended by a single mechanic has an average of 4 customers per hour who bring small appliances for a repair. The mechanic inspects them for defects and quite often can fix them right way or otherwise render a diagnosis. This takes him 6 minutes on the average. Arrivals are Poisson and service time has the exponential distribution. You are required to find (i) the proportion of time during which the shop is empty, (ii) the probability of finding at least one customer in the shop, (iii) the average number of customers in the system, and (iv) the average time, including service, spent by a customer.
[Ans. (i) 0.6, (ii) $1 - P_0 = 0.4$, (iii) 2.3, (iv) 10 minutes]
43. A duplicating machine maintained for office is used and operated by people in the office who need to make copies, mostly secretaries. Since the work to be copied varies in length (number of pages of the original) and copies required, the service rate is randomly but it does approximate a Poisson having a mean service rate of 10 jobs an hour. Generally, the requirements for use are random over the entire 8-hour working day but arrive at the rate of 5 per hour. Several people have noted that a waiting line develops occasionally and have questioned the policy of maintaining only one unit. If the time of a secretary is valued at ₹ 3.50 per hour, make an analysis to find (i) the equipment utilization, (ii) the per cent time an arrival has to wait, (iii) the average system time, and (iv) the average cost of waiting and operating the machine.
[Ans. (i) 0.5, (ii) 50%, (iii) 12 minutes (iv) cost/day = ₹ 28]
44. At a one-man barbershop, customers arrive according to Poisson distribution with mean arival rate of 5 per hour and his haircutting time was exponentially distributed with an average haircut taking 10 minutes. It is assumed that because of his excellent reputation, customers were always willing to wait. Calculate the following: (i) the average number of customers

in the shop, (ii) the average number of customers waiting for a haircut, (iii) the percent of time an arrival can walk right in without having to wait, (iv) the percentage of customers who have to wait prior to getting into the barber's chair.

[Ans. (i) 4.8 or 5, (ii) 4 approx., (iii) 16.7%, (iv) 83.3%]

45. In a bank, operating from 10 a.m. to 2 p.m., the cheques are cashed at a single counter. Customer wishing to cash cheques arrive according to a Poisson process at the rate of 20 customers a day. The cashier at the counter takes on an average 10 minutes to cash the cheque. The service time has been shown to be exponentially distributed. Compute (i) the percentage of time the cashier is busy, (ii) the average time a customer is expected to wait, and (iii) the average number of customers waiting in the queue.

[Ans. (i) 83.33%, (ii) 50 minutes, (iii) 4 1/6]

46. The mean arrival rate to a service centre is 3 per hour. The mean service time is found to be 10 minutes per service. Assuming Poisson arrival and exponential service time, find (i) the utilization factor for this service facility, (ii) the probability of 2 units in the system, (iii) the expected number of customers in the queue, and (iv) the expected time in minutes that a customer has to spend in the system.

[Ans. (i) 0.5, (ii) 0.125, (iii) 1, (iv) 0.333]

47. In a bank, there is only one window. A solitary employee performs all the service required and the window continuously remains open from 7.00 a.m. to 1.00 p.m. It has been discovered that the average number of clients is 54 during the day and that the average service time is of 5 minutes per person. Calculate (i) the average number of clients in the system, (ii) the average number of clients in the waiting line (excluding the one being served), and (iii) the average waiting time.

[Ans. (i) 3, (ii) 2.25, (iii) 20 minutes/client]

48. An oil company is constructing a service station on a highway. Traffic analysis indicates that customer's arrivals over most of the days would approximate a Poisson distribution with a mean of 30 automobiles per hour. Previous studies show that one pump could service a mean of 10 automobiles per hour, with the service time distribution approximating the negative exponential. If 4 pumps are installed, (i) what is the probability that an arrival would have to wait in line? Find (ii) the average waiting time, (iii) the average time spent in the system, and (iv) the average number of automobiles in the system.

[Ans. (i) $1 - P_0 - P_1 - P_2 - P_3 - P_4 = 0.3826$,
(ii) 0.0509 hour or 3.05 minutes,
(iii) 0.1509 hour, (iv) 4.53 automobiles]

49. A two-channel queueing system with Poisson arrival has a mean arrival rate of 50 per hour and exponential service with mean service rate of

75 per hour for each channel. Find (i) the probability of an empty system, and (ii) the probability that an arrival in the system will have to wait.

[Ans. (i) 0.83, (ii) 0.167]

50. For a queueing system with 2 service stations each having exponential service time distribution $\mu = 5$ per minute feed by a queue build-up of arrival rate $\lambda = 8$ per minutes, find (i) the average number of customers in the system, and (ii) the average waiting time of a customer in the system.
[Ans. (i) 4.44, (ii) 0.56 mintues]
51. A two-channel waiting line with Poisson arrival has a mean rate of 50 per hour and exponential service with mean rate of 75 per hour for each channel. Find (i) the probability of an empty system, and (ii) the probability that an arrival in the system will have to wait.

[Ans. (i) 0.5, (ii) 0.1114]

52. Find the probability that there are no customers in the system given that the number of channels in parallel is 3, mean arrival rate 24 per hour and mean service rate in each channel is 10 per hour. Compare the average time that a customer is in the system for the following two systems: (i) Five channels in parallel with mean service rate of 10 per hour. (ii) One channel with mean service rate of 30 per hour.

[Ans. 0.019 (i) 6 minutes, (ii) 10 minutes]

53. A railway goods traffic section has 4 claim assistants. Customers with claims against the railway are found to arrive in Poisson fashion at an average rate of 24 per 8-hour day for 6 days. The amount of time that an assistant spends with a claimant is found to have negative exponential distribution with mean service time of 40 minutes. Claimants are processed in the order of their appearance. (i) How many hours a week can an assistant expect to spend with claimants? (ii) How much time, on the average, does claimant spend in the goods train office.

[Ans. (i) 72 hours (ii) 47.2 minutes]

54. For a queueing system with k service stations, each having exponential distribution with mean service rate μ , feed by a queue with built up arrival rate λ , find (i) the average number of customers in the system, and (ii) the average waiting time of a customer in the system if $c = 2$, $\lambda = 8$ per minute and $\mu = 5$ per minute.

[Ans. (i) 4.44, (ii) 0.56 minute]

55. A tax councelling firm has 4 service counters in the office to receive people who have problems and complaints about their income, wealth and sales taxes. Arrivals are at an average 80 persons in an 8-hour service day. Each tax advisor spends an irregular amount of time in servicing the arrivals which have been found to have exponential distribution. The average service time is 20 minutes. Calculate

- (i) The average number of customers in the system.
- (ii) The average number of customers waiting for service.
- (iii) The average time a customer waits in the system and in queue.
- (iv) The number of hours each week a tax consultant spends with customer.
- (v) What is the probability that a customer has to wait for service?
- (vi) What is the expected number of idle tax advisors at any specified time?

[Ans. (i) 6.61, (ii) 3.28, (iii) 0.66 hour; 0.33 hour,
 (iv) 33.3 hours, (v) 0.55, (vi) 0.666]

- 56.** A telephone exchange has 2 long distance operators. The telephone company finds that during the peak load, long distance calls arrive in a Poisson fashion at an average rate of 15 per hour. The length of service on these calls is approximately exponentially distributed with mean length 5 minutes. (i) What is the probability that a subscriber will have to wait for his long distance call during the peak hours of the day? (ii) If the subscribers will wait and are serviced in turn, what is the expected waiting time?
 [Ans. (i) 0.48, (ii) 3.2 minutes]

- 57.** Assume that the goods trains are coming in a yard at the rate of 30 trains per day and suppose that the inter-arrival times follow an exponential distribution. The service time for each train is assumed to be exponential with an average of 36 minutes. If the yard can admit 9 trains at a time, calculate the probability that the yard is empty, and the average queue length.
 [Ans. 0.28, 1.55]

- 58.** A car park contains 5 cars. The arrival of cars is in Poisson at a mean rate of 10 per hour. The length of time each car spends in the car park has negative exponential distribution with mean of 2 minutes. How many cars are in the car park on an average, and what is the probability of a newly arriving customer finding the car park full and leaving to park his car elsewhere?
 [Ans. 0.49, 0.0027]

- 59.** A stenographer is attached to 5 officers for whom she performs stenographic work. She gets calls from the officers at the rate of 4 per hour and takes on the average 10 minutes to attend to each call. If arrival rate is Poisson and service time is exponential, find (i) the average waiting time for an arriving call, (ii) the average number of waiting calls, and (iii) the average time an arriving call spends in the system.

[Ans. (i) 12.45 minutes, (ii) 0.79, (iii) 22.42 minutes]

- 60.** Consider a single-server queueing system with Poisson input, exponential service times. Suppose the mean arrival rate is 3 calling units per hour, the expected service time is 0.25 hours and the maximum permissible calling units in the system is two. Calculate the expected number in the system.
 [Ans. 0.81]

for W_Q was obtained in Reference 6 by using the foregoing approach and then approximating $E[R]$:

$$W_Q \approx \frac{\lambda^k E[S^2] (E[S])^{k-1}}{2(k-1)!(k-\lambda E[S])^2 \left[\sum_{n=0}^{k-1} \frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k-\lambda E[S])} \right]} \quad (8.63)$$

The preceding approximation has been shown to be quite close to W_Q when the service distribution is gamma. It is also exact when G is exponential.

Exercises

1. For the $M/M/1$ queue, compute

- (a) the expected number of arrivals during a service period and
- (b) the probability that no customers arrive during a service period.

Hint: “Condition.”

- *2. Machines in a factory break down at an exponential rate of six per hour. There is a single repairman who fixes machines at an exponential rate of eight per hour. The cost incurred in lost production when machines are out of service is \$10 per hour per machine. What is the average cost rate incurred due to failed machines?
3. The manager of a market can hire either Mary or Alice. Mary, who gives service at an exponential rate of 20 customers per hour, can be hired at a rate of \$3 per hour. Alice, who gives service at an exponential rate of 30 customers per hour, can be hired at a rate of $\$C$ per hour. The manager estimates that, on the average, each customer's time is worth \$1 per hour and should be accounted for in the model. Assume customers arrive at a Poisson rate of 10 per hour
- (a) What is the average cost per hour if Mary is hired? If Alice is hired?
 - (b) Find C if the average cost per hour is the same for Mary and Alice.

4. Suppose that a customer of the $M/M/1$ system spends the amount of time $x > 0$ waiting in queue before entering service.
- (a) Show that, conditional on the preceding, the number of other customers that were in the system when the customer arrived is distributed as $1 + P$, where P is a Poisson random variable with mean λ .
 - (b) Let W_Q^* denote the amount of time that an $M/M/1$ customer spends in queue. As a by-product of your analysis in part (a), show that

$$P\{W_Q^* \leq x\} = \begin{cases} 1 - \frac{\lambda}{\mu} & \text{if } x = 0 \\ 1 - \frac{\lambda}{\mu} + \frac{\lambda}{\mu}(1 - e^{-(\mu-\lambda)x}) & \text{if } x > 0 \end{cases}$$

5. It follows from Exercise 4 that if, in the $M/M/1$ model, W_Q^* is the amount of time that a customer spends waiting in queue, then

$$W_Q^* = \begin{cases} 0, & \text{with probability } 1 - \lambda/\mu \\ \text{Exp}(\mu - \lambda), & \text{with probability } \lambda/\mu \end{cases}$$

where $\text{Exp}(\mu - \lambda)$ is an exponential random variable with rate $\mu - \lambda$. Using this, find $\text{Var}(W_Q^*)$.

6. Suppose we want to find the covariance between the times spent in the system by the first two customers in an $M/M/1$ queueing system. To obtain this covariance, let S_i be the service time of customer i , $i = 1, 2$, and let Y be the time between the two arrivals.
 - (a) Argue that $(S_1 - Y)^+ + S_2$ is the amount of time that customer 2 spends in the system, where $x^+ = \max(x, 0)$.
 - (b) Find $\text{Cov}(S_1, (S_1 - Y)^+ + S_2)$.

Hint: Compute both $E[(S - Y)^+]$ and $E[S_1(S_1 - Y)^+]$ by conditioning on whether $S_1 > Y$.
- *7. Show that W is smaller in an $M/M/1$ model having arrivals at rate λ and service at rate 2μ than it is in a two-server $M/M/2$ model with arrivals at rate λ and with each server at rate μ . Can you give an intuitive explanation for this result? Would it also be true for W_Q ?
8. A facility produces items according to a Poisson process with rate λ . However, it has shelf space for only k items and so it shuts down production whenever k items are present. Customers arrive at the facility according to a Poisson process with rate μ . Each customer wants one item and will immediately depart either with the item or empty handed if there is no item available.
 - (a) Find the proportion of customers that go away empty handed.
 - (b) Find the average time that an item is on the shelf.
 - (c) Find the average number of items on the shelf.
9. A group of n customers moves around among two servers. Upon completion of service, the served customer then joins the queue (or enters service if the server is free) at the other server. All service times are exponential with rate μ . Find the proportion of time that there are j customers at server 1, $j = 0, \dots, n$.
10. A group of m customers frequents a single-server station in the following manner. When a customer arrives, he or she either enters service if the server is free or joins the queue otherwise. Upon completing service the customer departs the system, but then returns after an exponential time with rate θ . All service times are exponentially distributed with rate μ .
 - (a) Find the average rate at which customers enter the station.
 - (b) Find the average time that a customer spends in the station per visit.
11. Consider a single-server queue with Poisson arrivals and exponential service times having the following variation: Whenever a service is completed a departure occurs only with probability α . With probability $1 - \alpha$ the customer, instead of leaving, joins the end of the queue. Note that a customer may be serviced more than once.
 - (a) Set up the balance equations and solve for the steady-state probabilities, stating conditions for it to exist.
 - (b) Find the expected waiting time of a customer from the time he arrives until he enters service for the first time.
 - (c) What is the probability that a customer enters service exactly n times, $n = 1, 2, \dots$?
 - (d) What is the expected amount of time that a customer spends in service (which does not include the time he spends waiting in line)?

Hint: Use part (c).

- (e) What is the distribution of the total length of time a customer spends being served?

Hint: Is it memoryless?

- *12. A supermarket has two exponential checkout counters, each operating at rate μ . Arrivals are Poisson at rate λ . The counters operate in the following way:
- One queue feeds both counters.
 - One counter is operated by a permanent checker and the other by a stock clerk who instantaneously begins checking whenever there are two or more customers in the system. The clerk returns to stocking whenever he completes a service, and there are fewer than two customers in the system.
- Find P_n , proportion of time there are n in the system.
 - At what rate does the number in the system go from 0 to 1? From 2 to 1?
 - What proportion of time is the stock clerk checking?

Hint: Be a little careful when there is one in the system.

13. Two customers move about among three servers. Upon completion of service at server i , the customer leaves that server and enters service at whichever of the other two servers is free. (Therefore, there are always two busy servers.) If the service times at server i are exponential with rate μ_i , $i = 1, 2, 3$, what proportion of time is server i idle?
14. Consider a queueing system having two servers and no queue. There are two types of customers. Type 1 customers arrive according to a Poisson process having rate λ_1 , and will enter the system if either server is free. The service time of a type 1 customer is exponential with rate μ_1 . Type 2 customers arrive according to a Poisson process having rate λ_2 . A type 2 customer requires the simultaneous use of both servers; hence, a type 2 arrival will only enter the system if both servers are free. The time that it takes (the two servers) to serve a type 2 customer is exponential with rate μ_2 . Once a service is completed on a customer, that customer departs the system.
- Define states to analyze the preceding model.
 - Give the balance equations.
- In terms of the solution of the balance equations, find
- the average amount of time an entering customer spends in the system;
 - the fraction of served customers that are type 1.
15. Consider a sequential-service system consisting of two servers, A and B . Arriving customers will enter this system only if server A is free. If a customer does enter, then he is immediately served by server A . When his service by A is completed, he then goes to B if B is free, or if B is busy, he leaves the system. Upon completion of service at server B , the customer departs. Assume that the (Poisson) arrival rate is two customers an hour, and that A and B serve at respective (exponential) rates of four and two customers an hour.
- What proportion of customers enter the system?
 - What proportion of entering customers receive service from B ?
 - What is the average number of customers in the system?
 - What is the average amount of time that an entering customer spends in the system?

16. Customers arrive at a two-server system according to a Poisson process having rate $\lambda = 5$. An arrival finding server 1 free will begin service with that server. An arrival finding server 1 busy and server 2 free will enter service with server 2. An arrival finding both servers busy goes away. Once a customer is served by either server, he departs the system. The service times at server i are exponential with rates μ_i , where $\mu_1 = 4$, $\mu_2 = 2$.
- What is the average time an entering customer spends in the system?
 - What proportion of time is server 2 busy?
17. Customers arrive at a two-server station in accordance with a Poisson process with a rate of two per hour. Arrivals finding server 1 free begin service with that server. Arrivals finding server 1 busy and server 2 free begin service with server 2. Arrivals finding both servers busy are lost. When a customer is served by server 1, she then either enters service with server 2 if 2 is free or departs the system if 2 is busy. A customer completing service at server 2 departs the system. The service times at server 1 and server 2 are exponential random variables with respective rates of four and six per hour.
- What fraction of customers do not enter the system?
 - What is the average amount of time that an entering customer spends in the system?
 - What fraction of entering customers receives service from server 1?
18. Arrivals to a three-server system are according to a Poisson process with rate λ . Arrivals finding server 1 free enter service with 1. Arrivals finding 1 busy but 2 free enter service with 2. Arrivals finding both 1 and 2 busy do not join the system. After completion of service at either 1 or 2 the customer will then either go to server 3 if 3 is free or depart the system if 3 is busy. After service at 3 customers depart the system. The service times at i are exponential with rate μ_i , $i = 1, 2, 3$.
- Define states to analyze the above system.
 - Give the balance equations.
 - In terms of the solution of the balance equations, what is the average time that an entering customer spends in the system?
 - Find the probability that a customer who arrives when the system is empty is served by server 3.
19. The economy alternates between good and bad periods. During good times customers arrive at a certain single-server queueing system in accordance with a Poisson process with rate λ_1 , and during bad times they arrive in accordance with a Poisson process with rate λ_2 . A good time period lasts for an exponentially distributed time with rate α_1 , and a bad time period lasts for an exponential time with rate α_2 . An arriving customer will only enter the queueing system if the server is free; an arrival finding the server busy goes away. All service times are exponential with rate μ .
- Define states so as to be able to analyze this system.
 - Give a set of linear equations whose solution will yield the long-run proportion of time the system is in each state.
- In terms of the solutions of the equations in part (b),
- what proportion of time is the system empty?
 - what is the average rate at which customers enter the system?

20. There are two types of customers. Type 1 and 2 customers arrive in accordance with independent Poisson processes with respective rate λ_1 and λ_2 . There are two servers. A type 1 arrival will enter service with server 1 if that server is free; if server 1 is busy and server 2 is free, then the type 1 arrival will enter service with server 2. If both servers are busy, then the type 1 arrival will go away. A type 2 customer can only be served by server 2; if server 2 is free when a type 2 customer arrives, then the customer enters service with that server. If server 2 is busy when a type 2 arrives, then that customer goes away. Once a customer is served by either server, he departs the system. Service times at server i are exponential with rate μ_i , $i = 1, 2$.

Suppose we want to find the average number of customers in the system.

- (a) Define states.
- (b) Give the balance equations. Do not attempt to solve them.

In terms of the long-run probabilities, what is

- (c) the average number of customers in the system?
- (d) the average time a customer spends in the system?

- *21. Suppose in Exercise 20 we want to find out the proportion of time there is a type 1 customer with server 2. In terms of the long-run probabilities given in Exercise 20, what is

- (a) the rate at which a type 1 customer enters service with server 2?
- (b) the rate at which a type 2 customer enters service with server 2?
- (c) the fraction of server 2's customers that are type 1?
- (d) the proportion of time that a type 1 customer is with server 2?

22. Customers arrive at a single-server station in accordance with a Poisson process with rate λ . All arrivals that find the server free immediately enter service. All service times are exponentially distributed with rate μ . An arrival that finds the server busy will leave the system and roam around “in orbit” for an exponential time with rate θ at which time it will then return. If the server is busy when an orbiting customer returns, then that customer returns to orbit for another exponential time with rate θ before returning again. An arrival that finds the server busy and N other customers in orbit will depart and not return. That is, N is the maximum number of customers in orbit.

- (a) Define states.
- (b) Give the balance equations.

In terms of the solution of the balance equations, find

- (c) the proportion of all customers that are eventually served;
- (d) the average time that a served customer spends waiting in orbit.

23. Consider the $M/M/1$ system in which customers arrive at rate λ and the server serves at rate μ . However, suppose that in any interval of length h in which the server is busy there is a probability $\alpha h + o(h)$ that the server will experience a breakdown, which causes the system to shut down. All customers that are in the system depart, and no additional arrivals are allowed to enter until the breakdown is fixed. The time to fix a breakdown is exponentially distributed with rate β .

- (a) Define appropriate states.
- (b) Give the balance equations.

In terms of the long-run probabilities,

- (c) what is the average amount of time that an entering customer spends in the system?

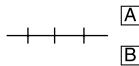


Figure 8.4

- (d) what proportion of entering customers complete their service?
 (e) what proportion of customers arrive during a breakdown?
- *24. Reconsider Exercise 23, but this time suppose that a customer that is in the system when a breakdown occurs remains there while the server is being fixed. In addition, suppose that new arrivals during a breakdown period are allowed to enter the system. What is the average time a customer spends in the system?
25. Poisson (λ) arrivals join a queue in front of two parallel servers A and B , having exponential service rates μ_A and μ_B (see Figure 8.4). When the system is empty, arrivals go into server A with probability α and into B with probability $1 - \alpha$. Otherwise, the head of the queue takes the first free server.
- Define states and set up the balance equations. Do not solve.
 - In terms of the probabilities in part (a), what is the average number in the system? Average number of servers idle?
 - In terms of the probabilities in part (a), what is the probability that an arbitrary arrival will get serviced in A ?
26. In a queue with unlimited waiting space, arrivals are Poisson (parameter λ) and service times are exponentially distributed (parameter μ). However, the server waits until K people are present before beginning service on the first customer; thereafter, he services one at a time until all K units, and all subsequent arrivals, are serviced. The server is then “idle” until K new arrivals have occurred.
- Define an appropriate state space, draw the transition diagram, and set up the balance equations.
 - In terms of the limiting probabilities, what is the average time a customer spends in queue?
 - What conditions on λ and μ are necessary?
27. Consider a single-server exponential system in which ordinary customers arrive at a rate λ and have service rate μ . In addition, there is a special customer who has a service rate μ_1 . Whenever this special customer arrives, she goes directly into service (if anyone else is in service, then this person is bumped back into queue). When the special customer is not being serviced, she spends an exponential amount of time (with mean $1/\theta$) out of the system.
- What is the average arrival rate of the special customer?
 - Define an appropriate state space and set up balance equations.
 - Find the probability that an ordinary customer is bumped n times.
- *28. Let D denote the time between successive departures in a stationary $M/M/1$ queue with $\lambda < \mu$. Show, by conditioning on whether or not a departure has left the system empty, that D is exponential with rate λ .

Hint: By conditioning on whether or not the departure has left the system empty we see that

$$D = \begin{cases} \text{Exponential}(\mu), & \text{with probability } \lambda/\mu \\ \text{Exponential}(\lambda) * \text{Exponential}(\mu), & \text{with probability } 1 - \lambda/\mu \end{cases}$$

where $\text{Exponential}(\lambda) * \text{Exponential}(\mu)$ represents the sum of two independent exponential random variables having rates μ and λ . Now use moment-generating functions to show that D has the required distribution.

Note that the preceding does not prove that the departure process is Poisson. To prove this we need show not only that the interdeparture times are all exponential with rate λ , but also that they are independent.

29. Potential customers arrive to a single-server hair salon according to a Poisson process with rate λ . A potential customer who finds the server free enters the system; a potential customer who finds the server busy goes away. Each potential customer is type i with probability p_i , where $p_1 + p_2 + p_3 = 1$. Type 1 customers have their hair washed by the server; type 2 customers have their hair cut by the server; and type 3 customers have their hair first washed and then cut by the server. The time that it takes the server to wash hair is exponentially distributed with rate μ_1 , and the time that it takes the server to cut hair is exponentially distributed with rate μ_2 .
 - (a) Explain how this system can be analyzed with four states.
 - (b) Give the equations whose solution yields the proportion of time the system is in each state.
 In terms of the solution of the equations of (b), find
 - (c) the proportion of time the server is cutting hair;
 - (d) the average arrival rate of entering customers.
30. For the tandem queue model verify that

$$P_{n,m} = (\lambda/\mu_1)^n (1 - \lambda/\mu_1) (\lambda/\mu_2)^m (1 - \lambda/\mu_2)$$

satisfies the balance equation (8.15).

31. Consider a network of three stations with a single server at each station. Customers arrive at stations 1, 2, 3 in accordance with Poisson processes having respective rates 5, 10, and 15. The service times at the three stations are exponential with respective rates 10, 50, and 100. A customer completing service at station 1 is equally likely to
 - (i) go to station 2,
 - (ii) go to station 3, or
 - (iii) leave the system.
 A customer departing service at station 2 always goes to station 3. A departure from service at station 3 is equally likely to either go to station 2 or leave the system.
 - (a) What is the average number of customers in the system (consisting of all three stations)?
 - (b) What is the average time a customer spends in the system?
32. Consider a closed queueing network consisting of two customers moving among two servers, and suppose that after each service completion the customer is equally likely to go to either server—that is, $P_{1,2} = P_{2,1} = \frac{1}{2}$. Let μ_i denote the exponential service rate at server i , $i = 1, 2$.
 - (a) Determine the average number of customers at each server.
 - (b) Determine the service completion rate for each server.
33. Explain how a Markov chain Monte Carlo simulation using the Gibbs sampler can be utilized to estimate
 - (a) the distribution of the amount of time spent at server j on a visit.

Hint: Use the arrival theorem.

- (b) the proportion of time a customer is with server j (i.e., either in server j 's queue or in service with j).
34. For open queueing networks
- state and prove the equivalent of the arrival theorem;
 - derive an expression for the average amount of time a customer spends waiting in queues.
35. Customers arrive at a single-server station in accordance with a Poisson process having rate λ . Each customer has a value. The successive values of customers are independent and come from a uniform distribution on $(0, 1)$. The service time of a customer having value x is a random variable with mean $3 + 4x$ and variance 5.
- What is the average time a customer spends in the system?
 - What is the average time a customer having value x spends in the system?
- *36. Compare the $M/G/1$ system for first-come, first-served queue discipline with one of last-come, first-served (for instance, in which units for service are taken from the top of a stack). Would you think that the queue size, waiting time, and busy-period distribution differ? What about their means? What if the queue discipline was always to choose at random among those waiting? Intuitively, which discipline would result in the smallest variance in the waiting time distribution?
37. In an $M/G/1$ queue,
- what proportion of departures leave behind 0 work?
 - what is the average work in the system as seen by a departure?
38. For the $M/G/1$ queue, let X_n denote the number in the system left behind by the n th departure.
- If

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n, & \text{if } X_n \geq 1 \\ Y_n, & \text{if } X_n = 0 \end{cases}$$

what does Y_n represent?

- (b) Rewrite the preceding as

$$X_{n+1} = X_n - 1 + Y_n + \delta_n \quad (8.64)$$

where

$$\delta_n = \begin{cases} 1, & \text{if } X_n = 0 \\ 0, & \text{if } X_n \geq 1 \end{cases}$$

Take expectations and let $n \rightarrow \infty$ in Equation (8.64) to obtain

$$E[\delta_\infty] = 1 - \lambda E[S]$$

- (c) Square both sides of Equation (8.64), take expectations, and then let $n \rightarrow \infty$ to obtain

$$E[X_\infty] = \frac{\lambda^2 E[S^2]}{2(1 - \lambda E[S])} + \lambda E[S]$$

- (d) Argue that $E[X_\infty]$, the average number as seen by a departure, is equal to L .

- *39. Consider an $M/G/1$ system in which the first customer in a busy period has the service distribution G_1 and all others have distribution G_2 . Let C denote the number of customers in a busy period, and let S denote the service time of a customer chosen at random.

Argue that

- (a) $a_0 = P_0 = 1 - \lambda E[S]$.
- (b) $E[S] = a_0 E[S_1] + (1 - a_0) E[S_2]$ where S_i has distribution G_i .
- (c) Use (a) and (b) to show that $E[B]$, the expected length of a busy period, is given by

$$E[B] = \frac{E[S_1]}{1 - \lambda E[S_2]}$$

- (d) Find $E[C]$.

40. Consider a $M/G/1$ system with $\lambda E[S] < 1$.

- (a) Suppose that service is about to begin at a moment when there are n customers in the system.
 - (i) Argue that the additional time until there are only $n - 1$ customers in the system has the same distribution as a busy period.
 - (ii) What is the expected additional time until the system is empty?
- (b) Suppose that the work in the system at some moment is A . We are interested in the expected additional time until the system is empty—call it $E[T]$. Let N denote the number of arrivals during the first A units of time.
 - (i) Compute $E[T|N]$.
 - (ii) Compute $E[T]$.

41. Carloads of customers arrive at a single-server station in accordance with a Poisson process with rate 4 per hour. The service times are exponentially distributed with rate 20 per hour. If each carload contains either 1, 2, or 3 customers with respective probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, compute the average customer delay in queue.

42. In the two-class priority queueing model of Section 8.6.2, what is W_Q ? Show that W_Q is less than it would be under FIFO if $E[S_1] < E[S_2]$ and greater than under FIFO if $E[S_1] > E[S_2]$.

43. In a two-class priority queueing model suppose that a cost of C_i per unit time is incurred for each type i customer that waits in queue, $i = 1, 2$. Show that type 1 customers should be given priority over type 2 (as opposed to the reverse) if

$$\frac{E[S_1]}{C_1} < \frac{E[S_2]}{C_2}$$

44. Consider the priority queueing model of Section 8.6.2 but now suppose that if a type 2 customer is being served when a type 1 arrives then the type 2 customer is bumped out of service. This is called the preemptive case. Suppose that when a bumped type 2 customer goes back in service his service begins at the point where it left off when he was bumped.

- (a) Argue that the work in the system at any time is the same as in the non-preemptive case.
- (b) Derive W_Q^1 .

Hint: How do type 2 customers affect type 1s?

- (c) Why is it not true that

$$V_Q^2 = \lambda_2 E[S_2] W_Q^2$$

- (d) Argue that the work seen by a type 2 arrival is the same as in the nonpreemptive case, and so

$$W_Q^2 = W_Q^2(\text{nonpreemptive}) + E[\text{extra time}]$$

where the extra time is due to the fact that he may be bumped.

- (e) Let N denote the number of times a type 2 customer is bumped. Why is

$$E[\text{extra time}|N] = \frac{NE[S_1]}{1 - \lambda_1 E[S_1]}$$

Hint: When a type 2 is bumped, relate the time until he gets back in service to a “busy period.”

- (f) Let S_2 denote the service time of a type 2. What is $E[N|S_2]$?
(g) Combine the preceding to obtain

$$W_Q^2 = W_Q^2(\text{nonpreemptive}) + \frac{\lambda_1 E[S_1] E[S_2]}{1 - \lambda_1 E[S_1]}$$

- *45. Calculate explicitly (not in terms of limiting probabilities) the average time a customer spends in the system in Exercise 24.
- 46. In the $G/M/1$ model if G is exponential with rate λ show that $\beta = \lambda/\mu$.
- 47. Verify Erlang's loss formula, Equation (8.60), when $k = 1$.
- 48. Verify the formula given for the P_i of the $M/M/k$.
- 49. In the Erlang loss system suppose the Poisson arrival rate is $\lambda = 2$, and suppose there are three servers, each of whom has a service distribution that is uniformly distributed over $(0, 2)$. What proportion of potential customers is lost?
- 50. In the $M/M/k$ system,
 - (a) what is the probability that a customer will have to wait in queue?
 - (b) determine L and W .
- 51. Verify the formula for the distribution of W_Q^* given for the $G/M/k$ model.
- *52. Consider a system where the interarrival times have an arbitrary distribution F , and there is a single server whose service distribution is G . Let D_n denote the amount of time the n th customer spends waiting in queue. Interpret S_n, T_n so that

$$D_{n+1} = \begin{cases} D_n + S_n - T_n, & \text{if } D_n + S_n - T_n \geq 0 \\ 0, & \text{if } D_n + S_n - T_n < 0 \end{cases}$$

53. Consider a model in which the interarrival times have an arbitrary distribution F , and there are k servers each having service distribution G . What condition on F and G do you think would be necessary for there to exist limiting probabilities?

References

- [1] J. Cohen, "The Single Server Queue," North-Holland, Amsterdam, 1969.
- [2] R. B. Cooper, "Introduction to Queueing Theory," Second Edition, Macmillan, New York, 1984.
- [3] D. R. Cox and W. L. Smith, "Queues," Wiley, New York, 1961.
- [4] F. Kelly, "Reversibility and Stochastic Networks," Wiley, New York, 1979.
- [5] L. Kleinrock, "Queueing Systems," Vol. I, Wiley, New York, 1975.
- [6] S. Nozaki and S. Ross, "Approximations in Finite Capacity Multiserver Queues with Poisson Arrivals," *J. Appl. Prob.* **13**, 826–834 (1978).
- [7] L. Takacs, "Introduction to the Theory of Queues," Oxford University Press, London and New York, 1962.
- [8] H. Tijms, "Stochastic Models: An Algorithmic Approach," Wiley, New York, 1994.
- [9] P. Whittle, "Systems in Stochastic Equilibrium," Wiley, New York, 1986.
- [10] Wolff, "Stochastic Modeling and the Theory of Queues," Prentice Hall, New Jersey, 1989.

so

$$E[T] = \frac{1 - e^{-\lambda\mu}}{\lambda e^{-\lambda\mu}}$$

42. (a) $F_e(x) = \frac{1}{\mu} \int_0^x e^{-y/\mu} dy = 1 - e^{-x/\mu}.$

(b) $F_e(x) = \frac{1}{c} \int_0^x dy = \frac{x}{c}, \quad 0 \leq x \leq c.$

- (c) You will receive a ticket if, starting when you park, an official appears within one hour. From Example 7.23 the time until the official appears has the distribution F_e which, by part (a), is the uniform distribution on $(0, 2)$. Thus, the probability is equal to $\frac{1}{2}$.

49. Think of each interarrival time as consisting of n independent phases—each of which is exponentially distributed with rate λ —and consider the semi-Markov process whose state at any time is the phase of the present interarrival time. Hence, this semi-Markov process goes from state 1 to 2 to 3 ... to n to 1, and so on. Also the time spent in each state has the same distribution. Thus, clearly the limiting probability of this semi-Markov chain is $P_i = 1/n$, $i = 1, \dots, n$. To compute $\lim P\{Y(t) < x\}$, we condition on the phase at time t and note that if it is $n - i + 1$, which will be the case with probability $1/n$, then the time until a renewal occurs will be sum of i exponential phases, which will thus have a gamma distribution with parameters i and λ .

Chapter 8

2. This problem can be modeled by an $M/M/1$ queue in which $\lambda = 6$, $\mu = 8$. The average cost rate will be

\$10 per hour per machine \times average number of broken machines

The average number of broken machines is just L , which can be computed from Equation (3.2):

$$\begin{aligned} L &= \frac{\lambda}{\mu - \lambda} \\ &= \frac{6}{2} = 3 \end{aligned}$$

Hence, the average cost rate = \$30/hour.

7. To compute W for the $M/M/2$, set up balance equations as follows:

$$\lambda P_0 = \mu P_1 \quad (\text{each server has rate } \mu)$$

$$(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2$$

$$(\lambda + 2\mu)P_n = \lambda P_{n-1} + 2\mu P_{n+1}, \quad n \geq 2$$

These have solutions $P_n = \rho^n / 2^{n-1} P_0$ where $\rho = \lambda/\mu$. The boundary condition $\sum_{n=0}^{\infty} P_n = 1$ implies

$$P_0 = \frac{1 - \rho/2}{1 + \rho/2} = \frac{(2 - \rho)}{(2 + \rho)}$$

Now we have P_n , so we can compute L , and hence W from $L = \lambda W$:

$$\begin{aligned} L &= \sum_{n=0}^{\infty} n P_n = \rho P_0 \sum_{n=0}^{\infty} n \left(\frac{\rho}{2}\right)^{n-1} \\ &= 2P_0 \sum_{n=0}^{\infty} n \left(\frac{\rho}{2}\right)^n \\ &= 2 \frac{(2 - \rho)}{(2 + \rho)} \frac{(\rho/2)}{(1 - \rho/2)^2} \quad (\text{See derivation of Equation (8.7).}) \\ &= \frac{4\rho}{(2 + \rho)(2 - \rho)} \\ &= \frac{4\mu\lambda}{(2\mu + \lambda)(2\mu - \lambda)} \end{aligned}$$

From $L = \lambda W$ we have

$$W = W(M/M/2) = \frac{4\mu}{(2\mu + \lambda)(2\mu - \lambda)}$$

The $M/M/1$ queue with service rate 2μ has

$$W(M/M/1) = \frac{1}{2\mu - \lambda}$$

from Equation (8.8). We assume that in the $M/M/1$ queue, $2\mu > \lambda$ so that the queue is stable. But then $4\mu > 2\mu + \lambda$, or $4\mu/(2\mu + \lambda) > 1$, which implies $W(M/M/2) > W(M/M/1)$. The intuitive explanation is that if one finds the queue empty in the $M/M/2$ case, it would do no good to have two servers. One would be better off with one faster server. Now let $W_Q^1 = W_Q(M/M/1)$ and $W_Q^2 = W_Q(M/M/2)$. Then,

$$W_Q^1 = W(M/M/1) - 1/2\mu$$

$$W_Q^2 = W(M/M/2) - 1/\mu$$

So,

$$W_Q^1 = \frac{\lambda}{2\mu(2\mu - \lambda)} \quad \text{from Equation (8.8)}$$

and

$$W_Q^2 = \frac{\lambda^2}{\mu(2\mu - \lambda)(2\mu + \lambda)}$$

Then,

$$W_Q^1 > W_Q^2 \Leftrightarrow \frac{1}{2} > \frac{\lambda}{(2\mu + \lambda)} \\ \lambda < 2\mu$$

Since we assume $\lambda < 2\mu$ for stability in the $M/M/1$ case, $W_Q^2 < W_Q^1$ whenever this comparison is possible, that is, whenever $\lambda < 2\mu$.

13. (a) $\lambda P_0 = \mu P_1$
 $(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2$
 $(\lambda + 2\mu)P_n = \lambda P_{n-1} + 2\mu P_{n+1}, \quad n \geq 2$

These are the same balance equations as for the $M/M/2$ queue and have solution

$$P_0 = \left(\frac{2\mu - \lambda}{2\mu + \lambda} \right), \quad P_n = \frac{\lambda^n}{2^{n-1}\mu^n} P_0$$

- (b) The system goes from 0 to 1 at rate

$$\lambda P_0 = \frac{\lambda(2\mu - \lambda)}{(2\mu + \lambda)}$$

The system goes from 2 to 1 at rate

$$2\mu P_2 = \frac{\lambda^2}{\mu} \frac{(2\mu - \lambda)}{(2\mu + \lambda)}$$

- (c) Introduce a new state cl to indicate that the stock clerk is checking by himself. The balance equation for P_{cl} is

$$(\lambda + \mu)P_{cl} = \mu P_2$$

Hence,

$$P_{cl} = \frac{\mu}{\lambda + \mu} P_2 = \frac{\lambda^2}{2\mu(\lambda + \mu)} \frac{(2\mu - \lambda)}{(2\mu + \lambda)}$$

Finally, the proportion of time the stock clerk is checking is

$$P_{cl} + \sum_{n=2}^{\infty} P_n = P_{cl} + \frac{2\lambda^2}{\mu(2\mu - \lambda)}$$

21. (a) $\lambda_1 P_{10}$.
(b) $\lambda_2(P_0 + P_{10})$.
(c) $\lambda_1 P_{10}/[\lambda_1 P_{10} + \lambda_2(P_0 + P_{10})]$.

- (d) This is equal to the fraction of server 2's customers that are type 1 multiplied by the proportion of time server 2 is busy. (This is true since the amount of time server 2 spends with a customer does not depend on which type of customer it is.) By (c) the answer is thus

$$\frac{(P_{01} + P_{11})\lambda_1 P_{10}}{\lambda_1 P_{10} + \lambda_2(P_0 + P_{10})}$$

24. The states are now $n, n \geq 0$, and $n', n' \geq 1$ where the state is n when there are n in the system and no breakdown, and n' when there are n in the system and a breakdown is in progress. The balance equations are

$$\lambda P_0 = \mu P_1$$

$$(\lambda + \mu + \alpha)P_n = \lambda P_{n-1} + \mu P_{n+1} + \beta P_{n'}, \quad n \geq 1$$

$$(\beta + \lambda)P_{1'} = \alpha P_1$$

$$(\beta + \lambda)P_{n'} = \alpha P_n + \lambda P_{(n-1)'}, \quad n \geq 2$$

$$\sum_{n=0}^{\infty} P_n + \sum_{n=1}^{\infty} P_{n'} = 1$$

In terms of the solution to the preceding,

$$L = \sum_{n=1}^{\infty} n(P_n + P_{n'})$$

and so

$$W = \frac{L}{\lambda_a} = \frac{L}{\lambda}$$

28. If a customer leaves the system busy, the time until the next departure is the time of a service. If a customer leaves the system empty, the time until the next departure is the time until an arrival *plus* the time of a service.

Using moment generating functions we get

$$\begin{aligned} E\{e^{sD}\} &= \frac{\lambda}{\mu} E\{e^{sD} \mid \text{system left busy}\} \\ &\quad + \left(1 - \frac{\lambda}{\mu}\right) E\{e^{sD} \mid \text{system left empty}\} \\ &= \left(\frac{\lambda}{\mu}\right) \left(\frac{\mu}{\mu-s}\right) + \left(1 - \frac{\lambda}{\mu}\right) E\{e^{s(X+Y)}\} \end{aligned}$$

where X has the distribution of interarrival times, Y has the distribution of service times, and X and Y are independent. Then

$$\begin{aligned} E\{e^{s(X+Y)}\} &= E\{e^{sX} e^{sY}\} \\ &= E\{e^{sX}\} E\{e^{sY}\} \quad \text{by independence} \\ &= \left(\frac{\lambda}{\lambda-s}\right) \left(\frac{\mu}{\mu-s}\right) \end{aligned}$$

So,

$$\begin{aligned} E\{e^{sD}\} &= \left(\frac{\lambda}{\mu}\right)\left(\frac{\mu}{\mu-s}\right) + \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\lambda-s}\right)\left(\frac{\mu}{\mu-s}\right) \\ &= \frac{\lambda}{(\lambda-s)} \end{aligned}$$

By the uniqueness of generating functions, it follows that D has an exponential distribution with parameter λ .

36. The distributions of the queue size and busy period are the same for all three disciplines; that of the waiting time is different. However, the means are identical. This can be seen by using $W = L/\lambda$, since L is the same for all. The smallest variance in the waiting time occurs under first-come, first-served and the largest under last-come, first-served.
39. (a) $a_0 = P_0$ due to Poisson arrivals. Assuming that each customer pays 1 per unit time while in service the cost identity of Equation (8.1) states that

$$\text{average number in service} = \lambda E[S]$$

or

$$1 - P_0 = \lambda E[S]$$

- (b) Since a_0 is the proportion of arrivals that have service distribution G_1 and $1 - a_0$ the proportion having service distribution G_2 , the result follows.
- (c) We have

$$P_0 = \frac{E[I]}{E[I] + E[B]}$$

and $E[I] = 1/\lambda$ and thus,

$$\begin{aligned} E[B] &= \frac{1 - P_0}{\lambda P_0} \\ &= \frac{E[S]}{1 - \lambda E[S]} \end{aligned}$$

Now from parts (a) and (b) we have

$$E[S] = (1 - \lambda E[S])E[S_1] + \lambda E[S]E[S_2]$$

or

$$E[S] = \frac{E[S_1]}{1 + \lambda E[S_1] + \lambda E[S_2]}$$

Substituting into $E[B] = E[S]/(1 - \lambda E[S])$ now yields the result.

- (d) $a_0 = 1/E[C]$, implying that

$$E[C] = \frac{E[S_1] + 1/\lambda - E[S_2]}{1/\lambda - E[S_2]}$$

45. By regarding any breakdowns that occur during a service as being part of that service, we see that this is an $M/G/1$ model. We need to calculate the first two moments of a service time. Now the time of a service is the time T until something happens (either a service completion or a breakdown) plus any additional time A . Thus,

$$\begin{aligned} E[S] &= E[T + A] \\ &= E[T] + E[A] \end{aligned}$$

To compute $E[A]$, we condition upon whether the happening is a service or a breakdown. This gives

$$\begin{aligned} E[A] &= E[A \mid \text{service}] \frac{\mu}{\mu + \alpha} + E[A \mid \text{breakdown}] \frac{\alpha}{\mu + \alpha} \\ &= E[A \mid \text{breakdown}] \frac{\alpha}{\mu + \alpha} \\ &= \left(\frac{1}{\beta} + E[S] \right) \frac{\alpha}{\mu + \alpha} \end{aligned}$$

Since $E[T] = 1/(\alpha + \mu)$ we obtain

$$E[S] = \frac{1}{\alpha + \mu} + \left(\frac{1}{\beta} + E[S] \right) \frac{\alpha}{\mu + \alpha}$$

or

$$E[S] = \frac{1}{\mu} + \frac{\alpha}{\mu\beta}$$

We also need $E[S^2]$, which is obtained as follows:

$$\begin{aligned} E[S^2] &= E[(T + A)^2] \\ &= E[T^2] + 2E[AT] + E[A^2] \\ &= E[T^2] + 2E[A]E[T] + E[A^2] \end{aligned}$$

The independence of A and T follows because the time of the first happening is independent of whether the happening was a service or a breakdown. Now,

$$\begin{aligned} E[A^2] &= E[A^2 \mid \text{breakdown}] \frac{\alpha}{\mu + \alpha} \\ &= \frac{\alpha}{\mu + \alpha} E[(\text{downtime} + S^*)^2] \\ &= \frac{\alpha}{\mu + \alpha} \{ E[\text{down}^2] + 2E[\text{down}]E[S] + E[S^2] \} \\ &= \frac{\alpha}{\mu + \alpha} \left\{ \frac{2}{\beta^2} + \frac{2}{\beta} \left[\frac{1}{\mu} + \frac{\alpha}{\mu\beta} \right] + E[S^2] \right\} \end{aligned}$$

Hence,

$$\begin{aligned} E[S^2] &= \frac{2}{(\mu + \beta)^2} + 2 \left[\frac{\alpha}{\beta(\mu + \alpha)} + \frac{\alpha}{\mu + \alpha} \left(\frac{1}{\mu} + \frac{\alpha}{\mu\beta} \right) \right] \\ &\quad + \frac{\alpha}{\mu + \alpha} \left\{ \frac{2}{\beta^2} + \frac{2}{\beta} \left[\frac{1}{\mu} + \frac{\alpha}{\mu\beta} \right] + E[S^2] \right\} \end{aligned}$$

Now solve for $E[S^2]$. The desired answer is

$$W_Q = \frac{\lambda E[S^2]}{2(1 - \lambda E[S])}$$

In the preceding, S^* is the additional service needed after the breakdown is over and S^* has the same distribution as S . The preceding also uses the fact that the expected square of an exponential is twice the square of its mean.

Another way of calculating the moments of S is to use the representation

$$S = \sum_{i=1}^N (T_i + B_i) + T_{N+1}$$

where N is the number of breakdowns while a customer is in service, T_i is the time starting when service commences for the i th time until a happening occurs, and B_i is the length of the i th breakdown. We now use the fact that, given N , all of the random variables in the representation are independent exponentials with the T_i having rate $\mu + \alpha$ and the B_i having rate β . This yields

$$\begin{aligned} E[S|N] &= \frac{N+1}{\mu+\alpha} + \frac{N}{\beta}, \\ \text{Var}(S|N) &= \frac{N+1}{(\mu+\alpha)^2} + \frac{N}{\beta^2} \end{aligned}$$

Therefore, since $1+N$ is geometric with mean $(\mu+\alpha)/\mu$ (and variance $\alpha(\alpha+\mu)/\mu^2$) we obtain

$$E[S] = \frac{1}{\mu} + \frac{\alpha}{\mu\beta}$$

and, using the conditional variance formula,

$$\text{Var}(S) = \left[\frac{1}{\mu+\alpha} + \frac{1}{\beta} \right]^2 \frac{\alpha(\alpha+\mu)}{\mu^2} + \frac{1}{\mu(\mu+\alpha)} + \frac{\alpha}{\mu\beta^2}$$

52. S_n is the service time of the n th customer; T_n is the time between the arrival of the n th and $(n+1)$ st customer.

Chapter 9

4. (a) $\phi(x) = x_1 \max(x_2, x_3, x_4)x_5$.
 (b) $\phi(x) = x_1 \max(x_2x_4, x_3x_5)x_6$.
 (c) $\phi(x) = \max(x_1, x_2x_3)x_4$.
6. A minimal cut set has to contain at least one component of each minimal path set. There are six minimal cut sets: $\{1, 5\}, \{1, 6\}, \{2, 5\}, \{2, 3, 6\}, \{3, 4, 6\}, \{4, 5\}$.

References

- ADAN, I., AND RESING, J. Queueing Theory .
<http://www.win.tue.nl/~iadan/queueing.pdf>, 2015.
- ALLEN, A. O. *Probability, statistics, and queueing theory with computer science applications*, 2nd ed. Academic Press, Inc., Boston, MA, 1990.
- BHAT, U. N. *An introduction to queueing theory: modeling and analysis in applications*. Birkhäuser, 2015.
- DAIGLE, J. *Queueing theory with applications to packet telecommunication*. Springer, New York, 2005.
- GROSS, D., SHORTLE, J., THOMPSON, J., AND HARRIS, C. *Fundamentals of queueing theory, 4th edition*. John Wiley & Sons, New York, 2008. .
- LEON-GARCIA, A. *Probability, statistics, and random processes for electrical engineering*. Pearson Education, 2017.
- JAIN, R. *The art of computer systems performance analysis*. Wiley & Sons, New York, 1991.
- KULKARNI, V. *Modeling, analysis, design, and control of stochastic systems*. Springer, New York, 1999.
- PALANIAMMAL, S. *Probability and Queueing Theory*. PHI Learning Pvt. Ltd., 2011.
https://content.kopykitab.com/ebooks/2016/06/7500/sample/sample_7500.pdf.
- ROSS, S. M. *Introduction to Probability Models*. Academic Press, Boston, 1989.