CHAPTER 4
Examples of Queueing Networks

In this chapter various examples of queueing networks will be described to illustrate the uses of earlier results and to indicate their limitations.

4.1 COMMUNICATION NETWORKS

An example of a communication network is a telegraph system such as that illustrated in Fig. 4.1(a), where a graph represents the system with vertices and directed edges corresponding to cities and directed channels respectively. Messages are generated in a city and require to be transmitted, possibly via intermediate relay cities, to their destination city. The transmission of message from a city cannot begin until the entire message has been received at that city. Each channel has a maximum capacity and hence the various messages interact. We can model this situation as a network of queues by regarding each message as a customer and each channel as a queue (see Fig. 4.1b). We shall suppose that messages arrive from outside the system (each with its route through the channels of the system) in independent Poisson streams. It is not obvious how the progress of a message through a channel can be represented by a customer passing through a queue, and we shall discuss two possible models.

The time taken for a message to pass along a channel depends on various factors, including the length of the message and random effects associated with the channel. In our first model we shall suppose that the time a message takes to pass along channel j is exponentially distributed with mean ϕ_j^{-1} and independent of the time it takes to pass along other channels along its route. We shall call ϕ_j the capacity of channel j. This model might be appropriate if the random effects associated with a channel are predominant; for example it will arise if the channels are noisy and a message has to be repeatedly transmitted over a channel until it is received without error. With this first model each channel behaves as a single-server queue with exponential service times. The most likely queue discipline is first come first served, but a possible alternative is service in random order. With either of these disciplines the system will behave as a network of queues of the sort discussed in Section 3.1. Thus if n_j is the number of messages waiting at channel j, including the message being transmitted, then in equilibrium

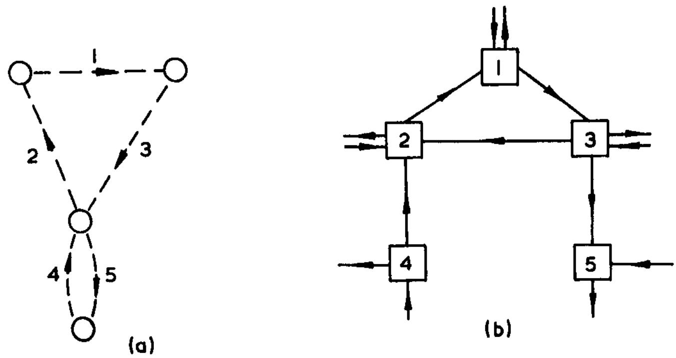

Fig. 4.1 (a) A telegraph system and (b) its representation as a network of queues

$n_1, n_2, \ldots, n_J$ are independent and

$$P(n_j = n) = \left(1 - \frac{a_j}{\phi_j}\right)\left(\frac{a_j}{\phi_j}\right)^n \tag{4.1}$$

where a_j is the average arrival rate at queue j.

In our second model we shall represent a message by a customer whose service requirement is the *same* at every queue along his route, is arbitrarily distributed with unit mean, and is independent from customer to customer. This model might be appropriate if the length of a message is important in determining the time taken to transmit it over a channel. If each channel can be modelled by a symmetric queue then the system will behave as an open network of quasi-reversible queues, with the type of a customer indicating his service requirement as well as his route through the network. The condition that a channel be modelled by a symmetric queue is a severe limitation but allows two queue disciplines which might be appropriate in the context. In the first the channel capacity is divided equally between all the messages waiting for transmission at the channel, corresponding to the server-sharing queue discussed in Section 3.3; this discipline might occur in computer networks and results in short messages being transmitted fairly quickly, even through a congested channel. In the second the channel capacity is devoted entirely to the last arriving message, corresponding to the stack discussed in Section 3.3; this discipline might be a reasonable approximation in a communication network in which the last message sent takes priority over earlier messages. If channel j can supply service effort at rate ϕ_j then in equilibrium $n_1, n_2, \ldots, n_J$ are independent and n_j again has the distribution (4.1).

In practice neither of the models described above might precisely represent the system under scrutiny, but the fact that the same distribution (4.1) emerges under a variety of different assumptions suggests that it might be a good first approximation. We shall now investigate some of its consequences.

The distribution (4.1) implies that the mean number of customers waiting at queue j is $a_j/(\phi_j - a_j)$ and that the average waiting time of a customer at queue j is $1/(\phi_j - a_j)$. Suppose now that we can choose the channel capacities $\phi_1, \phi_2, \ldots, \phi_J$ subject to an overall cost constraint

$$\sum_j f_j \phi_j = F \tag{4.2}$$

How should we allocate the resource F between the competing channels in order to minimize the average time spent in the network by a customer or, equivalently, the mean number of customers in the network? To answer this question we proceed just as in Section 2.4.

Theorem 4.1. *The optimal allocation is*

$$\phi_j = a_j + \frac{\sqrt{a_j f_j}}{\sum_k \sqrt{a_k f_k}} \frac{F - \sum_k a_k f_k}{f_j}$$

Proof. The mean number of customers in the network is

$$\sum_j \frac{a_j}{\phi_j - a_j}$$

and our task is to minimize this subject to the constraint (4.2). Introduce the Lagrange multiplier y and let

$$L = \sum_j \frac{a_j}{\phi_j - a_j} + y\left(\sum_j f_j \phi_j - F\right)$$

Setting $\partial L/\partial \phi_j = 0$ we find that L is minimized by the choice

$$\phi_j = a_j + \sqrt{\frac{a_j}{y f_j}}$$

Substitution of this in constraint (4.2) shows that we should choose

$$\frac{1}{\sqrt{y}} = \frac{F - \sum_k a_k f_k}{\sum_k \sqrt{a_k f_k}}$$

which establishes the result.

There are various other situations where the model discussed in this section and the optimal allocation obtained in Theorem 4.1 might prove helpful. For example the model might be appropriate for a manufacturing job-shop, with customers representing items of work which require to be processed by a number of machines, or a road traffic network, with queues representing bottlenecks in the system.

Exercises 4.1

1. Two cities are connected by two directed channels each of capacity ϕ (Fig. 4.2a). Each channel carries messages which are initiated at rate a. It is proposed that the two channels be replaced by a single channel capable of carrying messages in either direction (Fig. 4.2b). Show that the mean waiting time of a message will be decreased if the capacity of the new channel is greater than $\phi + a$.
2. The two communication networks illustrated in Fig. 4.3 are being considered to link three cities. In the first each of the six channels has capacity ϕ, while in the second each of the three channels has capacity 2ϕ. It is anticipated that the rate at which messages will be sent from one city to another city is a. Show that the mean time a message spends in the system will be less for the first network if $\phi < 3a$.
3. Suppose that the service requirements of a customer at successive queues are not identical but are random variables which may depend upon each other and upon the route of the customer. Show that if the queues are symmetric then distribution (4.1) remains valid, with a_j being the average amount of service requirement arriving at queue j per unit time.
4. Let the routes through the network be labelled and suppose that each unit of time a customer on route i remains in the system costs $g(i)$. Show that the optimal allocation is now

$$\phi_j = a_j + \frac{\sqrt{b_j f_j}}{\sum_k \sqrt{b_k f_k}} \frac{F - \sum_k a_k f_k}{f_j}$$

where $b_j = \sum_i g(i) a_j(i)$ and $a_j(i)$ is the average arrival rate at queue j of customers on route i.
5. If the queues are symmetric and if the service requirement of a customer on route i may depend upon his type, as in Exercise 4.1.3, show that the optimal allocation remains as in Exercise 4.1.4, but with $a_j(i)$ interpreted as the average rate at which service requirement for customers on route i arrives at queue j.

Fig. 4.2 Alternative communication networks

Fig. 4.3 Alternative communication networks

6. Suppose that service effort is supplied at queue j at rate

$$\phi_j(n) = \frac{n}{n+r-1}\phi_j$$

where r is a positive constant. The case $r=1$ corresponds to the model discussed in this section. Show that if $\phi_1, \phi_2, \ldots, \phi_J$ can be chosen subject to the constraint (4.2) then the optimal allocation is that given in Theorem 4.1.

7. Suppose that service effort is supplied at queue j at rate

$$\phi_j(n) = n\phi_j$$

This might be appropriate if any number of messages can be transmitted at the same time by a channel. Show that if $\phi_1, \phi_2, \ldots, \phi_J$ can be chosen subject to the constraint (4.2) then the optimal allocation is

$$\phi_j = \frac{\sqrt{a_j f_j}}{\sum_k \sqrt{a_k f_k}} \frac{F}{f_j}$$

Show that if $\phi_1, \phi_2, \ldots, \phi_J$ can be chosen subject to the constraint

$$\sum_j \log \phi_j = F$$

then in the optimal allocation $\phi_j/\phi_k = a_j/a_k$.

8. The model of a communication network discussed in this section is, of course, not the only available model, and other models may lead to quite different conclusions. For example suppose arrivals at the series of queues illustrated in Fig. 2.2 form a Poisson process, and the service requirements of customers at queues are all equal to the same fixed value. The model of this section deals with the case where the queues are symmetric, but is inadequate if each queue is a first come first served single-server queue. Show that in this case there will never be more than one customer in queue j, $j=2, 3, \ldots, J$.

4.2 MACHINE INTERFERENCE

The basic form of the machine interference problem is as follows. There are N machines under the care of a single operative. From time to time a machine stops and requires the attention of the operative before it can resume running. The operative can only attend one machine at a time, and so if two or more machines are stopped the others must wait for attention. Thus the machines interfere with one another, and a matter of interest is the extent of this interference.

Suppose initially that the N machines are identical, that the running time of a machine before it stops is exponentially distributed with mean R and that the service time a machine requires from the operative before it can resume running is exponentially distributed with mean S. All running times and service times are assumed independent. The machines that are stopped queue to receive the attention of the operative and the two most common disciplines are for the operative to attend to them in the order of their stopping or in a random order. We can regard the machines that are running as forming a queue also, a queue with an infinite number of servers where a machine remains until it next stops. Observe that service requirements in this queue are in fact machine running times.

The system can thus be represented by the closed queueing network shown in Fig. 4.4. Both queues are quasi-reversible: they would behave in isolation as an $M/M/1$ queue and an $M/M/\infty$ queue respectively. If n_0 is the number of stopped machines and $n_1 = N - n_0$ the number of running machines then in equilibrium

$$\pi(n_0, n_1) = BS^{n_0}\frac{R^{n_1}}{n_1!} \tag{4.3}$$

and from this distribution it is possible to calculate quantities of interest such as the proportion of time the operative or a machine is busy. The above system is in fact a simple example of a closed migration process, and indeed the equilibrium distribution could easily have been determined from the observation that n_1 is a birth and death process. The reason for considering the system as a closed network of quasi-reversible queues is that viewed in this light it becomes readily apparent which of the assumptions underlying the model are crucial in determining the distribution (4.3) and which are unnecessary.

The assumption that running times are exponentially distributed is clearly unnecessary. If running times have a general distribution then the queue containing running machines is equivalent to one which would in isolation

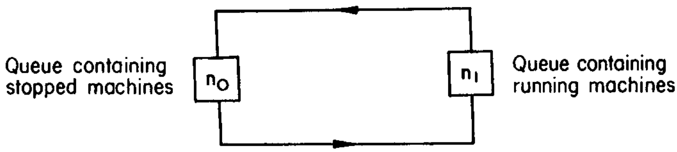

Fig. 4.4 The basic machine interference model

behave as an $M/G/\infty$ queue. Thus the queue remains quasi-reversible and expression (4.3) remains the equilibrium distribution, with R the mean of the running time distribution. The assumption that service times are exponentially distributed cannot be so easily relaxed since the queue containing stopped machines has just one server; with the queue disciplines of interest a single-server will not be quasi-reversible unless service times are exponentially distributed.

In terms of the closed queueing networks of Section 3.4 the present model has just one type of customer whose route consists of two stages with $r(1,1)=0$ and $r(1,2)=1$. We shall now consider the effect on the model of allowing longer routes. Suppose, for example, that the route consists of four stages, with $r(1,1)=0$, $r(1,2)=1$, $r(1,3)=0$, and $r(1,4)=1$. Since queue 1 is a symmetric queue we can allow the service requirement there to have a general distribution and to depend upon whether the customer has reached stage 2 or stage 4 of his route. Let R' and R'' be the mean service requirement at stage 2 and stage 4 respectively. Thus the mean running time of a machine alternates between R' and R'', changing after each service. Since queue 0 is not a symmetric queue we must retain the condition that service times be exponentially distributed, and they must not depend upon the stage reached. Expression (4.3) remains valid, with R defined as the overall mean running time $(R'+R'')/2$ (Exercise 3.4.4). Indeed the result will remain true even if the service requirements at stages 2 and 4 are dependent: in terms of Section 3.4 this corresponds to a customer being allocated a random type each time he begins his route. By fully exploiting the types and routes of the last chapter it is possible to allow a machine's running time to depend upon any number of its previous running times; expression (4.3) will remain valid with R defined as the overall mean running time.

Until now we have supposed that the N machines are identical. When the machines differ it is helpful to view the system as the network shown in Fig. 4.5. Machine i $(i=1,2,\ldots,N)$ remains in queue i while it is running and moves to queue 0 when it stops. After it has been serviced it returns to

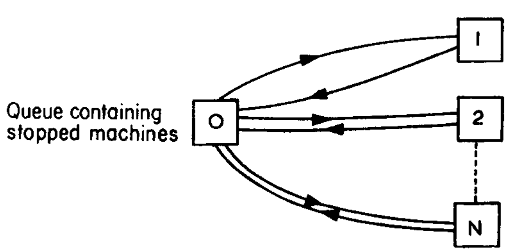

Fig. 4.5 Machine interference with differing machines

queue i. Assume that service times are exponentially distributed with mean S. Initially we shall also assume that the running time of machine i is exponentially distributed with mean R_i, with all running times and service times independent. The state of the system can be written as $(\mathbf{x}_0, x_1, x_2, \ldots, x_N)$ where $\mathbf{x}_0 = (t(1), t(2), \ldots, t(n_0))$ is a listing in order of the machines waiting for service, and $x_j = 1$ or 0 depending on whether machine j is running or not. From Theorem 3.12 we can deduce that the equilibrium distribution is

$$\pi(\mathbf{x}_0, x_1, x_2, \ldots, x_J) = B'S^{n_0} \prod_{i=1}^{N} R_i^{x_i} \tag{4.4}$$

Thus, given that machines $t(1), t(2), \ldots, t(n_0)$ are stopped, each possible ordering of them within queue 0 is equally likely. To obtain the probability that, say, machines $1, 2, \ldots, n$ are stopped and machines $n+1, n+2, \ldots, N$ are running we need to sum expression (4.4) over the $n!$ different orderings of machines $1, 2, \ldots, n$, giving

$$B'n!S^n \prod_{i=n+1}^{N} R_i \tag{4.5}$$

Observe that queue i $(i = 1, 2, \ldots, N)$ is quasi-reversible even when service requirements at this queue are not exponentially distributed: queue i can be considered to be an example of any one of the four symmetric queues described in Section 3.3, since at most one customer is ever present in it. Thus we can generalize the model to allow a machine's running time to have an arbitrary distribution and to depend upon its previous running times. The state of the system will become more complicated since it will need to record more information about each machine. Nevertheless, if the system is in equilibrium the probability that machines $1, 2, \ldots, n$ are stopped and the others running will still be given by expression (4.5), with R_i the overall mean running time of machine i (Exercise 3.4.2). If the overall mean running times of the machines are equal, so that

$$R_1 = R_2 = \cdots = R_N = R$$

then to obtain the probability that n machines are stopped we need to multiply expression (4.5) by the number of different ways n machines can be chosen from N, giving

$$B' \frac{N!}{(N-n)!} S^n R^{N-n} \tag{4.6}$$

This is consistent with expression (4.3); the normalizing constants bear the relation $B = B'N!$ If the overall mean running times are not equal then it might be hoped that distribution (4.6) would still hold with R taken as an average of $R_1, R_2, \ldots, R_N$. In fact this is not true except in the approximate sense explored in Exercise 4.2.3.

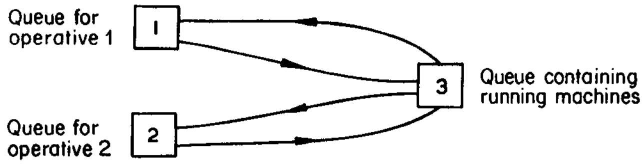

Fig. 4.6 Machine interference with two forms of stoppage

In our final example from the area of machine interference we shall suppose there are N identical machines and that there are two possible reasons for a machine stopping. We shall suppose there are two operatives, one to deal with each form of stoppage. Figure 4.6 illustrates the system; queues 1 and 2 contain machines awaiting service from operatives 1 and 2 respectively and queue 3 contains running machines. Assume that service times at queues 1 and 2 are exponentially distributed with means S_1 and S_2 respectively and that running times are arbitrarily distributed with mean R. Initially we shall also assume that all running times and service times are independent and that when a machine stops it requires attention from operative 1 with probability p_1 and from operative 2 with probability p_2 $(=1-p_1)$ independently of its past history. From Theorem 3.12 it is a simple matter to deduce that in equilibrium

$$\pi(n_1, n_2, n_3) = B(p_1S_1)^{n_1}(p_2S_2)^{n_2}\frac{R^{n_3}}{n_3!} \tag{4.7}$$

As might be expected the above assumptions can be considerably relaxed. If a machine alternately visits operatives 1 and 2 then expression (4.7) remains valid with $p_1 = p_2 = \frac{1}{2}$. If a machine's running time and its reason for stopping at the end of that running time are dependent then we let R be the overall mean running time of a machine. Indeed we can even allow a machine's running time and its reason for stopping to depend upon previous running times and previous reasons for stopping for that machine. Expression (4.7) remains valid with R the overall mean running time of a machine and p_1 the overall proportion of stoppages that require operative 1.

In all of the examples discussed in this section it has been necessary to assume that service times are exponentially distributed and independent of each other and of running times, since the queue for a server is not a symmetric queue. In the next section we shall consider a model closely related to the model of this section, but where it is reasonable to suppose that all the queues involved are symmetric.

Exercises 4.2

1. Deduce from expression (4.3) that n_1 has the same distribution as the number of calls connected in the telephone exchange model of Section 1.3.

2. Consider the basic machine interference model governed by Fig. 4.4 and expression (4.3). Suppose now that the machines are arranged in a priority order and that the operative always works on the machine with the highest priority of those stopped, even if this involves interrupting the service of another machine. Show that the probability that the machine which is kth in the priority order is running is

$$\left[k-(k-1)\frac{f_{k-2}}{f_{k-1}}\right]\frac{1}{f_k}$$

where f_k is given by the recursion

$$f_k = 1+\frac{Sk}{R}f_{k-1}$$

with $f_0 = 1$.

3. Consider the machine interference model with differing machines governed by Fig. 4.5, in which the mean running times of the machines are $R_1, R_2, \ldots, R_N$. Show that if the mean service time S is small a good approximation to the expected number of machines stopped and to the probability that no machines are stopped can be obtained from the basic machine interference model if in that model the mean running time of each machine, R, is given by

$$\frac{1}{R}=\frac{1}{N}\left(\frac{1}{R_1}+\frac{1}{R_2}+\cdots+\frac{1}{R_N}\right)$$

If S is large show that a good approximation to the expected number of machines stopped can be obtained with

$$R=\frac{1}{N}(R_1+R_2+\cdots+R_N)$$

and to the probability that no machines are stopped with

$$R=(R_1R_2\cdots R_N)^{1/N}$$

(A good approximation when S is small or large is one which is accurate to within $o(S)$ or $o(1/S)$ respectively.)

4. Consider the machine interference model illustrated in Fig. 4.5. Suppose now that while the number of stopped machines is n the remaining $N-n$ machines work at a reduced rate $\psi(n)$, and their remaining running times decrease at rate $\psi(n)$ rather than unity. Use Exercise 3.5.8 to find the equilibrium distribution for the system.

5. Consider the machine model with differing machines illustrated in Fig. 4.5. Suppose now that whenever the number of stopped machines reaches M the remaining $N-M$ machines pause, i.e. they cease running

until the operative has finished serving one of the stopped machines and they then resume running where they left off. Deduce from the previous exercise that the equilibrium distribution for the system still takes the form (4.4), but over a smaller state space. Observe that in the case $M=1$ queue 0 is a symmetric queue and so the equilibrium distribution will be the same if the service time of a machine is arbitrarily distributed and dependent on that machine's earlier service and running times provided its overall mean is still S. This system could be viewed as a model of a complex device comprising N units which stops functioning if any one of the units fails.

6. Outline how the machine interference models described can be generalized to allow more than one or two operatives.

4.3 TIMESHARING COMPUTERS

Figure 4.7 illustrates how a queueing network may arise as a much simplified model of a timesharing computer. Queue 0 represents the central processing unit of the computer and queues $1, 2, \ldots, N$ represent terminals. The N customers in the queueing network correspond to jobs, and job i $(i = 1, 2, \ldots, N)$ is either being dealt with by the central processing unit or is with the computer user at terminal i. The model as described so far is equivalent to the machine interference model illustrated in Fig. 4.5. The models diverge when we consider the appropriate queue discipline for queue 0. For the present application the most natural assumption is that queue 0 is a single-server queue operating with the server-sharing discipline described in Section 3.3. Queue 0 will then be a symmetric queue and we can allow service requirements at this queue as well as at queues $1, 2, \ldots, N$ to be arbitrarily distributed. Let S_i and R_i be the mean service requirement of customer i at queue 0 and at queue i respectively. Thus S_i and R_i could be

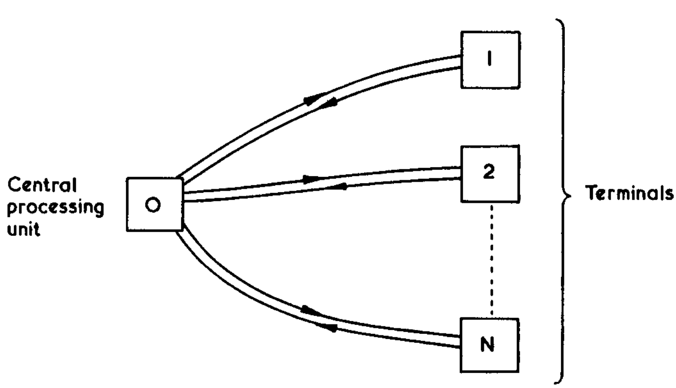

Fig. 4.7 A timesharing computer

called the mean processor requirement and the mean think time respectively for job i. The equilibrium probability that, say, jobs $1, 2, \ldots, n$ are with the central processing unit and jobs $n+1$, $n+2, \ldots, N$ are at terminals is

$$Bn!S_1S_2 \cdots S_nR_{n+1}R_{n+2} \cdots R_N$$

and is insensitive to the form of the distributions involved. This is true even if the service requirements of customer i at queue 0 or i depend upon his earlier service requirements at either or both queues. Note, however, that service requirements of different customers must be independent. The above equilibrium probability allows us to calculate quantities of interest such as the probability that the central processing unit is idle or the proportion of his time a user spends waiting for his job to return to the terminal.

We shall now consider an extension of the model which allows a user, with his job, to leave the terminal. Suppose there is a finite source population of potential users who may wish to use the computer. For simplicity assume these users are identical. Let n_2 be the number of users not using the computer, let n_1 be the number of users at terminals whose jobs are awaiting a response from them, and let n_0 be the number of jobs being dealt with by the central processing unit (Fig. 4.8). Thus if there are in total N users and M terminals then

$$n_0+n_1+n_2=N$$

and (4.8)

$$n_0+n_1 \leq M$$

If one of the users not using the computer attempts to find a terminal and they are all occupied, that is $n_0+n_1=M$, then he tries again later. If he finds a free terminal then he occupies it for a period while his job oscillates between the terminal and the central processing unit. During this period let the total think time have mean R and let the total processor requirement have mean S. Let the time between a user leaving a terminal and next attempting to find one have mean T.

Regarding the system as a network of queues we see that queue 2, containing potential users, behaves as an infinite-server queue at which service requirements are arbitrarily distributed with mean T. Note that there is a capacity constraint on queues 0 and 1: if a customer leaves queue 2 to

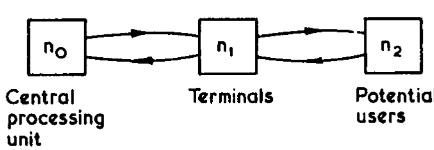

Fig. 4.8 A timesharing computer and its users

find M customers already in queues 0 and 1 he immediately returns to queue 2. Apart from this capacity constraint queue 1 behaves as an infinite-server queue and queue 0 as a server-sharing queue. Hence even with the capacity constraint queues 0 and 1 considered together behave as a quasi-reversible system (Exercise 3.5.5). Thus we can deduce that the equilibrium distribution is

$$\pi(n_0, n_1, n_2) = B_{M,N} S^{n_0} \frac{R^{n_1}}{n_1!} \frac{T^{n_2}}{n_2!} \tag{4.9}$$

over triples (n_0, n_1, n_2) satisfying conditions (4.8). Observe that during the period a user is at a terminal the precise pattern of the oscillations of his job between the terminal and the central processing unit do not affect the result; the distribution (4.9) depends only on the mean quantities R and S. Although it is possible to allow dependencies between the service requirements of a customer at different queues it is not possible to allow the time a user remains away from the terminals to depend upon whether or not he was successful the last time he attempted to find a terminal. This is because a customer entering queue 2 carries with him an indication of his past service requirements but no indication beyond this of his past experience; essentially a customer leaving queue 2 to find M customers already in queues 0 or 1 has his service requirements at these queues met, but instantaneously.

The probability that a user attempting to find a terminal is successful can be determined using part (iii) of Theorem 3.12. It is just the equilibrium probability that queues 0 and 1 contain M customers when there are only $N-1$ customers in the system altogether, and is hence

$$B_{M,N-1} \frac{T^{N-M-1}}{(N-M-1)!} \sum_{n=0}^{M} S^{M-n} \frac{R^n}{n!}$$

Exercises 4.3

1. Extend the model just described to the case where users are not identical.
2. If in the model just described $N, T \to \infty$ with N/T held fixed at ν, check that the equilibrium distribution for (n_0, n_1) becomes

$$\pi(n_0, n_1) = B(\nu S)^{n_0} \frac{(\nu R)^{n_1}}{n_1!}$$

over pairs (n_0, n_1) satisfying $n_0 + n_1 \leq M$. Show directly that this is the equilibrium distribution for an open system in which the points in time at which users attempt to find a terminal form a Poisson process.
3. Deduce from Exercise 3.4.6 that for the model illustrated in Fig. 4.7 the mean time a given user spends waiting before his job returns from the central processing unit is proportional to its processor requirement.

Observe that for the model illustrated in Fig. 4.8 the restriction to a given user can be dropped; since the users are identical the mean time taken by any job to pass through the central processing unit is proportional to its processor requirement.

4. In the model illustrated in Fig. 4.8 it was assumed that a customer leaving queue 2 to find M customers in queues 0 and 1 immediately returns to queue 2. Suppose the model is amended in the following way: when the number of customers present in queue 2 drops to $N-M$ the service effort provided at queue 2 becomes zero. Show that the equilibrium distribution is unaltered. Observe that if the time a user spends away from a terminal is exponentially distributed and independent of his experience elsewhere then the two models are equivalent. Unless this assumption is a reasonable one to make, it is unlikely that either model adequately represents the real response of users when all terminals are in use.

4.4 TELETRAFFIC MODELS

We have already discussed models of a telephone exchange in Sections 1.3, 2.1, and 3.3. We shall begin this section by showing how these models can be viewed as special cases of the machine interference models described in Section 4.2. We shall then discuss some further more complicated teletraffic models.

In the simple telephone exchange model of Section 1.3 calls are initiated as a Poisson process, the exchange has K lines, a call initiated when all the lines are busy is lost, and a connected call lasts for an exponentially distributed length of time. This corresponds to the basic machine interference model illustrated in Fig. 4.4 with n_1 the number of busy lines and n_0 the number of idle lines, where $n_0+n_1=K$. Note how the assumptions that calls are initiated as a Poisson process and that calls are lost when all lines are busy correspond to the assumption of exponential service times in the machine interference model. The more general telephone exchange model of Section 3.3 in which call lengths are arbitrarily distributed corresponds to the machine interference model in which running times are arbitrarily distributed. This relationship between telephone exchange models and machine interference models points the way to various generalizations, one of which we will discuss now (others will be considered in Exercises 4.4). Suppose that while a line is busy there is a possibility that it may develop a fault. If this happens the call in progress is allowed to finish, but immediately afterwards the line undergoes repair. We can represent the system by Fig. 4.9 where n_0 is the number of idle lines, n_1 the number of busy lines, and n_2 the number of lines undergoing repair. Suppose that calls are initiated as a Poisson process of rate ν, are independent of each other, and are arbitrarily

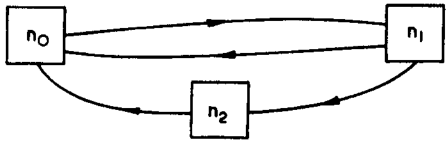

Fig. 4.9 A model of a telephone exchange with unreliable lines

distributed with mean μ^{-1}. Allow the probability intensity that a line develops a fault and the subsequent repair time to depend upon the total time for which the line has been busy and the number of calls it has dealt with since its last repair. Let the mean repair time be λ^{-1} and let m be the mean number of calls a line can handle between successive repairs. Regarding the system as a network of queues we see that queue 0 behaves as a single-server queue at which service requirements are exponentially distributed and queues 1 and 2 behave as infinite-server queues. Observe how the dependence allowed between call lengths, fault occurrence, and repair times is modelled by a dependence between the route and the service requirements at the two symmetric queues of a given customer. In equilibrium

$$\pi(n_1, n_2) = B_K \frac{1}{n_1!}\left(\frac{\nu}{\mu}\right)^{n_1} \frac{1}{n_2!}\left(\frac{\nu}{\lambda m}\right)^{n_2} \qquad 0 \le n_1 + n_2 \le K$$

We shall now consider a teletraffic model which makes simple use of the more general arrival rates discussed in Section 3.5. Suppose there are J exchanges $A_1, A_2, \ldots, A_J$ connected to exchange C via a transit exchange B (Fig. 4.10). Let there be R_j lines between A_j and B and K lines between B and C where $R_j \le K$ and $K < R_1 + R_2 + \cdots + R_J$. Suppose that calls requiring a line between A_j and C are initiated as a Poisson process of rate ν_j and that such calls are lost when all the lines from A_j to B or all the lines from B to C are busy. If a call between A_j and C is connected suppose that the call lasts for an arbitrarily distributed length of time, with mean μ_j^{-1}. Figure 4.10 includes a representation of the system as a network of queues. If n_j is the number of calls in progress between A_j and C for $j = 1, 2, \ldots, J$ then the

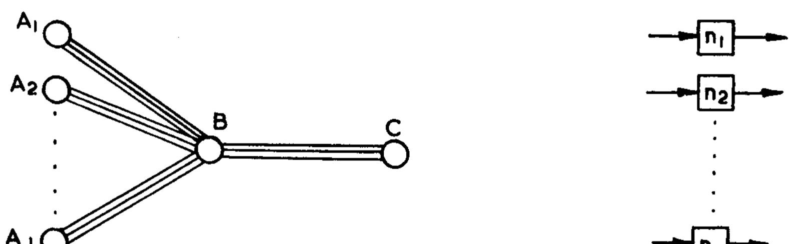

Fig. 4.10 Merging of teletraffic: the exchanges and their representation as a network of queues

probability intensity that a call is connected between A_j and C can be written as

$$\nu_j \frac{\Psi(n_1, n_2, \ldots, n_j+1, \ldots, n_J)}{\Psi(n_1, n_2, \ldots, n_j, \ldots, n_J)}$$

where

$$\Psi(n_1, n_2, \ldots, n_J) = \begin{cases} 1 & \text{if } n_j \le R_j,\ j=1,2,\ldots,J;\ n_1+n_2+\cdots+n_J \le K \\ 0 & \text{otherwise} \end{cases}$$

Thus Theorem 3.14 shows that in equilibrium

$$\pi(n_1, n_2, \ldots, n_J) = B \prod_{j=1}^{J} \left(\frac{\nu_j}{\mu_j}\right)^{n_j} \frac{1}{n_j!}$$

$$n_j \le R_j,\ j=1,2,\ldots,J;\ n_1+n_2+\cdots+n_J \le K$$

Note that in this application of Theorem 3.14 customers of class j visit only queue j.

A more demanding use of Theorem 3.14 will be needed for the following system. A call distributor consists of R_1 switches connected to a first group of K_1 lines and R_2 switches connected to a second group of K_2 lines (see Fig. 4.11), where $R_j \ge K_j$, for $j=1, 2$. Calls are initiated as a Poisson process of rate ν. When a call is initiated it is allocated an idle switch at random, so that if n_j calls are in progress on group j then the probability that the call is routed to group j is $(R_j-n_j)/(R_1+R_2-n_1-n_2)$. A call routed to group j that finds all K_j lines busy is lost. Connected calls last for a time which is arbitrarily distributed with mean μ^{-1}. The probability intensity a call is connected on group j is $\nu(R_j-n_j)/(R_1+R_2-n_1-n_2)$ is $n_j<K_j$ and is zero otherwise. This can be written in the form appropriate for an application of Theorem 3.14, viz. form (3.27), with

$$\Psi(n_1, n_2) = \begin{cases} \dfrac{(R_1+R_2-n_1-n_2)!}{(R_1-n_1)!(R_2-n_2)!} & n_1 \le K_1,\ n_2 \le K_2 \\ 0 & \text{otherwise} \end{cases}$$

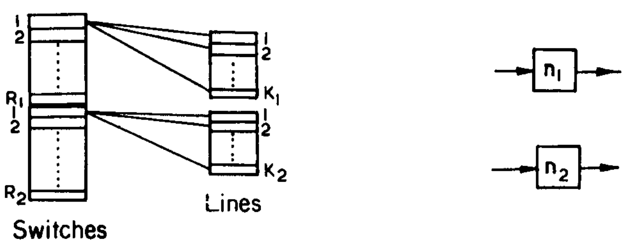

Fig. 4.11 A switching system and its representation as a network of queues

Thus in equilibrium

$$\pi(n_1, n_2) = B\frac{(R_1+R_2-n_1-n_2)!}{(R_1-n_1)!(R_2-n_2)!}\left(\frac{\nu}{\mu}\right)^{n_1+n_2}\frac{1}{n_1!n_2!} \qquad n_1 \le K_1,\ n_2 \le K_2$$

Exercises 4.4

1. Consider a telephone exchange with K lines. Suppose the mean call length on line j is μ_j^{-1}. Show that if the rate at which calls are initiated, ν, is small then a good approximation to the probability a call is lost is given by Erlang's formula

$$\frac{(1/K!)(\nu/\mu)^K}{\sum_{j=0}^{K}(1/j!)(\nu/\mu)^j} \tag{4.10}$$

with $\mu = (\mu_1\mu_2\cdots\mu_K)^{1/K}$, while if ν is large a good approximation is given by the same expression with $\mu = (1/K)(\mu_1+\mu_2+\cdots+\mu_K)$.

2. A switchboard has K lines and one operator. Calls arrive at the switchboard as a Poisson process of rate ν, but calls arriving while all K lines are in use are lost. A call finding a free line has to wait for the operator to answer. The operator deals with waiting calls one at a time and takes an exponentially distributed amount of time with mean λ^{-1} to connect a call to the correct extension, after which the call lasts for an arbitrarily distributed length of time with mean μ^{-1}. Show that the probability k lines are busy is proportional to

$$\left(\frac{\nu}{\lambda}\right)^k \sum_{n=0}^{k}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \qquad k = 0, 1, \ldots, K$$

3. Consider the model of a telephone exchange with unreliable lines. Show that the probability a call is lost is the same as in the simple telephone exchange model of Section 1.3 if the mean call length in that model is $\mu^{-1}+(\lambda m)^{-1}$.

4. Consider the model of a telephone exchange with unreliable lines. How will the equilibrium distribution be affected if when a fault occurs the call in progress is lost?

5. Consider the model of a telephone exchange with unreliable lines. Suppose that lines with faults must be repaired one at a time and that the time taken to repair a line is exponentially distributed. Show that the probability k lines are busy takes the same form as in Exercise 4.4.2 and discuss the relationship between the two models.

6. Extend the model of the merging of teletraffic to allow calls to be made between exchanges B and C and between exchanges A_j and B. Extend the model to allow more than one transit exchange. Deal with the case where there may be more than one possible route for a call and make the assumption (familiar from the repair shop model of Section 3.5) that

a call is connected if calls already in progress can be reshuffled to make room for it.

7. Extend the model of a switching system to allow J groups of lines. Show that the probability a call is lost is less than in a system in which an arriving call is allocated to group j with probability $R_j/\sum_{i=1}^{J} R_i$. Show that the probability a call is lost is greater than expression (4.10) with $K=\sum_{j=1}^{J} K_j$, unless $R_j=K_j$, $j=1,2,\ldots,J$.
8. The models of this section have assumed that calls are initiated as a Poisson process. Explain how the telephone exchange model considered in Exercise 4.4.2 can be amended to deal with a finite source population of potential callers, as described in Exercise 1.3.5, by allowing idle lines to form a queue with $\phi(n)=N-K+n$ rather then $\phi(n)=1$, for $n>0$, where N is the total size of the source population.
9. In this section we have viewed lines as customers. If there is a finite source population it may be more useful to regard the callers as customers. For example consider a telephone exchange with K lines serving a population of N distinguishable callers. Let μ_i^{-1} be the mean length of a call from caller i, $i=1,2,\ldots,N$, and let λ_i^{-1} be the mean time between the end of his last call (whether connected or lost) and his next attempt to call. Show that in equilibrium the probability callers $i_1, i_2, \ldots, i_n$ are connected is

$$B\prod_{j=1}^{n}\frac{\lambda_{i_j}}{\mu_{i_j}}.$$

Note that in this model intercall times and call lengths can be arbitrarily distributed, but whether a caller is connected or lost must have no effect on his future behaviour.

10. Although in this section we have viewed lines as customers the resulting systems are sometimes quasi-reversible with calls viewed as customers. Consider, for example, the basic telephone exchange model analogous to the machine interference model illustrated in Fig. 4.4. In this model the K lines are regarded as customers and so we can allow the successive call lengths on a given line to be dependent. Suppose now that we view the whole system as a single queue at which calls are the customers. Show that, counting lost calls, this queue is quasi-reversible provided the class of an arriving call does not affect its progress through the queue. Suppose now that this queue and an infinite-server queue form a closed network containing N callers. Since the infinite-server queue is symmetric we can allow the intercall times of a given caller to be dependent. Let λ_i^{-1} be the mean intercall time for caller i, $i=1,2,\ldots,N$, and let μ_k^{-1} be the mean call length for line k, $k=1,2,\ldots,K$. Show that in equilibrium the probability callers $i_1, i_2, \ldots, i_n$

are connected and lines $k_1, k_2, \ldots, k_n$ are busy is

$$B\prod_{i=1}^{n}\frac{\lambda_{i_i}}{\mu_{k_i}}$$

Contrast the patterns of dependence which can be allowed in this model with those that can be allowed in the model of Exercise 4.4.9.

11. Show that in the models of the preceding two exercises the probability a particular caller finds n lines busy when he attempts to make a call is equal to the equilibrium probability that n lines would be busy were the system to contain just the other $N-1$ callers.

4.5 COMPARTMENTAL MODELS

In this section we shall consider systems which have the property that after customers (or particles or individuals) have entered the system they move independently through it. Viewing the system as a network of queues, each queue would in isolation behave as an $M/G/\infty$ queue. The analysis of such systems in equilibrium is fairly straightforward, and there are various applications. We shall mention compartmental models in biology, birth–illness–death processes, and models of manpower systems. Finally, we shall show that in a special case the transient behaviour of the system can be completely described.

In biology compartmental models are used to represent the movement of particles through the various parts of an animal's body. The system is assumed to consist of J compartments with particles entering the system in a Poisson stream and independently moving around between the various compartments before leaving the system. Let ν be the arrival rate of particles at the system and let α_j be the mean time a particle passing through the system spends in compartment j. Then it is a simple consequence of Theorem 3.7 that in equilibrium the number of particles in compartment j is independent of the number of particles in the other compartments and has a Poisson distribution with mean $\nu\alpha_j$. This is true no matter how complicated the motion of the individual particles. For example a particle's stay in a particular compartment may be arbitrarily distributed and may depend upon its past history, as may the compartment the particle chooses to visit next. The essential assumption is that the movements of different particles are independent.

Birth–illness–death processes have the same structure as the above model but the interpretation is rather different. The idea is that an individual is born and passes through various states of health before eventually dying. Given their times of birth the individuals move independently through the J states of the system. The life history of an individual (the states he will pass through and the length of time he will remain in each) is chosen at birth from an arbitrary distribution over all such life histories. If there are N individuals alive let the probability intensity of a birth be $\nu(N)$. The previous model thus corresponds to the case $\nu(N)=\nu$ for all N. Let n_j be the number of individuals in state j and let α_j be the mean time an individual spends in state j throughout its lifetime. Theorem 3.14 shows that when an equilibrium distribution exists it is given by

$$\pi(n_1, n_2, \ldots, n_J) = B\left(\prod_{l=0}^{N-1} \nu(l)\right)\prod_{j=1}^{J} \frac{\alpha_j^{n_j}}{n_j!}$$

Obvious consequences of this are that the equilibrium distribution for N is

$$\pi(N) = B\frac{\alpha^N}{N!}\prod_{l=0}^{N-1} \nu(l)$$

where $\alpha = \sum_{j=1}^{J} \alpha_j$ is the average lifetime of an individual, and that given N the distribution of $(n_1, n_2, \ldots, n_J)$ is multinomial.

The above model could also be used to represent the flow of individuals through a manpower system. The various states would then correspond to grades within the organizational hierarchy. The assumption that individuals move independently of one another prevents the model from dealing with systems in which promotions occur to fill vacancies, rather than when an individual is ready (cf. Exercise 6.3.2). The rate of recruitment $\nu(N)$ will generally be a decreasing function of N, in contrast to the birth–illness–death process where the birth rate $\nu(N)$ will usually be an increasing function of N. As an example of the sort of result which might be useful in this particular application, suppose the distribution function for the total time an individual spends in grade j is $F(u)$. Then Theorems 3.10 and 3.14 show that in equilibrium the amount of experience in that grade which a typical individual there has already acquired has distribution function

$$F^*(x) = \frac{1}{\alpha_j}\int_0^x (1-F(u))\,du$$

Until now we have concerned ourselves with the equilibrium behaviour of compartmental models. If $\nu(N)=\nu$, so that the arriving stream of individuals is Poisson, it is possible to analyse the transient behaviour of the model. Let $p_j(t)$ be the probability that an individual is, a time t after his arrival, in compartment j.

Theorem 4.2. *If the arrival stream is Poisson and if the system is empty at time* 0 *then at time t the number of individuals in compartment j is independent of the number in the other compartments and has a Poisson distribution* *with mean $\nu\alpha_j(t)$ where*

$$\alpha_j(t) = \int_0^t p_j(u)\,du$$

Proof. Let M be the number of arrivals in the interval $(0, t)$. Conditional on M, the instants of arrival $t_1, t_2, \ldots, t_M$ are independent random variables uniformly distributed on $(0, t)$; this follows from the assumption that the arrival process is Poisson. The probability that the arrival at t_r is in compartment j at time t is $p_j(t-t_r)$. Because the individuals move independently we can deduce that

$$E(z_1^{n_1(t)}z_2^{n_2(t)}\cdots z_J^{n_J(t)} \mid M, t_1, t_2, \ldots, t_M) = \prod_{r=1}^{M}\left\{\sum_j [1-p_j(t-t_r)+z_jp_j(t-t_r)]\right\}$$
$$= \prod_{r=1}^{M}\left\{1-\sum_j (1-z_j)p_j(t-t_r)\right\}$$

Averaging this over $t_1, t_2, \ldots, t_M$, conditional on M,

$$E(z_1^{n_1(t)}z_2^{n_2(t)}\cdots z_J^{n_J(t)} \mid M) = \prod_{r=1}^{M}\left\{1-\sum_j (1-z_j)t^{-1}\int_0^t p_j(t-u)\,du\right\}$$
$$= \left\{1-\sum_j (1-z_j)t^{-1}\int_0^t p_j(t-u)\,du\right\}^M$$

Averaging over M, which has a Poisson distribution with mean νt,

$$E(z_1^{n_1(t)}z_2^{n_2(t)}\cdots z_J^{n_J(t)}) = \exp\left[-\nu\sum_j (1-z_j)\int_0^t p_j(t-u)\,du\right]$$
$$= \prod_{j=1}^{J} \exp[-(1-z_j)\nu\alpha_j(t)]$$

Hence $n_1(t), n_2(t), \ldots, n_J(t)$ are independent Poisson variables with means $\nu\alpha_j(t)$, which proves the result.

Letting $t \to \infty$ we obtain the previous equilibrium result, since $\int_0^\infty p_j(u)\,du$ is the mean time an individual spends in compartment j throughout its lifetime.

Exercises 4.5

1. In the birth, death, and immigration process considered in Section 1.3 the lifetimes of individuals were exponentially distributed with mean μ^{-1}. Show that the equilibrium distribution (1.14) remains the same if lifetimes are arbitrarily distributed with mean μ^{-1}. Look now at the family size process $(n_1, n_2, \ldots)$ introduced in Section 2.4. By considering

the mean time for which a family has j members alive throughout its entire existence show that the equilibrium distribution for the process $(n_1, n_2, \ldots)$ also remains the same if lifetimes are arbitrarily distributed with mean μ^{-1}. (An alternative proof of this result will be given in Exercise 7.1.9.)

2. In this section we have supposed that individuals are all of the same type. Theorem 3.14 shows that with more than one type of individual more general arrival rates can be allowed. We shall give two examples. Consider two birth–illness–death processes and let $N(i)$ be the number of individuals alive in process i, $i = 1, 2$. Find the equilibrium distribution if the birth rate in process i is altered to $(N(i)+1)/(N(1)+N(2)+2)$, $i = 1, 2$, and show that although the overall birth rate is constant the total number of individuals alive does not necessarily have a Poisson distribution. Find the equilibrium distribution if the birth rates in processes 1 and 2 are altered to $x^{N(2)}$ and $x^{N(1)}$ respectively $(x < 1)$ and show that conditional on the number of individuals alive in one process the number alive in the other process has a Poisson distribution.
3. Consider a birth–illness–death process with two states. Suppose individuals are born into state 1 where they remain for a mean time α_1 and then move to state 2 where they remain for a mean time α_2 before dying. Suppose the birth rate depends only on n_1, the number of individuals in state 1. Use Little's result to find an expression for the mean number of individuals in state 2. Show that it equals $\alpha_2/(1-\nu\alpha_1)$ if the birth rate is $\nu(n_1+1)$.

   Patients arrive at a hospital in a Poisson stream of rate ν, but the hospital redirects them if its K beds are all occupied. An accepted patient stays in the hospital for an average of α_1 days; after he leaves the hospital he attends an outpatients' department for an average of α_2 days. Find the mean number of patients attending the outpatients' department.
4. Cars arrive at the beginning of a long road in a Poisson stream of rate ν from time $t = 0$ onwards. A car has a fixed velocity $V > 0$ which is a random variable. The velocities of different cars are independent. Show that the number of cars on the first x miles of the road at time t has a Poisson distribution with mean $\nu E[V^{-1} \min\{x, Vt\}]$. What is the distribution of the number of cars between x and y miles along the road at time t?
5. Show that the conclusions contained in Theorem 4.2 are not valid if at time $t = 0$ there are already individuals within the system.
6. (Hard) A monkey attempts to climb a tree with a constant positive velocity, but at the points in time of a Poisson process the monkey suffers instantaneous negative displacements, the lengths of which are independent of each other and of the Poisson process, and have a common distribution. The expected net velocity of the monkey is positive. Let $n(s)$ be the number of times the monkey slips backwards past the point s

on the tree, so that $n(s)+1$ is the number of the times the monkey climbs past s in the forward direction. Show that the stochastic process $n(s)$ is identical to the number of individuals alive at time s in a birth, death, and immigration process (Exercise 4.5.1) with the immigration rate equal to the birth rate.

4.6 MISCELLANEOUS APPLICATIONS

Road traffic. Consider an infinitely long road on which there are two kinds of vehicle travelling in the same direction (Fig. 4.12). Some of the vehicles are lorries, travelling at a constant velocity u. The rest of the vehicles are cars which travel at a constant velocity $v(>u)$ unless they are held up behind lorries. Suppose that the cars behind a lorry overtake it one at a time, that the car immediately behind the lorry has to wait for an exponentially distributed period of time before it can overtake, and that these periods are independent with mean μ^{-1}. If we apply a velocity $-u$ to all the vehicles, so that the lorries are reduced to rest, then the lorries can be regarded as single-server queues and the cars as customers with exponential service times at these queues. The gaps between lorries can similarly be regarded as infinite-server queues. If we assume that the points in time at which cars catch up with a given lorry form a Poisson process, then the whole system will behave as an infinite series of quasi-reversible queues. At any given time the positions of the cars not held up behind lorries will form a Poisson process of rate λ_1, say. Let λ_2^{-1} be the mean distance between successive lorries. Each lorry behaves as an $M/M/1$ queue with arrival rate $\lambda_1(v-u)$ and service rate μ, so the mean queue size behind a lorry is

$$\frac{\lambda_1(v-u)}{\mu-\lambda_1(v-u)}$$

provided $\lambda_1(v-u)<\mu$. The average time taken for a car to pass a lorry is $[\mu-\lambda_1(v-u)]^{-1}$ and to catch up with the next lorry is $[\lambda_2(v-u)]^{-1}$. Thus the average speed of a car is

$$u+\frac{\lambda_2^{-1}}{[\mu-\lambda_1(v-u)]^{-1}+[\lambda_2(v-u)]^{-1}}=u+\frac{[\mu-\lambda_1(v-u)](v-u)}{\mu+(\lambda_2-\lambda_1)(v-u)} \tag{4.11}$$

Let ν denote the average rate at which cars pass a given point on the road; then

$$\nu=\lambda_1 v+\frac{\lambda_1\lambda_2(v-u)u}{\mu-\lambda_1(v-u)} \tag{4.12}$$

Fig. 4.12 Road traffic

the first term coming from cars passing the point singly and the second from cars passing the point in bunches behind lorries. Expressions (4.11) and (4.12) can be used to explore the effect of varying the parameters of the system. Suppose, for example, that ν varies, with u, v, μ, and λ_2 held constant, corresponding to an increase in the volume of car traffic. Expression (4.12) shows that as λ_1 increases from zero to $\mu/(v-u)$ the parameter ν increases from zero to infinity. Thus for a given value of ν there is a unique solution for λ_1, and using expression (4.11) it can be shown that as ν increases from zero to infinity the average speed of a car decreases from v to u. Other examples are given in Exercises 4.6.

Conveyor belt inspection. Consider a continuously moving conveyor belt carrying items past a quality control office (Fig. 4.13). The office contains K inspectors. When an item reaches the office it enters the office if any of the K inspectors are free—otherwise it continues along the belt. The time taken by an inspector to check an item is arbitrarily distributed with mean μ^{-1}, and after an item has been checked it is replaced on the belt. If items arrive at the office in a Poisson stream at rate ν then this model is equivalent to the queue with no waiting room of Section 3.3, and so the equilibrium probability that j inspectors are occupied is

$$\pi(j) = b_K \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} \qquad j = 0, 1, 2, \ldots, K$$

where b_K is the normalizing constant. Further, items pass a point on the conveyer belt downstream from the office in a Poisson stream.

Suppose now that the inspectors are not concentrated in one office but are spread along the conveyor belt as illustrated in Fig. 4.14. Suppose that an item reaching the kth inspector is picked up for checking by that inspector if he is free and if that item has not been checked already by an earlier inspector. It is apparent that in equilibrium the gaps between the inspectors will make no difference to whether or not a particular item is picked up by a given inspector. The stream of items reaching the kth inspector will be the same as if the first $k-1$ inspectors were positioned in a single office. Thus items pass any point along the conveyor belt in a Poisson stream. Despite this, the complete analysis of the system is difficult: for example the probability that the first and third inspectors are both busy will depend on more than just the first moment of the checking time. Exercise 4.6.4 obtains the probability that the kth inspector is busy.

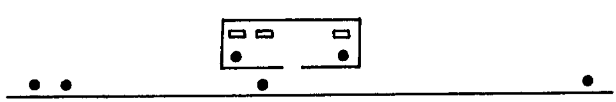

Fig. 4.13 A quality control office

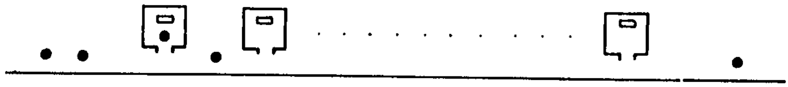

Fig. 4.14 *K* separate offices

If we suppose an item reaching the kth inspector is picked up if he is free and whether or not it has been checked by an earlier inspector, then the stream of items passing any point on the conveyor belt is again Poisson; the system is then a series of quasi-reversible queues.

Electronic counters. A source emits a stream of particles according to a Poisson process of rate ν. An electronic counter is exposed to the stream of particles, but not all the particles are registered. When a particle is registered it causes an aftereffect which lasts for a mean time α. If the counter is suffering from the aftereffects of n particles the probability that an arriving particle will be registered is $p(n)$. Thus n behaves as does the compartmental model of the previous section, and if the counter is in equilibrium the probability that it is suffering from the aftereffects of n particles is

$$\pi(n) = B\frac{(\nu\alpha)^n}{n!}\prod_{l=0}^{n-1} p(l) \tag{4.13}$$

where B is a normalizing constant. The long-run rate at which particles are registered is

$$\nu^* = \nu \sum_{n=0}^{\infty} \pi(n)p(n) \tag{4.14}$$

which can thus be calculated as a function of ν. It is in fact an increasing function of ν (Exercise 4.6.5), and hence from an observed rate ν^* the true rate ν can be determined. For example if $p(0)=1$ and $p(n)=0$, $n>0$ (a type I counter), then expressions (4.13) and (4.14) imply that

$$\nu = \frac{\nu^*}{1-\nu^*\alpha}$$

Some counters may accept a particle and suffer its aftereffect without registering it. Let $p(n)$ be the probability that a counter accepts an arriving particle when it is suffering from the aftereffects of n particles and let $r(n)$ be the probability that it registers it. Then expression (4.13) again gives the equilibrium probability that the counter is suffering from the aftereffects of n particles, but

$$\nu^* = \nu \sum_{n=0}^{\infty} \pi(n)r(n) \tag{4.15}$$

gives the long-run rate at which particles are registered. Expression (4.15) is not necessarily a monotonic function of ν and so an observed rate ν^* may not correspond to a unique value ν. For example if $r(0)=1$, $r(n)=0$, $n\geq 1$, and $p(n)=1$, $n\geq 0$ (a type II counter), then expressions (4.13) and (4.15) imply that ν is one of the two roots of the equation

$$\nu^* = \nu e^{-\nu\alpha}$$

A garage. A garage employs two mechanics and is fed by a Poisson stream of cars at rate ν. If when a car arrives both mechanics are free the car is equally likely to be assigned to either of the two mechanics. If one mechanic is free the car is assigned to him. If both mechanics are busy the car is lost. The time taken by mechanic i to repair a car has an arbitrary distribution with mean μ_i^{-1}, for $i=1, 2$.

If we view the cars as customers we can obtain the equilibrium distribution for the system from Theorem 3.14. Alternatively, we can view the mechanics as the customers in a closed queueing network. Either approach shows that the equilibrium probability that mechanic i is busy is

$$\frac{\nu(\nu+\mu_i)}{(\nu+\mu_1)(\nu+\mu_2)+\mu_1\mu_2} \tag{4.16}$$

and is insensitive to the form of the repair time distributions. The second approach additionally shows that if the times taken by mechanic i to repair cars form a dependent sequence then this probability is unaltered, with μ_i being the overall mean repair time for mechanic i.

Consider now the stream of cars leaving the garage, including lost cars. Augmenting the state of the network with a flip-flop variable to signal when cars are lost shows that this stream is Poisson and that if we now regard the cars as customers, all of the same class, then the system is quasi-reversible. If there are different classes of cars the system is still quasi-reversible provided the class of a car does not affect its progress through the garage. In this case the class of a car cannot affect its repair time, and hence if the garage is part of a network of quasi-reversible queues a car's repair time cannot depend upon its route or its service requirements at the symmetric queues in the network. In this respect the system is similar to the queues considered in Section 3.1 rather than a symmetric queue, even though its operation involves arbitrary distributions and its equilibrium distribution exhibits a form of insensitivity.

Is it possible to allow the class of a car to affect its progress through the garage? We shall now show that it is. Suppose there are J classes of car. Consider the closed queueing network illustrated in Fig. 4.15; the customers in this network are the two mechanics. The presence of a mechanic in queue 0 indicates that he is idle. The presence of mechanic 1 (respectively 2) in queue jA (respectively jB) indicates that he is repairing a car of class j,

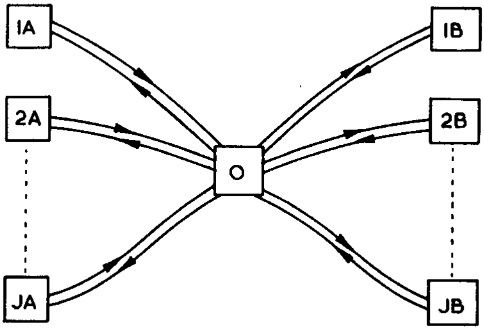

Fig. 4.15 Representation of a garage

$j = 1, 2, \ldots, J$. Cars of class j arrive at the garage in a Poisson stream at rate ν_j, $j = 1, 2, \ldots, J$, and these streams are independent: the implications of this for the queueing network illustrated in Fig. 4.15 are that queue 0 would in isolation behave as an $M/M/1$ queue and that when mechanic 1 (respectively 2) leaves queue 0 he goes to queue jA (respectively jB) with probability ν_j/ν, where $\nu = \sum \nu_j$, independently of the previous history of the system. This determines the routing behaviour of the mechanics. The assumption that if both mechanics are idle an arriving car is equally likely to be assigned to either of them is compatible with the queue disciplines allowed in Section 3.1, as are the alternative assumptions that the car is assigned to the mechanic who has been idle the longest, or the shortest, time. Note that which mechanic is assigned the car cannot depend on the car's class. Queues jA, jB, $j = 1, 2, \ldots, J$, are symmetric queues—they can each contain at most one customer. The service requirement of mechanic 1 at queue jA can depend upon j, upon his previous route, and upon his service requirements at queues 1A, 2A, . . . , JA. Thus the repair time of a car of class j assigned to mechanic 1 can depend upon the class of the car, upon the classes of the cars previously repaired by mechanic 1, and upon the repair times of these cars. The equilibrium probability that mechanic i is busy is given by expression (4.16) where μ_i is the overall mean repair time for mechanic i.

There is an important difference between the above process and the process obtained from it by time reversal. In the original process the probability that an arriving car is of class j is independent of the state of the process, but the time taken to repair the car may depend upon the state of the process, in particular upon the class of the car previously repaired by the mechanic assigned to the car. Thus in the reversed process the probability that on leaving queue 0 mechanic 1 goes to queue jA may depend upon the state of the process, in particular upon the time spent by mechanic 1 on his last excursion from queue 0. Hence the system will not in general be quasi-reversible with respect to the J classes of cars.

Let us now restrict the dependencies allowed in the queueing network illustrated in Fig. 4.15. Suppose that each time a car is assigned to mechanic i he is allocated a repair capability, where the successive values allocated to him may be dependent upon each other but not upon the classes of the cars. Let the time spent by a mechanic repairing a car (the repair time) have a distribution determined by the repair capability of the mechanic and the class of the car (we can imagine that a car of class j has a repair requirement whose distribution depends on j and that the repair time is a function of the car's repair requirement and the mechanic's repair capability). Queues jA, jB, $j=1,2,\ldots,J$, thus behave as symmetric queues at which the service requirement of a customer has a distribution determined by the corresponding repair capability of the mechanic. The system just described is a restricted form of the system previously discussed, and so the equilibrium probability that mechanic i is busy is still given by expression (4.16). Observe, though, that while the time taken to repair a car may depend upon previous repair times through the sequence of repair capabilities of a mechanic it cannot depend on the classes of the cars previously repaired by him. Hence the reversed process has the property that when mechanic 1 (respectively 2) leaves queue 0 the probability that he goes to queue jA (respectively jB) is ν_j/ν, independently of his past experience and hence of the state of the process. This is enough to show that the system, appropriately augmented to signal lost cars, is quasi-reversible with respect to the J classes of car.

If the garage is part of a network of quasi-reversible queues then the patterns of dependence which can emerge take an interesting form. The time taken by a mechanic to repair a car can be dependent upon the mechanic's experience elsewhere because it can depend on his repair capability, and it can be dependent on the car's experience elsewhere because it can depend on the car's class. However, the experience elsewhere of the car (respectively the mechanic) can depend on that particular repair time only through its class (respectively his repair capability).

Exercises 4.6

1. Consider the road traffic model discussed in this section. Allow ν to vary, with u, v, and μ held fixed and with λ_2 held equal to ν/ku; this corresponds to varying the overall volume of traffic with the ratio of cars to lorries held fixed at k to 1. Show that as ν increases from zero to infinity the average speed of a car decreases from v to u and the mean queue size behind a lorry increases from zero to k.
2. Suppose that in the road traffic model discussed in this section there are only finitely many lorries. In this case the volume of car traffic will be best measured by $\nu=\lambda_1 v$. Investigate the effect of varying the velocity

of the cars on T, the mean time taken by a car to pass the string of lorries, with u, μ, λ_2, and ν held fixed. Show that

(a) If $\mu > \nu$ then as v tends to infinity the mean overtaking time T tends to a finite limit.

(b) If $\mu < \nu$ then as v increases from u to $\nu u/(\nu - \mu)$ the mean overtaking time T decreases from infinity to a minimum and then increases back to infinity.

3. Show how the model of road traffic discussed in this section can be extended to allow cars to travel with different velocities, with cars overtaking each other freely.
4. Show that the probability the kth inspector in the sequence illustrated in Fig. 4.14 is busy is equal to the probability calculated in Exercise 4.2.2, with R/S replaced by λ/μ.
5. Observe that the mean of the distribution (4.13) is an increasing function of ν. Deduce that the rate of registrations ν^* given by expression (4.14) is an increasing function of the rate of arrivals ν.
6. Consider a variant of the type I counter in which $p(0) = p$, $p(n) = 1$, $n > 0$. Show that

$$\nu^* = \frac{\nu p}{p + (1-p)e^{-\nu\alpha}}$$

Consider a variant of the type II counter in which $r(n) = r^n$, $n \geq 0$. Show that

$$\nu^* = \nu e^{-\nu\alpha(1-r)}$$

7. For the garage represented in Fig. 4.15 show that if mechanic i is busy the probability he is repairing a car of class j is

$$\frac{\nu_j \mu_i}{\nu \mu_{ji}}$$

where μ_{ji}^{-1} is the overall mean repair time of a class j car with mechanic i. Show that all the results obtained in this section for a garage with two mechanics can be generalized to a garage with N mechanics.

8. Generalize the model described in Exercise 4.4.10 to allow a call length to depend upon the caller and the line, and discuss the extent to which it can depend upon the previous experience of each of them.
9. A garage employs two mechanics and is fed by two independent Poisson streams of cars. If when a car from the ith stream arrives at the garage the ith mechanic is free he repairs it; if he is busy and the other mechanic is free then the other mechanic repairs it; if both are busy the car is lost. The time taken to repair a car from the ith stream has an arbitrary distribution, for $i = 1, 2$. Use Erlang's formula and Little's result to find the probability the ith mechanic is busy. Observe that it is

insensitive to the form of the repair time distributions and depends only on their means. Observe that the stream of cars leaving the garage, counting lost cars, is Poisson.

10. Suppose the system of the previous exercise is represented by a Markov process $\mathbf{x}(t)$. Show that if from $\mathbf{x}(t_0)$ it is possible to deduce which mechanics are busy at time t_0 then $\mathbf{x}(t_0)$ is *not* independent of the departure process prior to time t_0. Thus if information about which mechanics are busy is included in the state then the system is not quasi-reversible.
11. A number of trucks and excavators are involved in an earthmoving operation. Trucks are loaded by the excavators at the site, after which they travel to a dump, unload the earth, and return to the site. Describe the various ways in which the operation could be modelled as a closed network of quasi-reversible queues.
12. A fleet of vessels operates between a number of loading and discharge ports. Describe the various ways in which the system could be modelled as a closed network of quasi-reversible queues.