CHAPTER 3

Queueing Networks

In the previous chapter some simple examples of queueing networks were introduced. This chapter will continue the discussion of queueing networks, but within a more general framework.

3.1 GENERAL CUSTOMER ROUTES

Consider the queueing network illustrated in Fig. 3.1. In this network there are five simple queues, and customers can enter the system at queues 1 or 2, arrivals at these queues forming two independent Poisson processes. Customers follow the route through queues 1, 3, and 4 or the route through queues 2, 3, and 5 before leaving the system. This might be a model of a manufacturing job-shop with customers representing items of work which require to be processed at a sequence of machines. This network *cannot* be represented by a migration process. The difficulty is that a customer leaving queue 3 does not choose at random between queues 4 and 5: he moves to queue 4 if he has previously been to queue 1. In a migration process the past route of a customer in a given queue is of no use in predicting his future route, and in this sense the customers in a queue are homogeneous. In this section we shall see that by dividing customers into different types we can deal with networks such as the one illustrated in Fig. 3.1.

Suppose that there are I different customer types and that a customer's type determines his route through the J queues of the system. More specifically, suppose that customers of type i $(i = 1, 2, \ldots, I)$ enter the system in a Poisson stream at rate $\nu(i)$ and pass through the sequence of queues

$$r(i, 1), r(i, 2), \ldots, r(i, S(i))$$

before leaving the system. Thus the queue which a customer of type i visits at stage s $(s = 1, 2, \ldots, S(i))$ of his route is queue $r(i, s)$. Note that the route of a customer may require him to visit the same queue more than once. For simplicity we shall not allow two successive stages of a customer's route to be identical. We shall assume that the I Poisson arrival streams are independent. It is not essential that I be finite, but we shall require that $\sum \nu(i)$ be finite.

By using more than one customer type we can represent the behaviour of a customer whose future route depends stochastically upon his past route: we simply use a different type for each possible route. Consider, for

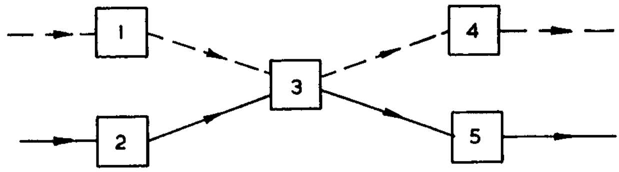

Fig. 3.1 A job-shop model

example, the network illustrated in Fig. 3.2. Suppose this differs from the network of Fig. 3.1 in that customers who have been through queue 2 are, after leaving queue 3, equally likely to move to queue 4 or queue 5. We require three customer types to model this network: customers of types 1, 2, and 3 follow the routes $1 \to 3 \to 4$, $2 \to 3 \to 4$, and $2 \to 3 \to 5$ respectively, and the arrival rates $\nu(2)$ and $\nu(3)$ are equal.

The above method can deal with the random routes which arise in an open migration process, but it will be more cumbersome than the approach of the previous chapter if the migration process allows a customer to visit the same queue more than once (Exercise 3.1.2). The advantage of the above method is that it allows much more general routing schemes than can arise in a migration process. To give two further examples, it can deal with a system in which a customer visits each queue exactly once, but in a random order, or a system in which each customer visits a certain queue exactly twice.

We have described how customers move between queues: we must now describe how the queues themselves operate. This is rather more complicated than it was for a migration process, since within each queue we must now keep track of the different types of customer. We shall suppose that the customers in each queue are ordered: thus queue j ($j = 1, 2, \ldots, J$) will contain customers in positions $1, 2, \ldots, n_j$, where n_j is the total number of customers in queue j. Assume queue j operates in the following manner:

(i) Each customer requires an amount of service which is a random variable exponentially distributed with unit mean.

(ii) A total service effort is supplied at the rate $\phi_j(n_j)$.

(iii) A proportion $\gamma_j(l, n_j)$ of this effort is directed to the customer in position l ($l = 1, 2, \ldots, n_j$); when this customer leaves the queue, his service

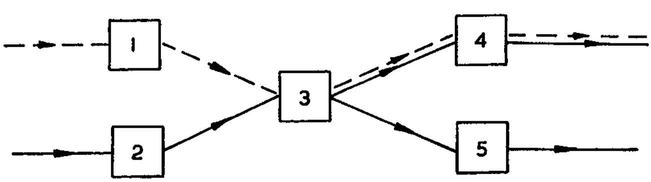

Fig. 3.2 Random routes

completed, customers in positions $l+1, l+2, \ldots, n_j$ move to positions $l, l+1, \ldots, n_j-1$ respectively.

(iv) When a customer arrives at queue j he moves into position l ($l=1,2,\ldots,n_j+1$) with probability $\delta_j(l, n_j+1)$; customers previously in positions $l, l+1, \ldots, n_j$ move to positions $l+1, l+2, \ldots, n_j+1$ respectively.

Of course

$$\sum_{l=1}^{n} \gamma_j(l, n) = 1$$

$$\sum_{l=1}^{n} \delta_j(l, n) = 1$$

and we shall insist that $\phi_j(n) > 0$ if $n > 0$. Call the amount of service a customer requires at a queue his *service requirement.* We shall assume that all service requirements, even of the same customer at different queues, are independent of each other and of the times at which customers enter the system. The way in which a customer's service requirement is satisfied can be visualized as follows. While the queue contains n_j customers, with him in position l, he receives service effort at the rate $\phi_j(n_j)\gamma_j(l, n_j)$ per unit time. When the amount of service effort he has received reaches his service requirement he leaves the queue. Since service requirements are exponentially distributed with unit mean, if queue j contains n_j customers then the probability intensity that the customer in position l leaves is $\phi_j(n_j)\gamma_j(l, n_j)$.

To illustrate the behaviour which can be allowed, if $\phi_j(n) = \lambda_j \min(K, n)$

$$\gamma_j(l, n) = \begin{cases} \dfrac{1}{n} & l = 1, 2, \ldots, n; \; n = 1, 2, \ldots, K \\ \dfrac{1}{K} & l = 1, 2, \ldots, K; \; n = K+1, K+2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_j(l, n) = \begin{cases} 1 & l = n \\ 0 & \text{otherwise} \end{cases}$$

then queue j behaves as a K-server queue in which customers have their service commenced in the order of their arrival and each customer has an exponentially distributed service time with mean λ_j^{-1}. In this example the service time of a customer can be identified with his service requirement, but this will not always be so. By varying ϕ_j we can allow the servers to work faster when the queue is large. By varying δ_j we can alter the queue discipline, making it, for example, last come first served or service in random order. A more subtle use of ϕ_j and γ_j will let a waiting customer defect at a rate depending upon his position in the queue. Note, however, that we cannot model a priority discipline based upon the type of a customer; nor can we allow a customer's service time to depend upon his service time at previous queues.

Let $t_j(l)$ be the type of the customer in position l in queue j and let $s_j(l)$ be the stage along his route that this customer has reached. We shall call $c_j(l) = (t_j(l), s_j(l))$ the class of this customer; if he can visit queue j more than once his class will contain more information than his type. The vector

$$\mathbf{c}_j = (c_j(1), c_j(2), \ldots, c_j(n_j))$$

describes the state of queue j and

$$\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_J)$$

is a Markov process representing the state of the system.

What are the transition rates of the process $\mathbf{C}$? If the customer in position l in queue j is at the last stage of his route then a possible event is that this customer may leave the system. Let $T_{jl}.\mathbf{C}$ be the state of the process after this event. The probability intensity of the event is

$$q(\mathbf{C}, l, \cdot, T_{jl}.\mathbf{C}) = \phi_j(n_j)\gamma_j(l, n_j) \tag{3.1}$$

It may be that $T_{jl}.\mathbf{C} = T_{jg}.\mathbf{C}$ for $l \neq g$, for example if all the customers in queue j are of the same type. The transition rate from the state $\mathbf{C}$ to the state $T_{jl}.\mathbf{C}$ is given by

$$q(\mathbf{C}, T_{jl}.\mathbf{C}) = \sum_g q(\mathbf{C}, g, \cdot, T_{jg}.\mathbf{C}) \tag{3.2}$$

where the summation runs over g such that $T_{jl}.\mathbf{C} = T_{jg}.\mathbf{C}$. If the customer in position l in queue j is not at the last stage of his route then let $k = r(t_j(l), s_j(l)+1)$ be the next queue he will visit. In this case a possible event is that this customer may leave queue j and move into position m in queue k. Let $T_{jlm}\mathbf{C}$ be the state of the process after this event. The probability intensity of the event is

$$q(\mathbf{C}, l, m, T_{jlm}\mathbf{C}) = \phi_j(n_j)\gamma_j(l, n_j)\delta_k(m, n_k+1) \tag{3.3}$$

The transition rate from the state $\mathbf{C}$ to the state $T_{jlm}\mathbf{C}$ is given by

$$q(\mathbf{C}, T_{jlm}\mathbf{C}) = \sum_g \sum_h q(\mathbf{C}, g, h, T_{jgh}\mathbf{C}) \tag{3.4}$$

where the summation runs over g and h such that $T_{jlm}\mathbf{C} = T_{jgh}\mathbf{C}$. Another possible event is that a customer of type i may enter the system and move into position m in queue k, where $k = r(i, 1)$. Let $T^{im}\mathbf{C}$ be the state of the process after this event. The probability intensity of the event is

$$q(\mathbf{C}, \cdot, m, T^{im}\mathbf{C}) = \nu(i)\delta_k(m, n_k+1) \tag{3.5}$$

The transition rate from the state $\mathbf{C}$ to the state $T^{iim}\mathbf{C}$ is given by

$$q(\mathbf{C}, T^{iim}\mathbf{C}) = \sum_h q(\mathbf{C}, \cdot, h, T^{ih}\mathbf{C}) \tag{3.6}$$

where the summation runs over h such that $T^{iim}\mathbf{C} = T^{ih}\mathbf{C}$.

Of course for a given state $\mathbf{C}$ it would not be appropriate to apply certain of the T operators defined above. However, we can say that any non-zero transition rate of the process $\mathbf{C}$ is of the form (3.2), (3.4), or (3.6).

Let

$$\alpha_j(i, s) = \begin{cases} \nu(i) & \text{if } r(i, s) = j \\ 0 & \text{otherwise} \end{cases}$$

and let

$$a_j = \sum_{i=1}^{I} \sum_{s=1}^{S(i)} \alpha_j(i, s)$$

If the system is in equilibrium then a_j will be the average number of customers arriving at queue j per unit time. Let

$$b_j^{-1} = \sum_{n=0}^{\infty} \frac{a_j^n}{\prod_{l=1}^{n} \phi_j(l)}$$

We shall assume that none of $b_1, b_2, \ldots, b_J$ is zero. This condition is imposed to ensure that an equilibrium distribution for the system exists, and if it is not satisfied at least one queue will be unable to cope with the number of customers arriving at it. Define

$$\pi_j(\mathbf{c}_j) = b_j \prod_{l=1}^{n_j} \frac{\alpha_j(t_j(l), s_j(l))}{\phi_j(l)}$$

Theorem 3.1. *The equilibrium distribution for the open network of queues described above is*

$$\pi(\mathbf{C}) = \prod_{j=1}^{J} \pi_j(\mathbf{c}_j)$$

Proof. First notice that $\pi(\mathbf{C})$ sums to unity, by the definition of the constants $b_1, b_2, \ldots, b_J$.

What might the Markov process $\mathbf{C}(t)$ look like if we reversed the direction of time? One possibility is that customers of type i might enter the system in a Poisson stream at rate $\nu(i)$ and pass through the sequence of queues

$$r(i, S(i)), r(i, S(i)-1), \ldots, r(i, 1)$$

before leaving the system, and that the queues of the system might behave as before but with the functions γ_j and δ_j interchanged. The reversed process $\mathbf{C}(-t)$ would then be of the same form as $\mathbf{C}(t)$, but with different parameters. With this in mind define, corresponding to the probability intensities (3.1), (3.3), and (3.5),

$$q'(T_{jl}.\mathbf{C}, \cdot, l, \mathbf{C}) = \nu(i)\gamma_j(l, n_j) \qquad \text{where } i = t_j(l)$$
$$q'(T_{jlm}\mathbf{C}, m, l, \mathbf{C}) = \phi_k(n_k+1)\delta_k(m, n_k+1)\gamma_j(l, n_j) \quad \text{where } k = r(t_j(l), s_j(l)+1)$$
$$q'(T^{im}\mathbf{C}, m, \cdot, \mathbf{C}) = \phi_k(n_k+1)\delta_k(m, n_k+1) \qquad \text{where } k = r(i, 1)$$

Similarly define the transition rates $q'(\mathbf{C}, \mathbf{D})$ by analogy with the definition of the transition rates $q(\mathbf{C}, \mathbf{D})$. By substituting the proposed form for $\pi(\mathbf{C})$ we see that

$$\pi(\mathbf{C})q(\mathbf{C}, l, m, T_{jlm}\mathbf{C}) = \pi(T_{jlm}\mathbf{C})q'(T_{jlm}\mathbf{C}, m, l, \mathbf{C})$$

Thus, by summation,

$$\pi(\mathbf{C})q(\mathbf{C}, T_{jlm}\mathbf{C}) = \pi(T_{jlm}\mathbf{C})q'(T_{jlm}\mathbf{C}, \mathbf{C})$$

In this way we can establish that for all $\mathbf{C}$ and $\mathbf{D}$,

$$\pi(\mathbf{C})q(\mathbf{C}, \mathbf{D}) = \pi(\mathbf{D})q'(\mathbf{D}, \mathbf{C})$$

We also find that

$$q(\mathbf{C}) = q'(\mathbf{C}) = \sum_{j=1}^{J} \phi_j(n_j) + \sum_{i=1}^{I} \nu(i)$$

Hence Theorem 1.13 allows us to deduce that $\pi(\mathbf{C})$ is the equilibrium distribution for the process $\mathbf{C}(t)$, which completes the proof of the present result.

We can also deduce from Theorem 1.13 that $\mathbf{C}(-t)$ does indeed take the form suggested; thus we obtain the following result.

Theorem 3.2. *If $\mathbf{C}(t)$ is a stationary open network of queues of the form described in this section then so is the reversed process $\mathbf{C}(-t)$.*

Theorems 3.1 and 3.2 parallel Theorems 2.4 and 2.5, and as in Chapter 2 there are some immediate consequences. Theorem 3.2 has the following corollary.

Corollary 3.3. *In equilibrium customers of type i ($i = 1, 2, \ldots, I$) leave the system in a Poisson stream at rate $\nu(i)$. These I Poisson streams are independent, and $\mathbf{C}(t_0)$ is independent of departures from the system prior to time t_0.*

If $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_J$ are possible states for the queues $1, 2, \ldots, J$ then $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_J)$ is a possible state for the system. This implies that the state space $\mathscr{S}$ has a product form and hence we can deduce from Theorem 3.1 that in equilibrium $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_J$ are independent.

Corollary 3.4. *In equilibrium the state of queue j is independent of the state of the rest of the system and is $\mathbf{c}_j$ with probability $\pi_j(\mathbf{c}_j)$. The probability that queue j contains n customers is*

$$P(n_j = n) = b_j \frac{a_j^n}{\prod_{l=1}^{n} \phi_j(l)} \tag{3.7}$$

If a customer is in position l in queue j then the probability that he is a type i customer at stage s of his route is $\alpha_j(i, s)/a_j$.

Equation (3.7) is exactly the expression we obtain if queue j is a single queue with customers arriving in a Poisson stream at rate a_j. Note, however, that in general arrivals at queue j do not form a Poisson process (cf. Exercise 2.4.2).

Corollary 3.5. *When a customer of type i reaches queue j at stage s of his route the probability that he finds queue j in state $\mathbf{c}_j$ is $\pi_j(\mathbf{c}_j)$. The probability that he finds n customers in queue j is given by expression (3.7).*

Proof. If $s = 1$ the result follows immediately from the fact that the arrival process of type i customers at the first queue on their route is Poisson. For $s > 1$ the proof proceeds along the same lines as the proof of Corollary 2.7. In equilibrium the probability flux that a customer of type i will depart from queue j after a given stage of his route and that the queue will be left in state $\mathbf{c}_j$ with n_j customers is

$$\sum_{l=1}^{n_j+1} \pi_j(\mathbf{c}_j) \frac{\nu(i)}{\phi_j(n_j+1)} \phi_j(n_j+1)\delta_j(l, n_j+1) = \nu(i)\pi_j(\mathbf{c}_j)$$

Thus if a customer of type i has just left queue j after a given stage of his route, the probability that he has left queue j in state $\mathbf{c}_j$ is $\pi_j(\mathbf{c}_j)$. Consideration of the reversed process now establishes the desired result.

If queue j is a first come first served K-server queue then Corollary 3.5 shows that the waiting time of a customer at this queue has the same distribution as if queue j were an isolated $M/M/K$ queue with a Poisson arrival process of rate a_j. Note that the waiting time of a customer at queue j will not in general be independent of his experience elsewhere in the network (cf. Exercise 2.2.5).

Exercises 3.1

1. Show that for the network illustrated in Fig. 3.1 the process $(n_1, n_2, \ldots, n_5)$ is not Markov.

2. Consider an open migration process with transition rates (2.8) where $\lambda_j = 1$, $j = 1, 2, \ldots, J$. Show that the process can be regarded as a queueing network with customers whose route will be $r(i, 1), r(i, 2), \ldots, r(i, S(i))$ arriving at rate

$$\nu_{r(i,1)}\lambda_{r(i,1),r(i,2)}\lambda_{r(i,2),r(i,3)} \cdots \lambda_{r(i,S(i)-1),r(i,S(i))}\mu_{r(i,S(i))}.$$

Observe that an infinite number of types will be required if a customer can visit the same queue more than once. Show that the quantities $a_1, a_2, \ldots, a_J$ calculated from the queueing network parameters are equal to the quantities $\alpha_1, \alpha_2, \ldots, \alpha_J$ determined by equations (2.9) from the migration process parameters.

3. Suppose that in the description given of a K-server queue the function δ_j is altered to

$$\delta_j(l, n) = \begin{cases} 1 & l = n;\ n = 1, 2, \ldots, K \\ 1 & l = K+1;\ n = K+1, K+2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Show that the resulting queue discipline is last come first served without preemption. If

$$\delta_j(l, n) = \frac{1}{n-K} \qquad l = K+1, K+2, \ldots, n;\ n = K+1, K+2, \ldots$$

show that the queue discipline is service in random order (i.e. that the queue is equivalent to one in which when a customer leaves the queue the next customer to be served is chosen at random from amongst those whose service has not yet commenced).

4. Suppose that $\phi_j(n) = \phi_j$ for all $n > 0$, $j = 1, 2, \ldots, J$, so that each queue is a single-server queue. Observe that n_j is a geometric random variable with mean $a_j/(\phi_j - a_j)$. Show that the number of type i customers at stage s of their route is also a geometric random variable, with mean $\nu(i)/(\phi_j - a_j)$ where $j = r(i, s)$. Observe that for differing values of i and s giving rise to the same value of j these random variables are dependent. Deduce from Little's result (1.12) that the mean time it takes a type i customer to pass through the system is

$$\sum_{s=1}^{S(i)} [\phi_{r(i,s)} - a_{r(i,s)}]^{-1}$$

5. Show that the restriction not allowing two successive stages of a customer's route to be identical can be removed.

6. The requirement that $\phi_j(n) > 0$ if $n > 0$ can be relaxed. Find the equilibrium distribution for a system in which

$$\phi_j(K) = 0$$
$$\phi_j(n) > 0 \qquad n > K$$

This form of the function ϕ_j would correspond to the servers at queue j only operating when more than K customers are present.

7. Show that the results of this section are unaltered if the functions $\gamma_j(l, n_j)$, $\delta_j(l, n_j)$ are replaced by functions $\gamma_j(l, \mathbf{c}_j)$, $\delta_j(l, \mathbf{c}_j)$, provided the functions γ_j and δ_j are invariant under permutations of $\mathbf{c}_j = (c_j(1), c_j(2), \ldots, c_j(n_j))$ and

$$\sum_{l=1}^{n_j} \gamma_j(l, \mathbf{c}_j) = \sum_{l=1}^{n_j} \delta_j(l, \mathbf{c}_j) = 1$$

3.2 OPEN NETWORKS OF QUASI-REVERSIBLE QUEUES

The routing mechanism introduced in the previous section is general enough for most purposes, but the queue described there is fairly limited in scope. In this section we shall show that essentially the same results can be obtained for any network of queues provided the queues have a certain important characteristic.

To define this characteristic we shall begin by considering a single isolated queue. We shall make quite weak assumptions about the nature of this queue; it could perhaps be visualized as a black box with a stream of customers entering the box and a further stream of customers leaving the box. Assume that every customer entering the queue leaves it but, for simplicity, not immediately. Assume also that at no point in time does more than one customer enter or leave the queue. Further assume that each customer has a class c chosen from a countable set $\mathscr{C}$ and that customers do not change class as they pass through the queue. Often the class of a customer will convey information about him; later we shall use it to provide an indication of his past and future route in a network and his service requirements at the various queues of the network. Suppose there is associated with the queue a Markov process $\mathbf{x}(t)$, which we shall call the state of the queue at time t. Assume that the state of the queue contains enough information for us to deduce how many customers of each class there are in the queue. Often the state will contain further information concerning, for example, the arrangement of customers within the queue or the amount of service still required by each customer. From now on we shall identify the queue with the Markov process $\mathbf{x}(t)$ giving its state. Observe that from a realization of the process $\mathbf{x}(t)$, $-\infty < t < \infty$, we can construct the arrival and departure processes of customers of class c, since such arrivals and departures are signalled by changes in the number of customers of class c in the queue.

Definition

A queue is *quasi-reversible* if its state $\mathbf{x}(t)$ is a stationary Markov process with the property that the state of the queue at time t_0, $\mathbf{x}(t_0)$, is independent

of:
(i) the arrival times of class c customers, $c \in \mathscr{C}$, subsequent to time t_0;
(ii) the departure times of class c customers, $c \in \mathscr{C}$, prior to time t_0.

Theorem 3.6. *If a queue is quasi-reversible then:*
(i) *arrival times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes;*
(ii) *departure times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes.*

Proof. Let $\mathscr{S}(c, \mathbf{x})$ be the set of states in which the queue contains one more customer of class c than in state $\mathbf{x}$, with the same numbers of customers of other classes. Thus a transition from the state $\mathbf{x}$ to a state $\mathbf{x}' \in \mathscr{S}(c, \mathbf{x})$ indicates the arrival of a customer of class c. Since the queue is quasi-reversible the probability a customer of class c arrives in the interval $(t_0, t_0+\delta t)$ is independent of the state $\mathbf{x}(t_0)$. Hence the probability intensity that a customer of class c arrives when the state is $\mathbf{x}$ depends only on c and not on $\mathbf{x}$; call it

$$\alpha(c) = \sum_{\mathbf{x}' \in \mathscr{S}(c,\mathbf{x})} q(\mathbf{x}, \mathbf{x}') \tag{3.8}$$

Since $\mathbf{x}(t)$ is a Markov process the realization $\mathbf{x}(t)$, $-\infty < t \leq t_0$, contains no more information than does $\mathbf{x}(t_0)$ about whether or not a class c customer will arrive in the interval $(t_0, t_0+\delta t)$. But this realization gives the arrival times of class c customers, for $c \in \mathscr{C}$, prior to time t_0. Hence the probability intensity that a customer of class c will arrive is $\alpha(c)$, even given all prior arrival times of class c customers, for $c \in \mathscr{C}$. Hence arrival times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes.

Consider now the reversed process $\mathbf{x}(-t)$. This can also be regarded as a queue: again transitions from the state $\mathbf{x}$ to a state $\mathbf{x}' \in \mathscr{S}(c, \mathbf{x})$ indicate the arrival of a customer of class c and transitions to the state $\mathbf{x}$ from a state $\mathbf{x}' \in \mathscr{S}(c, \mathbf{x})$ indicate the departure of a customer of class c. Observe that arrivals at the reversed queue $\mathbf{x}(-t)$ subsequent to time $-t_0$ correspond to departures from the queue $\mathbf{x}(t)$ prior to time t_0. Similarly, departures from the reversed queue $\mathbf{x}(-t)$ prior to time $-t_0$ correspond to arrivals at the queue $\mathbf{x}(t)$ subsequent to time t_0. Since the queue $\mathbf{x}(t)$ is quasi-reversible it therefore follows that the reversed queue $\mathbf{x}(-t)$ is also quasi-reversible. Thus at the reversed queue $\mathbf{x}(-t)$ arrival times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes. Thus at the queue $\mathbf{x}(t)$ departure times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes.

Although conclusions (i) and (ii) of Theorem 3.6 are the most obvious features of a quasi-reversible queue they cannot be taken as the definition of quasi-reversibility. These conclusions include no mention of the Markov process $\mathbf{x}(t)$ defining the state of the queue, and it is possible to construct systems satisfying conclusions (i) and (ii) which are not quasi-reversible (Exercises 3.2.2 and 3.2.3).

It usually follows from the definition of the process $\mathbf{x}(t)$ that $\mathbf{x}(t_0)$ is independent of subsequent arrivals. Often the form of the reversed process $\mathbf{x}(-t)$ allows us to deduce that $\mathbf{x}(t_0)$ is also independent of prior departures and hence that the queue is quasi-reversible. An example of a quasi-reversible queue is an $M/M/1$ queue with one class of customer and with $\mathbf{x}(t)$ the number in the queue at time t. Theorem 2.1 establishes that this queue is quasi-reversible. More generally, if a queue has one class of customer, a Poisson arrival process, and the state of the queue is a reversible Markov process independent of future arrivals, then the queue will be quasi-reversible. More complicated examples are provided by the networks of the previous section. If the class of a customer is taken to be its type then Corollary 3.3 shows that a network, considered in its entirety as a single system, is a quasi-reversible queue. The special case in which the network consists of just one queue shows that a single queue of the form discussed in the previous section is quasi-reversible. Further examples of quasi-reversible queues will be discussed in the next section.

If $\pi(\mathbf{x})$ is the equilibrium distribution of the queue $\mathbf{x}(t)$ then the transition rates of the reversed queue $\mathbf{x}(-t)$ are given by

$$\pi(\mathbf{x})q'(\mathbf{x}, \mathbf{x}') = \pi(\mathbf{x}')q(\mathbf{x}', \mathbf{x}) \tag{3.9}$$

Departures of class c customers from the queue $\mathbf{x}(t)$ form a Poisson process; the rate of this process must be $\alpha(c)$ since this is the arrival rate and the queue is in equilibrium. Hence the arrival rate of class c customers at the reversed queue $\mathbf{x}(-t)$ is also $\alpha(c)$, and so

$$\alpha(c) = \sum_{\mathbf{x}' \in \mathscr{S}(c,\mathbf{x})} q'(\mathbf{x}, \mathbf{x}') \tag{3.10}$$

This is an important result; relations (3.8) and (3.10) characterize the property of quasi-reversibility for a stationary Markov process $\mathbf{x}(t)$. Using equations (3.8), (3.9), and (3.10) we can obtain the partial balance equations

$$\pi(\mathbf{x}) \sum_{\mathbf{x}' \in \mathscr{S}(c,\mathbf{x})} q(\mathbf{x}, \mathbf{x}') = \sum_{\mathbf{x}' \in \mathscr{S}(c,\mathbf{x})} \pi(\mathbf{x}')q(\mathbf{x}', \mathbf{x}) \tag{3.11}$$

Thus in equilibrium the probability flux out of a state due to a customer of class c arriving is equal to the probability flux into that same state due to a customer of class c departing. Since the probability flux that a customer of class c arrives at the queue is equal to the probability flux that a customer of class c departs from the queue, this shows that the distribution over states found by an arriving customer of class c is the same as that left behind by a departing customer of class c. If the process $\mathbf{x}(t)$ is reversible then the partial balance equations (3.11) are automatically satisfied; however equations (3.8) and (3.10) will only be satisfied if the arrival rate of class c customers is independent of the state of the queue. Thus quasi-reversibility differs from reversibility in that a stronger condition (3.8) is imposed on the arrival rates and a weaker condition (3.11) is imposed on the probability fluxes.

In the remainder of this section we shall extend our previous results on open networks of queues to apply to the case where the queues are quasi-reversible. If the network is of a certain fairly simple form this can be done easily. Suppose that customers pass through the network in accordance with routes determined by their types as described in the previous section. Associate with each customer arriving at queue j its class (i, s), i.e. its type and the stage of its route it has reached. Thus $j = r(i, s)$. Note that a customer's class does not alter while it passes through a queue, but changes as it moves from one queue to another. If the routes through the system allow the queues to be ordered so that a customer leaving a queue always moves to a queue later in the order (as in the case in Figs. 3.1 and 3.2) then the assumption of quasi-reversibility together with the arguments of Section 2.2 show that in equilibrium the states of the queues are independent. In this simple case the arrival streams at each queue are Poisson; we cannot hope for this to be true in more general networks, and so for these a different approach is required.

Let $\pi_j(\mathbf{x}_j)$ be the equilibrium distribution of a quasi-reversible queue at which arrivals of customers of class (i, s) form a Poisson process of rate $\alpha_j(i, s)$. Let $q_j(\mathbf{x}_j, \mathbf{x}_j')$ be the transition rates of this process and let $S_j(i, s, \mathbf{x}_j)$ be the set of states in which the queue contains one more customer of class (i, s) than in state $\mathbf{x}_j$, with the same number of customers of other classes. Consider now a Markov process $\mathbf{X}(t) = (\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_J(t))$ whose transition rates are defined as follows. The probability intensity that a customer of type i enters the system and causes queue $k = r(i, 1)$ to change from state $\mathbf{x}_k$ to state $\mathbf{x}_k' \in \mathcal{S}_k(i, 1, \mathbf{x}_k)$ is $q_k(\mathbf{x}_k, \mathbf{x}_k')$. The probability intensity that a customer of type i leaves the system and causes queue $j = r(i, S(i))$ to change from state $\mathbf{x}_j' \in \mathcal{S}_j(i, S(i), \mathbf{x}_j)$ to state $\mathbf{x}_j$ is $q_j(\mathbf{x}_j', \mathbf{x}_j)$. The probability intensity that a customer of class (i, s), $s < S(i)$, leaves queue $j = r(i, s)$ and enters queue $k = r(i, s+1)$ as a customer of class $(i, s+1)$, causing queue j to change from state $\mathbf{x}_j' \in \mathcal{S}_j(i, s, \mathbf{x}_j)$ to state $\mathbf{x}_j$ and queue k to change from state $\mathbf{x}_k$ to state $\mathbf{x}_k' \in \mathcal{S}_k(i, s+1, \mathbf{x}_k)$, is

$$q_j(\mathbf{x}_j', \mathbf{x}_j)\frac{q_k(\mathbf{x}_k, \mathbf{x}_k')}{\sum_{\mathbf{x}' \in \mathcal{S}_k(i, s+1, \mathbf{x}_k)} q_k(\mathbf{x}_k, \mathbf{x}')} = q_j(\mathbf{x}_j', \mathbf{x}_j)\frac{q_k(\mathbf{x}_k, \mathbf{x}_k')}{\alpha_k(i, s+1)}$$

using equation (3.8). Finally, the probability intensity that there is an internal change in queue j from state $\mathbf{x}_j$ to state $\mathbf{x}_j'$, without the arrival or departure of any customer, is $q_j(\mathbf{x}_j, \mathbf{x}_j')$. The transition rates are thus defined in the obvious way: a queue behaves just as it would in isolation except that arrivals of class (i, s) customers, for $s > 1$, are triggered by departures from another queue rather than by an independent Poisson process. If queues

$1, 2, \ldots, J$ would in isolation be quasi-reversible and if the process $\mathbf{X}$ is in equilibrium then we shall call $\mathbf{X}$ an open network of quasi-reversible queues. Note that the jth queue of the network will not in general satisfy the conditions required for it to be quasi-reversible, and indeed the jth component of $\mathbf{X}$, $\mathbf{x}_j$, will not in general be a Markov process. Nevertheless, we shall occasionally abuse terminology and call queue j quasi-reversible—it would be if it were in isolation.

What might the reversed process $\mathbf{X}(-t)$ look like? The obvious possibility is that customers of type i might enter the system in a Poisson stream and pass backwards along their route and that the jth component of the system, $\mathbf{x}_j(-t)$, might be derived from the reversed version of queue j considered in isolation. Using Theorem 1.13 it becomes a routine matter to establish that this is indeed the reversed process and that the equilibrium distribution is

$$\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) = \pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2) \cdots \pi_J(\mathbf{x}_J) \tag{3.12}$$

The suggested probability intensity for the reversed process that a customer of class $(i, s+1)$ leaves queue $k = r(i, s+1)$ and enters queue $j = r(i, s)$ as a customer of class (i, s), causing queue k to change from state $\mathbf{x}'_k \in \mathcal{S}_k(i, s+1, \mathbf{x}_k)$ to state $\mathbf{x}_k$ and queue j to change from state $\mathbf{x}_j$ to state $\mathbf{x}'_j \in \mathcal{S}_j(i, s, \mathbf{x}_j)$, is

$$q'_k(\mathbf{x}'_k, \mathbf{x}_k) \frac{q'_j(\mathbf{x}_j, \mathbf{x}'_j)}{\sum_{\mathbf{x}' \in \mathcal{S}_j(i, s, \mathbf{x}_j)} q'_j(\mathbf{x}_j, \mathbf{x}')} = q'_k(\mathbf{x}'_k, \mathbf{x}_k) \frac{q'_j(\mathbf{x}_j, \mathbf{x}'_j)}{\alpha_j(i, s)}$$

from equation (3.10), which in turn followed from the quasi-reversibility of queue j. To establish condition (1.28) of Theorem 1.13 for transitions arising from the movement of customers from one queue to another we need therefore to show that

$$\frac{\pi_j(\mathbf{x}'_j)\pi_k(\mathbf{x}_k)q_j(\mathbf{x}'_j, \mathbf{x}_j)q_k(\mathbf{x}_k, \mathbf{x}'_k)}{\alpha_k(i, s+1)} = \frac{\pi_j(\mathbf{x}_j)\pi_k(\mathbf{x}'_k)q'_k(\mathbf{x}'_k, \mathbf{x}_k)q'_j(\mathbf{x}_j, \mathbf{x}'_j)}{\alpha_j(i, s)} \tag{3.13}$$

But this follows from equation (3.9) and the observation that $\alpha_k(i, s+1) = \alpha_j(i, s) = \nu(i)$. Condition (1.28) is established even more easily for transitions associated with the arrival at or departure from the system of a customer. The only remaining transitions are those where a single queue changes its state without the arrival or departure of a customer. Equation (3.9) establishes condition (1.28) directly for such transitions. We must finally check condition (1.27) of Theorem 1.13:

$$\begin{aligned} q(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) &= \sum_{j=1}^{J} \left(q_j(\mathbf{x}_j) - \sum_{(i,s)} \alpha_j(i,s) \right) + \sum_{i=1}^{I} \nu(i) \\ &= \sum_{j=1}^{J} \left(q'_j(\mathbf{x}_j) - \sum_{(i,s)} \alpha_j(i,s) \right) + \sum_{i=1}^{I} \nu(i) \\ &= q'(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) \end{aligned}$$

Thus the reversed process does take the conjectured form and the equilibrium distribution is given by expression (3.12).

As usual the form of the reversed process allows much to be deduced about the original process. The probability flux that a customer of class (i, s) departs from queue $j = r(i, s)$ and that queue j is left in state $\mathbf{x}_j$ is

$$\sum_{\mathbf{x}' \in \mathcal{S}_j(s,\, i,\, \mathbf{x}_j)} \pi_j(\mathbf{x}')q_j(\mathbf{x}'_j, \mathbf{x}_j) = \pi_j(\mathbf{x}_j)\alpha_j(i, s)$$

from equations (3.9) and (3.10). Thus if a customer of class (i, s) has just left queue j the probability he has left queue j in state $\mathbf{x}_j$ is $\pi_j(\mathbf{x}_j)$. The corresponding statement also holds for the reversed process, and hence a customer of class (i, s) arriving at queue $j = r(i, s)$ finds the queue in state $\mathbf{x}_j$ with probability $\pi_j(\mathbf{x}_j)$.

We can summarize the results of this section as follows.

Theorem 3.7. *An open network of quasi-reversible queues has the following properties:*

(i) *The states of the individual queues are independent.*

(ii) *For an individual queue the equilibrium distribution and the distribution over states found by an arriving customer of a given class are identical and are both as they would be if the queue were in isolation with arrivals of customers of each class forming independent Poisson processes.*

(iii) *Under time reversal the system becomes another open network of quasi-reversible queues.*

(iv) *The system itself is quasi-reversible and so departures from the system of customers of each type form independent Poisson processes, and the state of the system at time t_0 is independent of departures from the system prior to time t_0.*

Exercises 3.2

1. In the description of a quasi-reversible queue it was assumed that every customer who entered the queue left it, that customers did not change class as they passed through the queue, and that the process $\mathbf{x}(t)$ recorded how many customers of each class the queue contained. While these assumptions help us to visualize the queue they are not necessary. Show that the analysis of this section is unaltered if they are replaced by the weaker assumptions that the arrivals and departures of class c customers are signalled by transitions of the process $\mathbf{x}(t)$ and that the equilibrium arrival and departure rates of class c customers are equal, for $c \in \mathscr{C}$.
2. The definition of quasi-reversibility characterizes the Markov process representing the state of the queue, rather than any more fundamental property of the queue itself. It is quite possible that there may be two

representations of the same physical mechanism, one of which is quasi-reversible and the other not. Consider, for example, an isolated queue of the form described in the last section. Let the state of the queue be $(\mathbf{c}, c)$ where c is the class of the last customer to leave the queue. Show that with this representation the queue is not quasi-reversible.

3. Consider a stationary $M/M/1$ queue. Suppose that when a customer arrives at the queue a clerk issues him with a ticket and that when the customer leaves the queue he returns the ticket to the clerk (the purpose of the tickets may be to maintain the queue discipline). Now regard the clerk's office as a system in its own right and regard the tickets entering and leaving the office as customers. Show that although the arrival and departure streams are Poisson processes the system is not quasi-reversible however its state is defined, even under the weaker assumptions of Exercise 3.2.1.
4. Consider a queue with a Poisson arrival process and a state which is a reversible Markov process independent of future arrivals, e.g. the two-server queue considered in Section 1.5. Suppose now that each customer arriving at the queue is randomly allocated a class from the set $\mathscr{C}$, so that arrival times of class c customers, for $c \in \mathscr{C}$, form independent Poisson processes. Suppose further that the passage of a customer through the queue is unaffected by his class. Show that if the state of the queue is now taken to be the original reversible Markov process together with the classes of the customers in the queue arranged in order of their arrival, then the queue is quasi-reversible. If the passage of a customer through the queue *is* affected by his class then the queue may not be quasi-reversible however its state is defined, as the next exercise shows.
5. Arrivals of customers of types 1 and 2 at a single-server queue form independent Poisson processes. The service requirements of customers are independent and all have the same exponential distribution. The server gives priority to customers of type 1, and will even interrupt the service of a type 2 customer if a type 1 customer arrives. Deduce that departures from the queue form a Poisson process and that departures of type 1 customers form a Poisson process. Show that departures of type 2 customers do not form a Poisson process.
6. It was assumed early in this section that a customer entering a queue could not leave it immediately. Certain systems, e.g. the telephone exchange model of Section 2.1 or the queue with balking considered in Exercise 2.1.1, satisfy all the conditions for quasi-reversibility apart from this assumption. The assumption can be relaxed provided we deal with two technical difficulties. The first is that we must require that all arrivals and departures of class c customers are signalled by changes in the state of the queue, for each $c \in \mathscr{C}$. If $\mathscr{C}$ is finite it is easy to comply with this requirement using flip-flop variables as described in Section 2.1. The

second difficulty is that our definition of a network requires that a customer who enters a queue and leaves it immediately must go on to the next queue on his route. There will thus exist transitions of the Markov process $\mathbf{X}$ involving more than two queues. Extend the analysis of this section to deal with this difficulty. Observe that both difficulties can be avoided by using the method of Exercise 2.1.1(ii), whereby a customer pauses momentarily instead of leaving the queue immediately.

3.3 SYMMETRIC QUEUES

The quasi-reversible queues considered in Section 3.1 possess the property that the service requirement of a customer is exponentially distributed. This property simplifies analysis, since it removes the need for the state of the queue to include information on the amount of service customers have received. A more general distribution which can be handled with a little more effort is the gamma distribution. This arises when a customer requires a number of stages of service, each of which consists of an independent exponentially distributed amount of service. In this section we shall consider a range of queues which turn out to be quasi-reversible even when service requirements are not exponentially distributed. Initially we shall allow only service requirements which have a gamma distribution, but later we shall remove this restriction.

Consider a queue within which customers are ordered, with the queue containing customers in positions $1, 2, \ldots, n$, where n is the total number of customers in the queue. We shall call such a queue *symmetric* if it operates in the following manner:

(i) The service requirement of a customer is a random variable whose distribution may depend upon the class of the customer.

(ii) A total service effort is supplied at the rate $\phi(n)$.

(iii) A proportion $\gamma(l, n)$ of this effort is directed to the customer in position l $(l = 1, 2, \ldots, n)$; when this customer leaves the queue customers in positions $l+1, l+2, \ldots, n$ move to positions $l, l+1, \ldots, n-1$ respectively.

(iv) When a customer arrives at the queue he moves into position l $(l = 1, 2, \ldots, n+1)$ with probability $\gamma(l, n+1)$; customers previously in positions $l, l+1, \ldots, n$ move to positions $l+1, l+2, \ldots, n+1$ respectively.

Of course

$$\sum_{l=1}^{n} \gamma(l, n) = 1$$

and we shall insist that $\phi(n) > 0$ if $n > 0$. The queue described differs from those of Section 3.1 in that service requirements are not restricted and the symmetry condition $\gamma \equiv \delta$ is imposed. This condition rules out many queue disciplines, e.g. first come first served, and indeed at a symmetric queue there will be little queueing at all, in the usual sense of the word. Nevertheless, some useful systems can be set up as symmetric queues, and we shall describe four examples.

A server-sharing queue. When

$$\gamma(l, n) = \frac{1}{n} \qquad l = 1, 2, \ldots, n;\ n = 1, 2, \ldots$$

the service effort is shared equally between all customers in the queue. If $\phi(n) = 1$ for $n > 0$ then the queue behaves as a single-server queue, and a customer's remaining service requirement decreases at rate $1/n$.

A stack. When

$$\gamma(l, n) = 1 \qquad l = n;\ n = 1, 2, \ldots$$

the total service effort is directed to the customer who last arrived. Such a queue is best visualized as a stack, with customers arriving at and departing from the top of the stack. If $\phi(n) = 1$ for $n > 0$ then we have a single-server queue at which the queue discipline is last come first served with preemption (cf. Exercise 1.3.8).

A queue with no waiting room. Consider the functions

$$\begin{aligned} \phi(n) &= n && n = 1, 2, \ldots, K \\ \phi(n) &= \xi && n = K+1, K+2, \ldots \\ \gamma(l, n) &= \frac{1}{n} && l = 1, 2, \ldots, n;\ n = 1, 2, \ldots, K \\ \gamma(l, n) &= 1 && l = n;\ n = K+1, K+2, \ldots \end{aligned}$$

where ξ is very large. We can regard this queue as one with K available servers at which a customer who arrives to find all K servers occupied leaves almost immediately. We have chosen not to make ξ infinite since this would entail a minor technical difficulty. It would allow an arrival and a departure to occur at the same time and not cause a change of state. This difficulty could be overcome using the flip-flop variable described in Section 2.1 in connection with the telephone exchange model, which would then be a special case.

An infinite-server queue. If

$$\begin{aligned} \phi(n) &= n && n = 1, 2, \ldots \\ \gamma(l, n) &= \frac{1}{n} && l = 1, 2, \ldots, n;\ n = 1, 2, \ldots \end{aligned}$$

then the queue behaves as a queue with an infinite number of servers, with each customer having a server to himself; in this case customers do not affect each other within the queue. An infinite-server queue can be regarded as a special case of either a server-sharing queue or a queue with no waiting room.

Consider now a symmetric queue at which customers of class c arrive in a Poisson stream at rate $\nu(c)$. Suppose that a class c customer requires $w(c)$ stages of service, each of which consists of an independent exponentially distributed amount of service with mean $d(c)$. The service requirement of a class c customer will then have a gamma distribution, with mean $w(c)d(c)$ and variance $w(c)d(c)^2$.

Let $c(l)$ be the class of the customer in position l and suppose that his service has reached stage $u(l)$, where $1 \le u(l) \le w(c(l))$. Let $\mathbf{c}(l) = (c(l), u(l))$. Then

$$\mathbf{c} = (\mathbf{c}(1), \mathbf{c}(2), \ldots, \mathbf{c}(n))$$

(where n is the number in the queue) is a Markov process representing the state of the queue. We will now show that its equilibrium distribution is

$$\pi(\mathbf{c}) = b \prod_{l=1}^{n} \frac{\nu(c(l))d(c(l))}{\phi(l)} \tag{3.14}$$

provided the normalizing constant given by

$$b^{-1} = \sum_{n=0}^{\infty} \frac{a^n}{\prod_{l=1}^{n} \phi(l)} \tag{3.15}$$

where

$$a = \sum_{c} \nu(c)d(c)w(c)$$

is positive. Note that a is the average amount of service requirement arriving at the queue per unit time. It is fairly easy to show that expression (3.14) is the equilibrium distribution, since there is an obvious candidate for the reversed process, namely a queue at which arrivals are Poisson and which operates in precisely the same manner but with $u(l)$ recording the number of stages *yet to be completed* before the customer in position l leaves the queue. We shall now verify this. The probability intensity that a customer of class c arrives at the original queue and moves into position l is $\nu(c)\gamma(l, n+1)$, where n was the number previously in the queue. Let this event cause a transition from the state $\mathbf{c}$ to the state $\mathbf{c}'$. The probability intensity that when the state of the reversed queue is $\mathbf{c}'$ the customer in position l departs from the queue is $\phi(n+1)\gamma(l, n+1)/d(c)$. From the form

(3.14) we see that

$$\pi(\mathbf{c}') = \pi(\mathbf{c}) \frac{\nu(c)d(c)}{\phi(n+1)}$$

and hence that

$$\pi(\mathbf{c})\nu(c)\gamma(l, n+1) = \frac{\pi(\mathbf{c}')\phi(n+1)\gamma(l, n+1)}{d(c)}$$

Hence we can show that condition (1.28) of Theorem 1.13 holds for transitions arising from arrivals at the queue. Similarly, we can show that it holds for transitions caused by departures from the queue. The only remaining transitions are those which occur when an intermediate stage of a customer's service is completed. But if this causes a transition from $\mathbf{c}$ to $\mathbf{c}'$ then the transition rates $q(\mathbf{c}, \mathbf{c}')$ in the original process and $q(\mathbf{c}', \mathbf{c})$ in the reversed process are equal, and so are $\pi(\mathbf{c})$ and $\pi(\mathbf{c}')$. (Observe that it is the possibility of such a transition which differentiates the queue from those considered in Section 3.1 and which necessitates the symmetry condition $\gamma \equiv \delta$.) Finally, it is clear that $q(\mathbf{c}) = q'(\mathbf{c})$, and hence Theorem 1.13 shows that expression (3.14) does indeed give the equilibrium distribution and that the reversed process is of the suggested form. This in turn establishes that a symmetric queue in equilibrium is quasi-reversible, at least when service requirements have gamma distributions. In fact the queue is dynamically reversible with the conjugacy relation defined by

$$u^+(l) = w(c(l)) - u(l) + 1$$
$$\mathbf{c}^+(l) = (c(l), u^+(l))$$
$$\mathbf{c}^+ = (\mathbf{c}^+(1), \mathbf{c}^+(2), \dots, \mathbf{c}^+(n))$$

If the sum in equation (3.15) is infinite then the queue cannot reach equilibrium: work arrives at the queue more quickly than it can be dealt with.

The equilibrium distribution (3.14) has some interesting implications. The probability there are n customers in the queue is

$$\frac{ba^n}{\prod_{l=1}^n \phi(l)} \tag{3.16}$$

Further, given there are n customers in the queue, $\mathbf{c}(1), \mathbf{c}(2), \dots, \mathbf{c}(n)$ are independent. The customer in position l is of class c with probability

$$\frac{\nu(c)d(c)w(c)}{a} \tag{3.17}$$

and $u(l)$ is equally likely to be any value in the range $1 \le u(l) \le w(c(l))$. The constant a and the probabilities (3.16) and (3.17) depend on the values $d(c)$

and $w(c)$ only through the product $d(c)w(c)$, which is the mean of the service requirement distribution.

Suppose now that when a customer of class c arrives at the queue he is allocated a finer classification, (c, z), with probability $p(c, z)$, where z belongs to a countable set $\mathscr{Z}$, and $\sum_z p(c, z) = 1$ for each c. Then arrivals at the queue of customers of class (c, z) form a Poisson process of rate $\nu(c)p(c, z)$. If the service requirement of a customer of class (c, z) has a gamma distribution with mean $w(c, z)d(c, z)$ and variance $w(c, z)d(c, z)^2$, then the preceding analysis still applies with regard to the finer classification. The service requirement distribution of a customer of class c is now a gamma distribution with mean $w(c, z)d(c, z)$ and variance $w(c, z)d(c, z)^2$ with probability $p(c, z)$ for $z \in \mathscr{Z}$, i.e. it is a *mixture* of gamma distributions. The mean service requirement of a customer of class c is

$$a(c) = \sum_z p(c, z)w(c, z)d(c, z)$$

and the average amount of service requirement arriving at the queue per unit time is

$$\begin{aligned} a &= \sum_c \sum_z \nu(c)p(c, z)w(c, z)d(c, z) \\ &= \sum_c \nu(c)a(c) \end{aligned}$$

Let $\mathbf{c}(l) = (c(l), z(l), u(l))$ where $(c(l), z(l))$ is the refined classification of the customer in position l, and again take $\mathbf{c} = (\mathbf{c}(1), \mathbf{c}(2), \ldots, \mathbf{c}(n))$ to be the state of the queue. Then the equilibrium distribution is now

$$\pi(\mathbf{c}) = b \prod_{l=1}^{n} \frac{\nu(c(l))p(c(l), z(l))d(c(l), z(l))}{\phi(l)} \tag{3.18}$$

where b is defined as before by equation (3.15). Thus the probability there are n customers in the queue is again (3.16), and if there is a customer in position l the probability he is of class c is $\nu(c)a(c)/a$. These probabilities depend on the parameters $p(c, z)$, $d(c, z)$, and $w(c, z)$ defining the distribution of service requirement for a class c customer only through the mean service requirement $a(c)$. If we are given the class c of the customer in position l then the probability his refined classification is (c, z) is $p(c, z)d(c, z)w(c, z)/a(c)$. If we are given the refined classification (c, z) of the customer in position l then we will know $w(c, z)$, i.e. how many stages his service consists of; the number of the stage he has reached, $u(l)$, is equally likely to be any number between 1 and $w(c, z)$. Let us record some of these conclusions in the following theorem.

Theorem 3.8. *A stationary symmetric queue* $\mathbf{c}$ *at which service requirement distributions are mixtures of gamma distributions has the following* *properties:*

(i) *The probability the queue contains n customers is*

$$\frac{ba^n}{\prod_{l=1}^{n} \phi(l)}$$

(ii) *Given there are n customers in the queue the classes of the customers are independent and the probability the customer in a given position is of class c is*

$$\frac{\nu(c)a(c)}{a}$$

(iii) *The queue is quasi-reversible with respect to either the classification c or the refined classification* (c, z).

Suppose now that $\mathscr{Z}$ is a collection of positive numbers. For each $z \in \mathscr{Z}$ let $d(c, z)$ become very small and $w(c, z)$ very large, with $w(c, z)d(c, z)$ fixed at the value z. The variance of the gamma distribution associated with the refined classification (c, z), $w(c, z)d(c, z)^2$, tends to zero, while the mean remains at z. We can thus approximate a service requirement of exactly z. By using several refined classifications (c, z) for the class c it is possible to approximate as closely as we please an arbitrary distribution of service requirement; this is stated more precisely in the next result, which we shall prove in Exercise 3.3.3.

Lemma 3.9. *Let $F(x)$ be the distribution function of a positive random variable. Then it is possible to choose a sequence of distribution functions $F_m(x)$, each of which corresponds to a mixture of gamma distributions, so that*

$$\lim_{m \to \infty} F_m(x) = F(x)$$

for all x at which F is continuous.

Lemma 3.9 strongly suggests that Theorem 3.8 will remain valid without the restriction that service requirement distributions be mixtures of gamma distributions. This is in fact the case although we will not be able to prove it here, since a symmetric queue with arbitrary service requirement distributions cannot be represented by a Markov process with a countable state space. A continuous state space is required, and we shall have to content ourselves with a brief sketch of the results.

Consider then a symmetric queue at which customers of class c arrive in a Poisson stream of rate $\nu(c)$, and suppose the service requirement distribution of a class c customer has distribution function $F_c(x)$, with mean $a(c)$. Let

$$a = \sum_c \nu(c)a(c)$$

and define b as before by equation (3.15). Let $c(l) \in \mathscr{C}$ be the class of the customer in position l, let $z(l) \in (0, \infty)$ be his service requirement, and let $u(l) \in (0, z(l))$ be the amount of service effort he has so far received. Let $\mathbf{c}(l) = (c(l), z(l), u(l))$ and take $\mathbf{c} = (\mathbf{c}(1), \mathbf{c}(2), \ldots, \mathbf{c}(n))$ to be the state of the queue. Observe that the process $\mathbf{c}$ is Markov with a continuous state space. Jumps in the process $\mathbf{c}$ occur when a customer arrives or departs, but between these jumps $\mathbf{c}$ changes continuously, with $u(l)$ increasing linearly at rate $\phi(n)\gamma(l, n)$. When $u(l)$ reaches $z(l)$ the customer in position l leaves the queue.

Theorem 3.10. *A stationary symmetric queue* $\mathbf{c}$ *at which service requirements are arbitrarily distributed has properties* (i) *to* (iii) *listed in Theorem 3.8. In addition:*

(iv) *Given the number of customers in the queue and the class of each of them, the amounts of service effort the customers have received are independent, and the probability a customer of class c has received an amount of service effort not greater than x is*

$$F_c^*(x) = \frac{1}{a(c)} \int_0^x (1 - F_c(z))\, dz$$

Outline of proof. The theorem is proved by showing that the equilibrium distribution is the probability density

$$b \prod_{l=1}^{n} \frac{du(l)\, dF_{c(l)}(z(l))}{\phi(l)}\, \nu(c(l)) \tag{3.19}$$

$n = 0, 1, 2, \ldots, c(l) \in \mathscr{C}$, $0 < z(l) < \infty$, $0 < u(l) < z(l)$, $l = 1, 2, \ldots, n$, and that the reversed process is an identical symmetric queue but with $u(l)$ recording the amount of service effort yet to be received by the customer in position l. These facts can be established by a direct consideration of the process $\mathbf{c}$ or by a limiting argument based on a sequence of symmetric queues chosen so that the limit of the sequence is the process $\mathbf{c}$, but where at each queue in the sequence service requirement distributions are mixtures of gamma distributions.

The equilibrium distribution (3.19) is a density with respect to $u(1), u(2), \ldots, u(n)$, $z(1), z(2), \ldots, z(n)$ for each value of n and for each arrangement $(c(1), c(2), \ldots, c(n))$. Properties (i), (ii), and (iv) follow by integrating this density over the appropriate values of these variables. For example the distribution (3.19) shows that, given the number of customers in the queue and the class of each of them, $z(l)$ is distributed with density

$$\frac{1}{a(c(l))} z(l)\, dF_{c(l)}(z(l)) \qquad 0 < z(l) < \infty$$

and, given $z(l)$, $u(l)$ is distributed uniformly on $(0, z(l))$. Hence the probability a customer of class c has received an amount of service effort greater than x is

$$\frac{1}{a(c)} \int_x^\infty \left[\int_x^z du \right] dF_c(z)$$

which reduces to $1 - F_c^*(x)$.

The probabilities in (i) and (ii) are *insensitive* to the form of the distribution functions $F_c(x)$ in that they depend upon them only through their means $a(c)$. Thus, for example, Erlang's formula (1.13), calculated for a telephone exchange model in which call lengths are exponentially distributed, holds even when call lengths are arbitrarily distributed since the model is a symmetric queue.

The distribution function $F_c^*(x)$ is familiar as the equilibrium age distribution of a renewal process in which components have lifetime distribution $F_c(x)$. We can in fact derive this from our results. Consider a queue with no waiting room and with just one server. Suppose that there is just one class of customer c and that the arrival rate $\nu(c)$ is very large. This queue is equivalent to a renewal process, since as soon as the single customer in the queue is served (a component fails) he is replaced by another customer (a new component). Thus in equilibrium the age of the component in use has distribution function $F_c^*(x)$.

The above example illustrates a minor difficulty which can arise with arbitrary distributions. If lifetimes are all the same fixed constant, the renewal process is periodic and will not approach equilibrium unless it starts there. Periodicity cannot arise when distributions are mixtures of gamma distributions, but must be watched for in general.

The number in a symmetric queue will not usually be a Markov process. Nevertheless, the form of the reversed process leads immediately to the following result.

Theorem 3.11. *The number in a stationary symmetric queue is a reversible stochastic process.*

This property, like quasi-reversibility, is lost when the queue is part of a network of queues.

A major difference between the symmetric queues considered in this section and the queues considered in Section 3.1 is that at a symmetric queue the service requirement of a customer can depend upon his class. In a network of quasi-reversible queues this has two important consequences which we shall explore further in the next chapter. First, a customer's service requirement at a symmetric queue can depend upon the queues he has previously visited and the queues he has yet to visit. This follows naturally since his class depends upon his type, which determines his route through the queues of the network. Hence if the type of a customer in a symmetric queue is unknown then his future route can depend upon his service requirement at that queue. Second, given a customer's route through the queues of the network, his service requirements at symmetric queues along that route may be dependent. This will happen when a variety of customer types correspond to the same route through the system but to different service requirements at the symmetric queues along that route. Indeed, by using enough customer types it is possible to approximate as closely as we please any desired pattern of dependence between the service requirements at the symmetric queues along a route (Exercise 3.3.12) and between these service requirements and the route itself. This strongly suggests that arbitrary patterns of dependence can be allowed, but once again to establish this would take us beyond the realm of countable state space Markov processes.

Although symmetric queues are not the only quasi-reversible queues whose operation involves arbitrary distributions (see, for example, Exercise 3.5.11), they form a class large enough to include all the special cases of such queues which we shall need in the next chapter when we discuss examples of queueing networks.

Exercises 3.3

1. The queue with no waiting room described above has the property that when a customer arrives to find all the servers occupied the servers pause momentarily until he leaves. Redefine the functions ϕ and γ so that this does not occur.
2. Show how the functions ϕ and γ can be chosen to represent a server-sharing queue with a maximum size of N, so that customers who arrive when N customers are being served are turned away.
3. Prove Lemma 3.9, using

$$F_m(x) = \sum_{k=1}^{\infty} \left[F\left(\frac{k}{m}\right) - F\left(\frac{k-1}{m}\right)\right] G_m^k(x) \qquad x \geq 0$$

   where
   (i) G_m^k is the distribution function of a gamma distribution with mean k/m and variance k/m^2 (i.e. the distribution function of the sum of k exponential random variables each with mean $1/m$),
   (ii) G_m^k is the distribution function of a gamma distribution with mean $g(k, m)$ and variance $g(k, m)/m$ where

$$g(k, m) = \int_{(k-1)/m}^{k/m} \frac{x \, dF(x)}{F(k/m) - F((k-1)/m)}$$

   Show that in this case the mean of F_m is equal to the mean of F.

4. Show that $F(x) = F^*(x)$ if and only if F is the exponential distribution.
5. If $F_c(x)$, $c \in \mathscr{C}$, correspond to mixtures of gamma distributions establish property (iv) of Theorem 3.10 directly from the equilibrium distribution (3.18).
6. Show that the service requirement yet to be received by a customer in a symmetric queue has mean
$$\frac{\mu^2 + \sigma^2}{2\mu}$$
where μ and σ^2 are the mean and variance respectively of the service requirement distribution.
7. Suppose that an $M/G/1$ queue has a queue discipline which allows it to be considered as a symmetric queue, e.g. last come first served with preemption or server sharing. Let W be the service requirement yet to be received summed over all the customers in the queue. Show that the distribution of W is the same as that of a geometric sum of independent random variables each with distribution function F^*. Use the previous exercise to deduce that
$$E(W) = \frac{\nu(\mu^2 + \sigma^2)}{2(1 - \nu\mu)} \tag{3.20}$$
where ν is the arrival rate at the queue. Observe that the distribution of W does not depend upon the queue discipline. If the queue discipline is first come first served W is called the virtual waiting time and is the time a typical customer would have to queue to before his service started; expression (3.20) is known as the Pollaczek–Khinchin formula.
8. Show that the examples of symmetric queues given in this section can each be constructed from the queues of Section 3.1. Begin by showing that a service requirement with a gamma distribution can be obtained by requiring a customer to pass through the same queue a fixed number of times before moving on to the next queue.
9. Show that the waiting time of a customer at a stack has the same distribution as the busy period in an $M/G/1$ queue.
10. Show that at either the server-sharing queue or the queue with no waiting room the order of the customers in the queue is independent of the order of arrival of these customers. Suppose now that customers of class c require an amount of service which is exponentially distributed with mean $1/\lambda(c)$. If the queue contains n customers show that the probability the customers arrived in a given order is
$$\prod_{l=1}^{n} \frac{\lambda(c(l))}{\sum_{m=1}^{l} \lambda(c(m))}$$
where $c(l)$, $l = 1, 2, \ldots, n$, is the class of the customer who is lth in the given order.

11. Suppose that in the telephone exchange model considered in Sections 1.3 and 2.1 call lengths are arbitrarily distributed. Show that the points in time at which a call is lost or is completed form a Poisson process. Show that the points in time at which a call is lost form a reversible point process. Establish the result stated in Exercise 2.1.4.
12. Let $F(x_1, x_2, \dots, x_n)$ be the joint distribution function of n positive random variables. Show that if

$$\begin{aligned} F_m(x_1, x_2, \dots, x_n) &= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \cdots \sum_{k_n=1}^{\infty} G_m^{k_1}(x_1) G_m^{k_2}(x_2) \cdots G_m^{k_n}(x_n) \\ &\quad \times \int_{(k_1-1)/m}^{k_1/m} \int_{(k_2-1)/m}^{k_2/m} \cdots \int_{(k_n-1)/m}^{k_n/m} \mathrm{d}F(z_1, z_2, \dots, z_n) \end{aligned}$$

where G_m^k is defined as in either part (i) or (ii) of Exercise 3.3.3 then

$$\lim_{m\to\infty} F_m(x_1, x_2, \dots, x_n) = F(x_1, x_2, \dots, x_n)$$

for all $(x_1, x_2, \dots, x_n)$ at which F is continuous.
13. Consider a symmetric queue within an open network of quasi-reversible queues. Suppose that customers enter the system at rate ν and that the mean service requirement of a customer at the symmetric queue, summed over all the customer's visits to the queue and averaged over all customer types, is a. Show that the equilibrium distribution for the number in the symmetric queue is just what it would be if that queue were in isolation and customers with mean service requirement a arrived in a Poisson stream at rate ν.
14. Consider a symmetric queue within an open network of quasi-reversible queues. Deduce from Theorem 3.8(ii) that the mean number of class c customers in the symmetric queue is proportional to $\nu(c)a(c)$. By supposing that the classification c is fine enough to determine the customer's service requirement deduce from Theorem 3.10 and Little's result that the mean period a customer spends in the symmetric queue is proportional to his service requirement there.
15. Show that for a stack the function $\phi(n)$ can be replaced by a function $\phi(c(1), c(2), \dots, c(n))$ without destroying the property of quasi-reversibility.

3.4 CLOSED NETWORKS

In the open networks considered previously in this chapter customers of type i entered the system in a Poisson stream at rate $\nu(i)$ and passed through the sequence of queues

$$r(i, 1), r(i, 2), \dots, r(i, S(i))$$

before leaving the system. Suppose now that customers of type i return to queue $r(i, 1)$ after leaving queue $r(i, S(i))$ and repeat their route through the system. The network will become closed, with customers neither entering nor leaving the system. The number of customers of type i within the system, $N(i)$ say, will remain fixed for $i = 1, 2, \ldots, I$.

If the queues of the system would in isolation be quasi-reversible the equilibrium distribution takes a simple form. Let $\pi_j(\mathbf{x}_j)$ again be the equilibrium distribution of the jth queue in the network if it were in isolation, with arrivals of customers of class (i, s) forming a Poisson process of rate $\alpha_j(i, s)$. Let

$$\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) = B\pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2) \cdots \pi_J(\mathbf{x}_J) \tag{3.21}$$

where B is chosen so that the distribution π sums to unity, with the sum taken over $\mathcal{S}$, the set of states $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ for which the total number of type i customers is $N(i)$, for $i = 1, 2, \ldots, I$. It is fairly easy to check that π is the equilibrium distribution for the system and that the reversed process consists of the reversed queues with customers moving backwards around their routes. Indeed we have done most of the work already in Section 3.2 when dealing with open networks. The relations established there for transitions arising from the movement of a customer from one queue to another (equation 3.13) or from internal changes in a queue (equation 3.9) apply here also. The movement of a type i customer from queue $r(i, S(i))$ to queue $r(i, 1)$ can be dealt with in precisely the same way as a movement from queue $r(i, s)$ to $r(i, s+1)$. Finally,

$$\begin{aligned} q(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) &= \sum_{j=1}^{J} \left(q_j(\mathbf{x}_j) - \sum_{(i,s)} \alpha_j(i,s) \right) \\ &= \sum_{j=1}^{J} \left(q_j'(\mathbf{x}_j) - \sum_{(i,s)} \alpha_j(i,s) \right) \\ &= q'(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) \end{aligned} \tag{3.22}$$

and so Theorem 1.13 establishes the desired result.

Some comments on the distribution (3.21) are in order. First, since $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ is constrained to lie in the set $\mathcal{S}$ it does not follow from (3.21) that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J$ are independent. Second, the distributions $\pi_1(\mathbf{x}_1), \pi_2(\mathbf{x}_2), \ldots, \pi_J(\mathbf{x}_J)$ depend upon $\nu(i)$, $i = 1, 2, \ldots, I$, yet the distribution $\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ cannot since in a closed network the values of these parameters do not affect the process. The resolution of this apparent contradiction lies in the role of the normalizing constant B. If $\nu(i)$ is changed then $\pi_1(\mathbf{x}_1), \pi_2(\mathbf{x}_2), \ldots, \pi_J(\mathbf{x}_J)$ do indeed change, but so does B, and B changes in such a way that $\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ remains unaltered (cf. the discussion following Theorem 2.3).

In an open network we represented the behaviour of a customer with a random route by using a set of types, one of which was allocated to him at random (see Fig. 3.2). In the closed network just described a customer's type never changes, and this prevents us from modelling the network illustrated in Fig. 3.3 where a customer leaving queue 2 chooses at random whether to go to queue 1 or queue 3. We shall solve this problem by allowing a customer's type to change randomly in a closed network. Thus for the network in Fig. 3.3 suppose there are two customer types, and suppose a customer of type 1 follows the route 1, 2 and a customer of type 2 the route 3, 2. If we allow a customer leaving queue 2 to choose his type at random from the set $\{1, 2\}$ then customers will behave in the required way.

More generally, suppose the set of types $\{1, 2, \ldots, I\}$ is divided into disjoint subsets $\mathscr{I}(1), \mathscr{I}(2), \ldots$, and that on leaving queue $r(i, S(i))$ a customer of type $i \in \mathscr{I}(m)$ becomes a customer of type $i' \in \mathscr{I}(m)$ with probability

$$\frac{\nu(i')}{\sum_{i \in \mathscr{I}(m)} \nu(i)} \tag{3.23}$$

He then proceeds through queues $r(i', 1), r(i', 2), \ldots, r(i', S(i'))$ before rechoosing his type again. His type will always belong to the set $\mathscr{I}(m)$ and thus

$$M(m) = \sum_{i \in \mathscr{I}(m)} N(i)$$

will remain constant for $m = 1, 2, \ldots$. The above mechanism will allow very general routing schemes and certainly those which can arise in closed migration processes (Exercise 3.4.1). It would be possible to regard m as the type of a customer and $i \in \mathscr{I}(m)$ as a finer indication of his progress: with this terminology the type of a customer would not keep changing. We prefer to call i the type since for open networks at least it is helpful to have the route of a customer determined by his type. Observe that in a closed network the points in time at which a customer rechooses his type can be regarded as regeneration points for him.

Consider then a system $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ with the above routing mechanism and containing queues which would in isolation be quasi-reversible. If the process $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ is in equilibrium then call $\mathbf{X}$ a closed network of quasi-reversible queues. An obvious possibility for the equilibrium distribution is

$$\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) = B(M(1), M(2), \ldots)\pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2) \cdots \pi_J(\mathbf{x}_J) \tag{3.24}$$

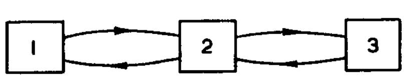

Fig. 3.3 A closed network

where the normalizing constant $B(M(1), M(2), \ldots)$ is chosen so that the distribution π sums to unity, with the sum taken over the state space

$$\mathscr{S}(M(1), M(2), \ldots) = \left\{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J) \,\middle|\, \sum_{i \in \mathscr{I}(m)} N(i) = M(m), \qquad m = 1, 2, \ldots\right\}$$

We shall use Theorem 1.13 to check that (3.24) is the equilibrium distribution and that the reversed process consists of the reversed queues with customers moving backwards around their routes and changing from type i to type i' after leaving queue $r(i, 1)$ with probability (3.23). Consider the transition which arises when a customer of type i leaves queue $j = r(i, S(i))$, changes to type i', and enters queue $k = r(i', 1)$. To establish condition (1.28) for this transition we must show that

$$\pi_j(\mathbf{x}_j')\pi_k(\mathbf{x}_k)q_j(\mathbf{x}_j', \mathbf{x}_j)\frac{q_k(\mathbf{x}_k, \mathbf{x}_k')}{\alpha_k(i', 1)}\frac{\nu(i')}{\sum_{i\in\mathscr{I}(m)}\nu(i)}$$

$$= \pi_j(\mathbf{x}_j)\pi_k(\mathbf{x}_k')q_k'(\mathbf{x}_k', \mathbf{x}_k)\frac{q_j'(\mathbf{x}_j, \mathbf{x}_j')}{\alpha_j(i, S(i))}\frac{\nu(i)}{\sum_{i\in\mathscr{I}(m)}\nu(i)} \tag{3.25}$$

All but the final terms in equation (3.25) arise in the same way as did the corresponding terms in equation (3.13). But since $\alpha_k(i', 1) = \nu(i')$ and $\alpha_j(i, S(i)) = \nu(i)$, equation (3.25) reduces to equation (3.13) which has already been established. For all the other transitions arising in a closed network condition (1.28) takes the same form as for the corresponding transitions in an open network and has thus already been established. Finally, condition (1.27) follows from equation (3.22). Hence the equilibrium distribution and the reversed process are of the suggested form.

Consider now the instant at which a customer of class (i, s), $i \in \mathscr{I}(m)$, leaves queue $j = r(i, s)$. Let $\mathbf{x}_k$, for $k \neq j$, be the state of queue k immediately before this instant. Call $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ the disposition of the other customers in the system; note that it is a member of the set $\mathscr{S}(M(1), M(2), \ldots, M(m)-1, \ldots)$, since one customer of type $i \in \mathscr{I}(m)$ is not included in the description $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$. The probability flux that a customer of class (i, s) departs from queue $j = r(i, s)$ with $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ the disposition of the other customers in the system is

$$B(M(1), M(2), \ldots)\left(\prod_{k\neq j}\pi_k(\mathbf{x}_k)\right)\sum_{\mathbf{x}'\in\mathscr{S}_j(s,i,\mathbf{x}_j)}\pi_j(\mathbf{x}')q_j(\mathbf{x}', \mathbf{x}_j)$$

$$= B(M(1), M(2), \ldots)\pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2)\cdots\pi_J(\mathbf{x}_J)\alpha_j(i, s)$$

from equations (3.8) and (3.9). Thus if a customer of class (i, s) has just left queue j the probability that the disposition of the other customers in the system is $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ is proportional to $\pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2)\cdots\pi_J(\mathbf{x}_J)$. The constant of proportionality is found by summing over the set

$\mathcal{S}(M(1), M(2), \ldots, M(m)-1, \ldots)$ and is hence $B(M(1), M(2), \ldots, M(m)-1, \ldots)$. This is an intriguing result: the disposition of the other customers in the system is distributed in accordance with the equilibrium distribution which would obtain if they were the only customers in the system. Consideration of the reversed process shows that the same statement is valid at the instant when a customer of class (i, s) *arrives* at queue $j = r(i, s)$.

We can summarize the results of this section as follows.

Theorem 3.12. *A closed network of quasi-reversible queues has the following properties:*

(i) *The equilibrium distribution is of the form (3.24).*
(ii) *Under time reversal the system becomes another closed network of quasi-reversible queues.*
(iii) *When a customer of a given class arrives at a queue the disposition of the other customers in the system is distributed in accordance with the equilibrium distribution which would obtain if they were the only customers in the system.*

If a closed network of quasi-reversible queues contains symmetric queues then by using various customer types as described in Section 3.3 it is possible to allow dependences between a customer's service requirements at the symmetric queues he visits and between these service requirements and his route. We shall discuss this point further in Section 4.2, where it can be illustrated with some simple examples.

Exercises 3.4

1. Consider a closed migration process with transition rates (2.1) where $\lambda_j = 1$, $j = 1, 2, \ldots, J$. Show that the process can be regarded as a closed queueing network with customers whose route will be $r(i, 1), r(i, 2), \ldots, r(i, S(i))$, where $r(i, 1) = 1$, $r(i, s) \neq 1$ for $s \neq 1$, having the parameter

$$\nu(i) = \lambda_{r(i,1),r(i,2)}\lambda_{r(i,2),r(i,3)} \cdots \lambda_{r(i,S(i)-1),r(i,S(i))}\lambda_{r(i,S(i)),r(i,1)}$$

Show that the quantities $a_1, a_2, \ldots, a_J$ calculated from the queueing network parameters are proportional to the quantities $\alpha_1, \alpha_2, \ldots, \alpha_J$ determined by equations (2.2) from the migration process parameters.

2. Let

$$\pi_j(M_j(1), M_j(2), \ldots) = \sum \pi_j(\mathbf{x}_j)$$

where the summation runs over all $\mathbf{x}_j$ such that queue j contains $M_j(m)$ customers whose type is in the set $\mathcal{I}(m)$, for $m = 1, 2, \ldots$. Show that the equilibrium distribution (3.24) implies that the probability queue j

contains $M_j(m)$ customers of type $i \in \mathscr{I}(m)$, for $m = 1, 2, \ldots$, $j = 1, 2, \ldots, J$, is

$$B(M(1), M(2), \ldots) \prod_{j=1}^{J} \pi_j(M_j(1), M_j(2), \ldots)$$

provided

$$\sum_j M_j(m) = M(m) \qquad \text{for } m = 1, 2, \ldots$$

3. The normalizing constant $B(M(1), M(2), \ldots)$ can be calculated using the generating function method introduced in Exercise 2.3.6. Define the generating functions

$$\Phi_j(z(1), z(2), \ldots) = \sum_m \sum_{M_j(m)} \pi_j(M_j(1), M_j(2), \ldots) z(1)^{M_j(1)} z(2)^{M_j(2)} \ldots$$

$$B(z(1), z(2), \ldots) = \sum_m \sum_{M(m)} \frac{z(1)^{M(1)} z(2)^{M(2)} \ldots}{B(M(1), M(2), \ldots)}$$

Show that

$$B(z(1), z(2), \ldots) = \prod_j \Phi_j(z(1), z(2), \ldots)$$

4. Consider a closed network of quasi-reversible queues containing a symmetric queue. The service requirement of a customer visiting the symmetric queue may well be dependent on his earlier service requirements at this and other symmetric queues, because these may all be related to his type. Use Exercise 3.4.2 to show that the equilibrium distribution for the number of customers in the various queues of the network is insensitive to these dependencies and that it will be unaltered if each customer arriving at the symmetric queue has a service requirement independent of his previous service requirements so long as the mean service requirement of each customer at the queue is unaltered.
5. By considering the probability flux that a customer of class (i, s), $i \in \mathscr{I}(m)$, leaves queue $j = r(i, s)$ show that in equilibrium the mean arrival rate of customers of class (i, s) at queue j is

$$\frac{\alpha_j(i, s) B(M(1), M(2), \ldots, M(m), \ldots)}{B(M(1), M(2), \ldots, M(m)-1, \ldots)}$$

6. Exercise 3.3.14 has a parallel for a closed network. Consider a symmetric queue within a closed network of quasi-reversible queues. Show that for a given customer entering the symmetric queue the mean period he will remain in the queue is proportional to his service requirement there. Observe that, unlike the parallel result for open networks, the constant of proportionality depends on the chosen customer. This is not surprising: when a given customer arrives at a symmetric queue the

disposition of the other customers does not depend on the class of the given customer if the network is open, but it does if the network is closed.

7. A haulage firm has a fleet of N lorries, but at any one time some of these are being overhauled. While there are n lorries available the firm is able to handle an amount $\phi(n)$ of work, measured in miles per day, and this is shared out equally between the n lorries. A lorry requires an overhaul when it has travelled X miles since its last overhaul. An overhaul takes Y days. The quantities X and Y may be random variables; if so these random variables are independent between different lorries, but values relating to the same lorry may be dependent. Derive an expression for the mean amount of work handled per day by the firm.
8. A complex device consists of n main units, all of which must be operative for the normal operation of the device. Each main unit may fail within time h with probability $\lambda h + o(h)$, independently of the other service units and of its previous service life. There are $m + l$ additional units, m of which are active, i.e. may fail with the same probability as the main units, while the remaining l are passive and cannot fail. Failed units are sent for repair and take a mean time μ^{-1} to repair. If some of the main units fail these are replaced by units from the active redundant system, and these in turn by units of the passive redundant system. Describe the system as a closed network of queues and obtain the equilibrium probability that the device is operative.
9. Consider the model of a mining operation described in Section 2.3. Suppose that after machine j has dealt with face i the face must be left for a period until the dust has settled before the next machine can start work on it; let this period have mean X_{ij}. Obtain the equilibrium distribution for the system.
10. Consider again the model of mining operation described in Section 2.3. The model assumed that it took no time for a machine to travel from one face to the next. Suppose now that it takes a mean period Y_{ij} for machine j to travel from face i to the next face. Obtain the equilibrium distribution for the system when $\phi_1 = \phi_2 = \cdots = \phi_J = \phi$.
11. All the queues we have considered so far which convert a Poisson arrival process into a Poisson departure process have the property that the number in the queue is a reversible stochastic process. This property does not follow necessarily, as the following example shows. Consider a closed network with two customers and two queues. Queue 1 would in isolation behave as a first come first served $M/M/1$ queue. Queue 2 would in isolation behave as a queue with one server and no waiting room. The first customer's service requirement at queue 2 is 1. The second customer's service requirement at queue 2 changes in the cycle

$2, 3, \ldots, s, 2, 3, \ldots$. Both customers' service requirements at queue 1 are independent exponentially distributed random variables with mean λ^{-1}. Show that the arrival process at and the departure process from queue 2 are both Poisson. Show that the number of customers in queue 2 is not a reversible stochastic process. Observe that these statements are true for queue 1 as well.

3.5 MORE GENERAL ARRIVAL RATES

In the open networks considered in Section 3.2 the state of the system did not affect the streams of customers entering the system. In this section we shall show that results can be obtained when the rates of arrival at the system are influenced by the state of the system, provided the influence takes a certain fairly restricted form.

It is illuminating to approach these results by consideration of the following queue. Suppose a queue is such that its state is $(n(1), n(2), \ldots)$, where $n(c)$ is the number of customers of class c that it contains, and suppose arrivals of class c customers form a Poisson stream of rate $\nu(c)$, $c \in \mathscr{C}$, where the arrival streams are independent of each other. Write

$$\nu(c)\phi_c(n(1), n(2), \ldots)$$

for the probability intensity that a customer of class c leaves the queue when its state is $(n(1), n(2), \ldots)$.

Lemma 3.13. *If the above queue is in equilibrium the following statements are equivalent:*

(i) *The process $(n(1), n(2), \ldots)$ is reversible.*
(ii) *The queue is quasi-reversible.*
(iii) *There exists a function $\Phi(n(1), n(2), \ldots)$ such that*

$$\Phi(n(1), n(2), \ldots, n(c), \ldots) = \phi_c(n(1), n(2), \ldots, n(c), \ldots) \times \Phi(n(1), n(2), \ldots, n(c)-1, \ldots) \qquad (3.26)$$

Proof. Let

$$\pi(n(1), n(2), \ldots) = \frac{b}{\Phi(n(1), n(2), \ldots)}$$

If relation (3.26) holds then π satisfies the detailed balance conditions; hence statement (iii) implies statement (i). Conversely, if the process is reversible the equilibrium distribution $\pi(n(1), n(2), \ldots)$ satisfies the detailed balance condition

$$\pi(n(1), n(2), \ldots, n(c)-1, \ldots)\nu(c) = \pi(n(1), n(2), \ldots, n(c), \ldots) \times \nu(c)\phi_c(n(1), n(2), \ldots, n(c), \ldots)$$

and hence

$$\Phi(n(1), n(2), \ldots) = \pi(n(1), n(2), \ldots)^{-1}$$

satisfies the relation (3.26). Thus statement (i) implies statement (iii).

Statement (i) immediately implies statement (ii). Conversely, if the queue is quasi-reversible then equation (3.11) shows that the detailed balance condition is satisfied, and thus statement (ii) implies statement (i).

Consider now a closed network in which customers do not change type, of the sort discussed early in the previous section. Suppose that a quasi-reversible queue of the above form is appended to the network as queue 0, and that a visit to this queue is added as an extra stage at the beginning of the route of each customer; thus $r(i, 0) = 0$, for $i = 1, 2, \ldots$. Let $N(i)$ be the number of customers of type i in queues $1, 2, \ldots, J$ and $N^*(i)$ the number in queues $0, 1, 2, \ldots, J$. Thus $N^*(i) = N(i) + n(i)$. Now define a function Ψ by

$$\Psi(N(1), N(2), \ldots) = \begin{cases} [\Phi(n(1), n(2), \ldots)]^{-1} & \text{if } N(i) \leq N^*(i),\ i = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $n(i) = N^*(i) - N(i)$. If we confine our attention to queues $1, 2, \ldots, J$ how does the system behave? If these queues contain between them $N(i)$ customers of type i, for $i = 1, 2, \ldots$, then the probability intensity a customer of type i leaves queue 0 and enters the system is

$$\nu(i) \frac{\Phi(n(1), n(2), \ldots, n(i), \ldots)}{\Phi(n(1), n(2), \ldots, n(i)-1, \ldots)} = \nu(i) \frac{\Psi(N(1), N(2), \ldots, N(i)+1, \ldots)}{\Psi(N(1), N(2), \ldots, N(i), \ldots)} \tag{3.27}$$

The equilibrium distribution for the system is, by Theorem 3.12, of the form

$$B \frac{1}{\Phi(n(1), n(2), \ldots)} \pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2) \cdots \pi_J(\mathbf{x}_J) = B\Psi(N(1), N(2), \ldots)\pi_1(\mathbf{x}_1)\pi_2(\mathbf{x}_2) \cdots \pi_J(\mathbf{x}_J) \tag{3.28}$$

The above discussion strongly suggests that in an open network of queues if we relax the assumption of Poisson arrival streams, and suppose instead that the probability intensity a customer of type i arrives takes the form (3.27), then the equilibrium distribution is given by expression (3.28). Indeed the discussion proves this assertion if there exist numbers $N^*(1), N^*(2), \ldots$ such that $\Psi(N(1), N(2), \ldots) = 0$ if $N(i) > N^*(i)$. This restriction can be removed by proving the assertion directly.

Theorem 3.14. *If in a stationary network of quasi-reversible queues the* *arrival rates at the system take the form (3.27) then the equilibrium distribution is given by expression (3.28).*

Proof. The reversed process might consist of the reversed queues with customers of type i entering the system at rate (3.27) and moving backwards through their route. We can use Theorem 1.13 to check that the reversed process does indeed take this form and that (3.28) is the equilibrium distribution. The conditions of Theorem 1.13 are readily verified; they take the same form as before apart from a straightforward embellishment for transitions caused by a customer entering or leaving the system.

To illustrate how we might use Theorem 3.14 let $\mathcal{I}(1), \mathcal{I}(2), \ldots$ be sets of customer types (not necessarily disjoint) and let

$$N(\mathcal{I}) = \sum_{i \in \mathcal{I}} N(i)$$

If Ψ is of the form

$$\Psi(N(1), N(2), \ldots) = \prod_{\mathcal{I}} \prod_{N=0}^{N(\mathcal{I})-1} \psi_{\mathcal{I}}(N)$$

then the arrival rate of customers of type i is

$$\prod_{\mathcal{I}: i \in \mathcal{I}} \psi_{\mathcal{I}}(N(\mathcal{I}))$$

If $\psi_{\mathcal{I}} \equiv 1$ for all $\mathcal{I}$ then customers of type i arrive at the system in a Poisson stream at rate $\nu(i)$, and we have the ordinary open network of queues. By varying the function $\psi_{\{i\}}$ we can make the rate of arrival of customers of type i depend upon the number of such customers already in the system. If $\mathcal{I}$ corresponds to the set of types used to represent a customer whose route is random then it may be appropriate to use the function $\psi_{\mathcal{I}}$ rather than the functions $\psi_{\{i\}}$, $i \in \mathcal{I}$. Of course the function $\psi_{\{1,2,\ldots\}}$ enables the rate of arrival of all customer types to be affected by the total number of customers in the system; for example if we let $\psi_{\{1,2,\ldots\}}(N^*) = 0$ then the system will saturate when the number of customers in it reaches N^*.

In spite of these examples the arrival rates allowed by Theorem 3.14 are of a fairly restricted form. They can depend upon the number of customers of each type in the system but not upon the position within the network of these customers. Further, the dependence upon the number of customers of each type must be expressible in terms of the function Ψ.

We shall end this section with a simple example to illustrate Theorem 3.14.

A repair shop. Consider a repair shop that accepts two types of job. The shop employs K repairmen altogether, of whom K_1 can deal with jobs of type 1, K_2 can deal with jobs of type 2, and the remainder, K_3, can deal with both types of job. The shop is interested in choosing the best arrangement (K_1, K_2, K_3). Suppose that jobs of type 1 and 2 arrive in independent Poisson streams of states ν_1 and ν_2 respectively. When a job arrives it is accepted if there is a repairman free who can deal with it; otherwise the job is lost. Jobs being dealt with can be reshuffled amongst the repairmen if this allows an extra job to be accepted. The time taken to deal with a job of type i is arbitrarily distributed with mean μ_i^{-1} and is not affected by any reshuffling that may be necessary.

Let n_i be the number of jobs of type i in the shop and let

$$\Psi(n_1, n_2) = \begin{cases} 1 & n_1 \le K_1 + K_3,\ n_2 \le K_2 + K_3,\ n_1 + n_2 \le K_1 + K_2 + K_3 \\ 0 & \text{otherwise} \end{cases}$$

The rules we have specified above imply that the probability intensity a job of type i will arrive and be accepted when n_j jobs of type j are already there is

$$\nu_1 \frac{\Psi(n_1+1, n_2)}{\Psi(n_1, n_2)} \qquad i=1$$

and

$$\nu_2 \frac{\Psi(n_1, n_2+1)}{\Psi(n_1, n_2)} \qquad i=2$$

The entire system thus behaves as an infinite-server queue with two types of customer whose arrival pattern is of the form (3.27). Theorem 3.14 thus shows that in equilibrium

$$\pi(n_1, n_2) = B\Psi(n_1, n_2) \frac{1}{n_1!}\left(\frac{\nu_1}{\mu_1}\right)^{n_1} \frac{1}{n_2!}\left(\frac{\nu_2}{\mu_2}\right)^{n_2}$$

Exercise 3.5.5 shows that the system appropriately augmented is quasi-reversible and that, counting lost jobs, the departure streams formed by jobs of types 1 and 2 are independent and Poisson.

Exercises 3.5

1. Write down Kolmogorov's criteria for the queue considered in Lemma 3.13. Observe that these conditions on the functions $\phi_c(n(1), n(2), \ldots)$ are equivalent to the existence of a function Φ satisfying equation (3.26).
2. Choose a function Ψ so that the queues considered in Exercise 1.6.1 have arrival rates of the form (3.27).
3. Show that if Ψ takes the value zero unless $N(\mathscr{I}) = N^*(\mathscr{I})$ for each $\mathscr{I}$ then the resulting network is equivalent to the closed network considered in Section 3.4. Observe that an appropriate choice of Ψ produces

a mixed system in which certain types of customer can enter and leave the system while customers of other types can do neither.

4. Show that when a customer of type i arrives at queue $j = r(i, s)$ at stage s of his route the probability that the disposition of the other customers in the system is $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$ is proportional to
$$\Psi(N(1), N(2), \ldots, N(i)+1, \ldots)\pi(\mathbf{x}_1)\pi(\mathbf{x}_2) \cdots \pi(\mathbf{x}_J)$$
where $(N(1), N(2), \ldots, N(i), \ldots)$ is calculated from the disposition $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J)$.
5. Suppose that expression (3.27) takes values not greater than $\nu(i)$. Then the apparent arrival rate (3.27) will arise if customers of type i arrive at the system in a Poisson stream of rate $\nu(i)$ but a customer of type i is lost with probability
$$1 - \frac{\Psi(N(1), N(2), \ldots, N(i)+1, \ldots)}{\Psi(N(1), N(2), \ldots, N(i), \ldots)}$$
Show that if I is finite and the state of the system is augmented so that it signals when customers are lost, then the resulting process is quasi-reversible with the class of a customer given by his type.
6. Generalize the model of a repair shop to allow I types of job, with each repairman able to deal with a subset of the I types.
7. Amend the queue considered in Lemma 3.13 so that service requirements at it can be arbitrarily distributed. Show that Lemma 3.13 still holds.
8. Let $\mathbf{x}$ be a quasi-reversible process with equilibrium distribution $\pi(\mathbf{x})$. Suppose now that all transition rates $q(\mathbf{x}, \mathbf{x}')$ which do not correspond to the arrival of a customer are multiplied by $\phi(N)$, where N is the number of customers in the system in state $\mathbf{x}$. Deduce that the resulting system is quasi-reversible with equilibrium distribution
$$B \frac{\pi(\mathbf{x})}{\prod_{l=1}^{N} \phi(l)}$$
either from the equilibrium equations and the partial balance equations (3.11) for the process $\mathbf{x}$, or from Theorem 3.14 and a dilation of the time scale.
9. There are quite subtle ways in which a system with the general arrival rates discussed in this section can be rendered quasi-reversible. Consider a system with three types of customer and with arrival rates of the form (3.27) where $\Psi(0,0,0) = \Psi(1,0,0) = \Psi(0,0,1) = \Psi(0,1,1) = 1$, and $\Psi(N(1), N(2), N(3)) = 0$ for other arguments. Then the apparent arrival rate (3.27) could arise as described in Exercise 3.5.5, but if $\nu(1) = \nu(2)$ it could also arise in the following way. Customers of classes 1 and 2 arrive in independent Poisson streams of rates $\nu(1)$ and $\nu(3)$

respectively. When a customer of class 2 arrives he is lost unless the system is empty, in which case he enters it as a customer of type 3. If when a customer of class 1 arrives the system is empty he enters the system as a customer of type 1; if the system contains one customer of type 3 he enters the system as a customer of type 2; otherwise he is lost. Show that, counting lost customers, this system is quasi-reversible with respect to the classes 1 and 2, but not with respect to the types 1, 2, and 3.

10. Consider the following loss priority queueing system with one server. Customers of classes 1 and 2 arrive at a single server in independent Poisson streams of rates ν_1 and ν_2. Customers of class 1 have a higher priority and an arriving customer of class 1 interrupts the service of a customer of class 2; the interrupted service is resumed when the customer of class 1 has his service completed. An arriving customer is lost if a customer of the same or higher priority is being served at the time of arrival. Service times of customers of class i are arbitrarily distributed with mean μ_i^{-1}, $i=1,2$. Show that this system arises when the system considered in the previous exercise contains just a stack. Deduce that the equilibrium probability the system is empty is

$$\frac{\mu_1\mu_2}{(\nu_1+\mu_1)(\nu_2+\mu_2)}$$

Generalize the model to deal with I priority classes.

11. Suppose that the system described in the preceding exercise is amended in the following way: the service time of a customer of type 1 depends upon whether or not he is the only customer in the system. If he is it is arbitrarily distributed with mean μ_1^{-1}; if he has interrupted the service of a customer of type 2 then his service time is arbitrarily distributed with mean μ^{-1}. Show that the equilibrium probability the system is empty is now

$$\frac{\mu_1\mu_2\mu}{\mu(\mu_1\mu_2+\nu_1\mu_2+\nu_2\mu_1)+\nu_1\nu_2\mu_1}$$