2WB40 Queueing Systems
Networks of queues

Céline Comte and Jacques Resing

Eindhoven University of Technology

April 19, 2021

In this chapter, we will consider networks of queues, in which each customer may receive service at several queues in a row. Each queue is similar to the M/M/1 queue introduced in Chapter 4. The only difference is that the input process is not Poisson in general, as each queue may receive customers from other queues. This chapter gives you a glimpse of the topics covered in the course 2MMS40 *Stochastic Networks*. We will make a distinction between two types of networks called *open* and *closed*.

1 Open Jackson networks

We first focus on open networks of queues, in which customers arrive from the outside and eventually leave the system.

1.1 Model description

Consider a network of k queues. Queue i has a single server and service times are exponential with mean $1/\mu_i$, for each $i = 1, \dots, k$. Customers enter queue i from outside according to a Poisson process with rate ν_i. Taking $\nu_i = 0$ means that there are no exogenous arrivals at queue i, and for now we only require that $\sum_{i=1}^k \nu_i > 0$. Customers can also move from queue to queue upon service completion. More specifically, for each $i = 1, \dots, k$, when a customer completes service at queue i, this customer enters queue j to receive service with probability $q_{i,j}$, for each $j = 1, \dots, k$, or leaves the system with probability $q_i = 1 - \sum_{j=1}^k q_{i,j}$. This is called *Markov routing*, as the routing is described by a Markov chain. We assume that the exogenous arrival rates and routing probabilities satisfy the following technical assumptions, often called *irreducibility* assumptions: each queue is visited with a positive probability by some customers, and each customer eventually leaves the system with probability one. We also assume that the service times of customers at different queues are independent. This network is called an *open Jackson network*.

The *effective* arrival rates $\lambda_1, \lambda_2, \dots, \lambda_k$ are defined by the following *traffic* equations:

$$\lambda_i = \nu_i + \sum_{j=1}^k \lambda_j q_{j,i}, \quad i = 1, \dots, k. \tag{1}$$

As the name suggests, these effective arrival rates will give the rates at which customers enter the queues (either from outside or from another queue) when the system is in equilibrium. The traffic equations simply mean that the arrival process at queue i is the superposition of two processes, the exogenous arrival process (Poisson with rate ν_i) and the internal routing process. In matrix form, the traffic equations can be rewritten as $\lambda = \nu + \lambda Q$, where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ and $\nu = (\nu_1, \nu_2, \dots, \nu_k)$ are row vectors and $Q = (q_{i,j})_{i,j=1,\dots,k}$ is the internal (sub-stochastic) routing matrix. The irreducibility assumptions guarantee that the matrix $I - Q$ (where I is the identity matrix of dimension k) is invertible, and that the traffic equations have a unique solution given by $\lambda = \nu(I - Q)^{-1}$. The proof of this statement, which consists of interpreting the traffic equations as the equilibrium equations of a discrete-time Markov chain, is postponed until Section 3. Summing the traffic equations also yields

$$\sum_{i=1}^{k} \nu_i = \sum_{i=1}^{k} \lambda_i q_i,$$

meaning that, in equilibrium, the overall arrival and departure rates are equal to each other.

Similarly to the single-queue model of Chapter 4, we require that

$$\rho_i = \frac{\lambda_i}{\mu_i} < 1, \quad i = 1, \dots, k,$$

since, otherwise, the length of at least one of the queues will grow without bound. We will see later that, under this condition, the quantity ρ_i is the fraction of time that the server at queue i is working, that is, the occupation rate of queue i.

Example 1 (Two M/M/1 queues in series)*.* Consider an open tandem network of $k = 2$ queues. The external arrival rates are $\nu_1 > 0$ and $\nu_2 = 0$. The Markov routing is defined by $q_{1,2} = q_2 = 1$, so that all customers visit first queue 1 and then queue 2 before leaving the system. The corresponding network of queues is shown in Figure 1. The traffic equations (1) simplify as $\lambda_1 = \nu_1$ and $\lambda_2 = \lambda_1$, which yields directly $\lambda_1 = \lambda_2 = \nu_1$. The stability condition is $\nu_1 < \mu_1$ and $\nu_1 < \mu_2$. If the first condition is not satisfied, the length of the first queue will grow without bound (and so may do the length of the second queue); if the second condition is not satisfied, the length of the second queue will grow without bound.

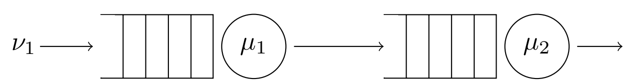

Figure 1: Two M/M/1 queues in series

If you take the course 2MMS40, you will see that, according to Burke's theorem, the output process of a stable M/M/1 queue is a Poisson process with the same rate as the input process. In our example, this implies that the output process of the first queue is a Poisson process with rate ν_1, so that the second queue is again an M/M/1 queue. The stationary behavior of each queue can therefore be analyzed in the same way as in Chapter 4. This simple yet insightful example motivated the first analysis of Jackson networks.

Example 2 (Open network of two queues)*.* Again consider an open network of $k = 2$ queues. The exogenous rates are $\nu_1 = 1$ and $\nu_2 = 2$. The Markov routing is given by $q_{1,1} = \frac{1}{4}$, $q_{1,2} = \frac{1}{2}$, and $q_{2,1} = q_{2,2} = \frac{1}{3}$. We also let $\mu_1 = \mu_2 = 8$. This network is shown in Figure 2. The traffic equations are

$$\lambda_1 = 1 + \frac{1}{4}\lambda_1 + \frac{1}{3}\lambda_2, \qquad \lambda_2 = 2 + \frac{1}{2}\lambda_1 + \frac{1}{3}\lambda_2.$$

Solving these equations yields $\lambda_1 = 4$ and $\lambda_2 = 6$. The system is stable, and the occupation rates of the queues are $\rho_1 = \frac{1}{2}$ and $\rho_2 = \frac{3}{4}$.

1.2 Equilibrium distribution

The network state is described by the vector $n = (n_1, n_2, \dots, n_k)$, where n_i is the number of customers at queue i, for each $i = 1, \dots k$. The set of possible values of this vector is $\mathcal{S} = \{0, 1, 2, \dots\}^k$. The memoryless property of the interarrival and service times guarantees that the evolution of this vector over time defines a Markov process. For each $i = 1, \dots, k$, we let e_i denote the k-dimensional vector with one in component i and zero elsewhere, corresponding to a network state with a single customer at queue i.

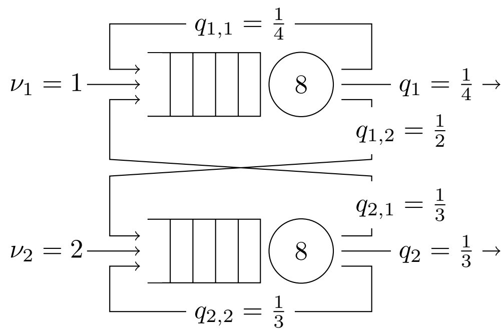

Figure 2: An open network of two queues

Let p_n denote the equilibrium probability that there are n_i customers at queue i, for each $i = 1, \dots, k$. We obtain the following set of equilibrium equations for p_n:

$$\left(\sum_{i=1}^{k} \nu_i + \sum_{\substack{i=1 \\ n_i>0}}^{k} \mu_i \right) p_n = \sum_{\substack{i=1 \\ n_i>0}}^{k} \nu_i p_{n-e_i} + \sum_{\substack{i=1 \\ n_i>0}}^{k} \sum_{j=1}^{k} \mu_j q_{j,i} p_{n-e_i+e_j} + \sum_{j=1}^{k} \mu_j q_j p_{n+e_j}, \quad n \in \{0, 1, 2, \dots\}^k, \tag{2}$$

The first factor on the left-hand side is the rate at which a transition occurs when the network is in state n. The right-hand side is made up of three sums that correspond to the probability flow into state n due to a customer who enters the network from the outside, a customer who enters a queue after a service completion, and a customer who leaves the system after a service completion, respectively.

Using the traffic equations (1), we can verify that the following distribution satisfies the balance equations (2), and is therefore the equilibrium distribution:

$$p_n = \prod_{i=1}^{k} (1 - \rho_i) {\rho_i}^{n_i}, \quad n \in \{0, 1, 2, \dots\}^k. \tag{3}$$

This distribution is said to have a *product form* because it is written as a product of factors. Referring back to Chapter 4, we recognize in these factors the equilibrium distributions of M/M/1 queues with occupation rates $\rho_1, \rho_2, \dots, \rho_k$. Therefore, when the system is in equilibrium, the queue states at any point in time are independent of each other, and their distributions are the same as in M/M/1 queues. This remark also proves that the distribution is indeed normalized.

Example 3 (Open network of two queues). Going back to Example 2, we obtain that the equilibrium distribution of the network state is given by

$$p_n = \left(\frac{1}{2} \right)^{n_1+1} \times \frac{1}{4} \left(\frac{3}{4} \right)^{n_2}, \quad n \in \{0, 1, 2, \dots\}^2.$$

Since the marginal stationary distribution of each queue is that of an M/M/1 queue, the mean numbers of customers at the first and second queues are given by $E(L_1) = \frac{\rho_1}{1-\rho_1} = 1$ and $E(L_2) = \frac{\rho_2}{1-\rho_2} = 3$. It follows from Little's law that the means of the sojourn times S in the network, S_1 at queue 1, and S_2 at queue 2 are given by

$$E(S) = \frac{E(L_1) + E(L_2)}{\nu_1 + \nu_2} = \frac{4}{3}, \qquad E(S_1) = \frac{E(L_1)}{\lambda_1} = \frac{1}{4}, \qquad E(S_2) = \frac{E(L_2)}{\lambda_2} = \frac{1}{2}.$$

These quantities are related by the equation

$$E(S) = \frac{\lambda_1}{\nu_1 + \nu_2} E(S_1) + \frac{\lambda_2}{\nu_1 + \nu_2} E(S_2) = \frac{4}{3} E(S_1) + 2E(S_2).$$

Consistently, one could verify that $\frac{\lambda_i}{\nu_1+\nu_2}$ is the expected number of times that a customer visits queue i, for $i \in \{1, 2\}$.

Remark. Even though the external arrival processes are Poisson and the service time distributions are exponential, the queues do not strictly behave as M/M/1 queues in general because of the additional arrivals coming from other queues. Surprisingly, as we have just seen, the stationary distribution of each queue still corresponds to that of an M/M/1 queue with the same occupancy rate. This explains why the queues of an open Jackson networks are (somewhat sloppily) sometimes referred to as M/M/1 queues in the literature.

1.3 Arrival theorem

Several results about the M/M/1 queue, like the distribution of the sojourn time of a customer and the mean-value analysis, rely on the PASTA property. An analogous property holds in an open Jackson network, namely, the probability that a customer arriving at a queue (either from outside or from another queue) sees the system in state n is equal to the equilibrium probability p_n. Consequently, the results derived in Chapter 4 for the M/M/1 queue still hold for each queue of an open Jackson network.

2 Closed Jackson networks

We now consider a modification of the previous network that is said to be *closed* because customers neither enter nor leave the system. We will mainly focus on highlighting the differences with the open network.

2.1 Model description

We again consider a network of k single-server queues with exponential service times with rates $\mu_1, \mu_2, \dots, \mu_k$. There are no exogenous arrivals and the overall number of customers in the system is denoted by ℓ. For each $i = 1, \dots, k$, when a customer completes service at queue i, this customer enters queue j with probability $q_{i,j}$, for each $j = 1, \dots, k$, with $\sum_{j=1}^{k} q_{i,j} = 1$. We assume that this Markov routing is irreducible in the sense that customers can eventually visit all queues with a positive probability, irrespective of the queue where they are currently. We also assume that the service times of customers at different queues are independent. This network is called a *closed Jackson network*.

The *relative* arrival rates $\lambda_1, \lambda_2, \dots, \lambda_k$ are defined by the traffic equations

$$\lambda_i = \sum_{j=1}^{k} \lambda_j q_{j,i}, \quad i = 1, \dots, k. \tag{4}$$

These equations can be seen as a special case of the traffic equations (1), stated for open networks, with $\nu_i = 0$ for $i = 1, \dots, k$. The traffic equations can be rewritten as $\lambda = \lambda Q$, where $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ is a row vector and $Q = (q_{i,j})_{i,j=1,\dots,k}$ is the (stochastic) routing matrix; this equation means that the vector λ is a left eigenvector of the matrix Q. The irreducibility assumption guarantees that the traffic equations have a unique solution *up to a multiplicative constant*. The proof of this statement, which again consists of interpreting the traffic equations as the equilibrium equations of a discrete-time Markov chain, is postponed until Section 3. For now, we simply define the relative arrival rates $\lambda_1, \lambda_2, \dots, \lambda_k$ as an arbitrary positive solution of these equations. These relative arrival rates only tell you that a fraction $\frac{\lambda_i}{\lambda_1+\cdots+\lambda_N}$ of the arrivals is an arrival at queue i. The *effective* arrival rates, which are also solutions to the traffic equations, will be determined later and depend on the number ℓ of customers in the system and the service rates $\mu_1, \mu_2, \dots, \mu_k$.

We again let $\rho_i = \frac{\lambda_i}{\mu_i}$ for each $i = 1, \dots, k$. The system is always stable because the overall number of customers in the system is a constant. The ratios $\rho_1, \rho_2, \dots, \rho_k$ are equal to the occupation rates only if the relative arrival rates are equal to the effective arrival rates.

Example 4. Consider a closed network of $k = 2$ queues. The Markov routing is given by $q_{1,1} = \frac{1}{3}$, $q_{1,2} = \frac{2}{3}$, and $q_{2,1} = q_{2,2} = \frac{1}{2}$. The service rates are $\mu_1 = \mu_2 = 8$. The traffic equations become

$$\lambda_1 = \frac{1}{3}\lambda_1 + \frac{1}{2}\lambda_2, \qquad \lambda_2 = \frac{2}{3}\lambda_1 + \frac{1}{2}\lambda_2.$$

These equations are linearly dependent and equivalent to $4\lambda_1 = 3\lambda_2$. We can choose for instance $\lambda_1 = 3$ and $\lambda_2 = 4$. The effective arrival rates, which will be denoted by Λ_1 and Λ_2 later in this section, are given by $\Lambda_1 = 3\alpha$ and $\Lambda_2 = 4\alpha$ for some proportionality constant $\alpha > 0$. We have in particular $\frac{\Lambda_1}{\Lambda_2} = \frac{\lambda_1}{\lambda_2}$.

2.2 Equilibrium distribution

We again focus on the differences with open networks. The state is still given by $n = (n_1, n_2, \ldots, n_k)$, where n_i is the number of customers at queue i, for each $i = 1, \ldots, k$. Since the overall number of customers in the network is equal to ℓ, the set of possible values of this vector is

$$\mathcal{S} = \left\{ n \in \{0, 1, 2, \ldots\}^k : n_1 + n_2 + \ldots + n_k = \ell \right\}.$$

The equilibrium equations for the closed network follow by injecting $\nu_i = q_i = 0$ in the equilibrium equations (2) of the open network. Consistently, we can verify that the equilibrium distribution for the closed network is similar to that of the open network, up to a multiplicative constant:

$$p_n = \frac{1}{c} \prod_{i=1}^{k} {\rho_i}^{n_i}, \quad n \in \mathcal{S}. \tag{5}$$

In general, and contrary to open networks, multiplying the term corresponding to queue i by $1 - \rho_i$ is not sufficient to normalize the equilibrium distribution because the state space is smaller. The normalization constant c is instead given by

$$c = \sum_{n \in \mathcal{S}} \prod_{i=1}^{k} {\rho_i}^{n_i}. \tag{6}$$

Since the number $n_1 + n_2 + \ldots + n_k = \ell$ of customers in the network is a constant, multiplying all relative arrival rates by the same (positive) constant a amounts to multiplying (5) and (6) by a^ℓ, so that the resulting equilibrium distribution is unchanged. This explains why the relative arrival rates can be chosen up to a multiplicative constant.

Brute-force use of (6) to calculate the normalization constant c leads to a time complexity that is exponential with the number of queues, which is practically unfeasible even for networks of reasonable size. However, there exist more efficient algorithms, such as Buzen's convolution algorithm that operates with a complexity $O(k\ell)$. Also note that the product form does no longer imply the independence of the queue states because the overall number of customers is constant; for instance, assuming that there are $n_1 = \ell$ customers in the first queue implies that all other queues are empty.

The previous paragraph, and in particular the absence of independence between queues, suggests that the derivation of the performance metrics will be more involved. In the remainder of this section, we will focus on deriving three average performance metrics, namely the effective arrival rates, mean numbers of customers, and mean sojourn times, by applying variants of the PASTA property and mean-value analysis.

2.3 Arrival theorem

A counterpart of the PASTA property holds for closed Jackson networks, but its statement is slightly more intricate than for open Jackson networks. Specifically, the probability that a customer arriving at a queue sees the system in state n is equal to the equilibrium probability that the same network, but with $\ell - 1$ customers instead of ℓ, is in state n. Informally speaking, a customer arriving at a queue sees the equilibrium distribution of the system as if this customer did not exist.

We now use the arrival theorem to design an algorithm, called the mean-value analysis algorithm, to calculate the above-mentioned performance metrics. The use of the phrase *mean-value analysis* in this chapter differs from previously. Indeed, this phrase referred so far to a method, based on Little's law and the PASTA property, to derive formulas for long-term performance metrics in single-server queues with Poisson arrivals; in this chapter, we similarly use Little's law and the above-mentioned arrival theorem to calculate long-term performance metrics, but the calculations are more intricate and lead to a recursive algorithm to calculate performance metrics in closed network of single-server queues by induction on the number of customers in the network.

For each $\ell = 1, 2, \ldots$ and $i = 1, \ldots, k$, we let $L_i(\ell)$ and $S_i(\ell)$ denote (random variables distributed like) the stationary numbers of customers and the sojourn time at queue i in a network with ℓ customers. We also let $\Lambda_i(\ell)$ denote the effective arrival rates at queue i in a network with ℓ customers. We know from Section 2.1 that $\frac{\Lambda_i(\ell)}{\Lambda_j(\ell)} = \frac{\lambda_i}{\lambda_j}$ for each $\ell = 1, 2, \ldots$ and $i, j = 1, \ldots, k$. We have the following sets of equations:

- Arrival relation at each queue (using the arrival theorem):

$$E(S_i(\ell)) = E(L_i(\ell - 1))\frac{1}{\mu_i} + \frac{1}{\mu_i}, \quad i = 1, \ldots, k. \tag{MVA-1}$$

- Little's formula for each queue:

$$E(L_i(\ell)) = \Lambda_i(\ell)E(S_i(\ell)), \quad i = 1, \ldots, k. \tag{MVA-3}$$

- Summing these equations over all queues and using the above-mentioned proportionality relation yields

$$\ell = \sum_{j=1}^{k} E(L_j(\ell)) = \sum_{j=1}^{k} \Lambda_j(\ell)E(S_j(\ell)) = \sum_{j=1}^{k} \Lambda_i(\ell)\frac{\lambda_j}{\lambda_i}E(S_j(\ell)), \quad i = 1, \ldots, k,$$

  so that we obtain

$$\Lambda_i(\ell) = \frac{\ell}{\sum_{j=1}^{k} \lambda_j E(S_j(\ell))}\lambda_i, \quad i = 1, \ldots, k. \tag{MVA-2}$$

The mean-value analysis algorithm consists of applying (MVA-1), (MVA-2), and (MVA-3) (in this order) to derive the values of $E(S_i(\ell))$, $\Lambda_i(\ell)$, and $E(L_i(\ell))$ by recursion over ℓ. More specifically, the recursion is initialized with the base case $E(L_i(0)) = 0$. Then we obtain $E(S_i(1)) = \frac{1}{\mu_i}$ from (MVA-1), $\Lambda_i(1)$ from (MVA-2), $E(L_i(1))$ from (MVA-3), $E(S_i(2))$ from (MVA-1), and so on. The time complexity of this algorithm is $O(k\ell)$.

Example 5. We illustrate the mean-value analysis algorithm on the closed network of $k = 2$ queues introduced in Example 4. We would like to calculate the performance metrics with $\ell = 3$ customers. Recall that, for this network, we have $\lambda_1 = 3$, $\lambda_2 = 4$, and $\mu_1 = \mu_2 = 8$. The recursion below is initialized with the base case $E(L_1(0)) = E(L_2(0)) = 0$. We can verify that, as mentioned earlier, we have $\frac{\Lambda_1(1)}{\Lambda_2(1)} = \frac{\Lambda_1(2)}{\Lambda_2(2)} = \frac{\Lambda_1(3)}{\Lambda_2(3)} = \frac{\lambda_1}{\lambda_2} = \frac{3}{4}$.

Calculations for $\ell = 1$:

(MVA-1)**:** $E(S_1(1)) = \dfrac{1}{\mu_1} = \dfrac{1}{8}$ and $E(S_2(1)) = \dfrac{1}{\mu_2} = \dfrac{1}{8}$

(MVA-2)**:** $\dfrac{\Lambda_1(1)}{\lambda_1} = \dfrac{\Lambda_2(1)}{\lambda_2} = \dfrac{1}{\frac{3}{8} + \frac{4}{8}} = \dfrac{8}{7}$, so that $\Lambda_1(1) = \dfrac{24}{7}$ and $\Lambda_2(1) = \dfrac{32}{7}$

(MVA-3): $E(L_1(1)) = \frac{24}{7}\frac{1}{8} = \frac{3}{7}$ and $E(L_2(1)) = \frac{32}{7}\frac{1}{8} = \frac{4}{7}$

Calculations for $\ell = 2$:

(MVA-1): $E(S_1(2)) = \frac{3}{7}\frac{1}{8} + \frac{1}{8} = \frac{5}{28}$ and $E(S_2(2)) = \frac{4}{7}\frac{1}{8} + \frac{1}{8} = \frac{11}{56}$

(MVA-2): $\frac{\Lambda_1(2)}{\lambda_1} = \frac{\Lambda_2(2)}{\lambda_2} = \frac{2}{3\frac{5}{28} + 4\frac{11}{56}} = \frac{56}{37}$, so that $\Lambda_1(2) = \frac{168}{37}$ and $\Lambda_2(2) = \frac{224}{37}$

(MVA-3): $E(L_1(2)) = \frac{168}{37}\frac{5}{28} = \frac{30}{37}$ and $E(L_2(2)) = \frac{224}{37}\frac{11}{56} = \frac{44}{37}$

Calculations for $\ell = 3$:

(MVA-1): $E(S_1(3)) = \frac{30}{37}\frac{1}{8} + \frac{1}{8} = \frac{67}{296}$ and $E(S_2(3)) = \frac{44}{37}\frac{1}{8} + \frac{1}{8} = \frac{81}{296}$

(MVA-2): $\frac{\Lambda_1(3)}{\lambda_1} = \frac{\Lambda_2(3)}{\lambda_2} = \frac{3}{3\frac{67}{296} + 4\frac{81}{296}} = \frac{296}{175}$, so that $\Lambda_1(3) = \frac{888}{175}$ and $\Lambda_2(3) = \frac{1184}{175}$

(MVA-3): $E(L_1(3)) = \frac{888}{175}\frac{67}{296} = \frac{201}{175}$ and $E(L_2(3)) = \frac{1184}{175}\frac{81}{296} = \frac{324}{175}$

Remark. As observed before, the traffic equations (4) define the effective arrival rates only up to a multiplicative constant. The mean-value analysis algorithm circumvents this difficulty by calculating the effective arrival rates $\Lambda_i(\ell)$ by induction over ℓ. Note that this algorithm does not give the normalization constant of the equilibrium distribution, so that it does not help calculate other performance metrics.

3 Appendix: Irreducibility assumptions

In this section, we discuss the uniqueness of the solutions of the traffic equations (1) and (4) for open and closed Jackson networks. In both cases, the proof of uniqueness consists of interpreting the traffic equations as the equilibrium equations of a (discrete-time) Markov chain. This Markov chain has a natural interpretation in the network: it describes the sequence of queues visited by a typical customer, while ignoring the amount of time spent by the customer at each queue. The irreducibility assumptions stated in Sections 1.1 and 2.1 guarantee that this Markov chain is irreducible, so that its equilibrium equations have a unique solution up to a multiplicative constant; the arrival rates to the queues are one of these solutions. These proofs are inspired by [1, Chapter 7].

3.1 Open Jackson networks

Consider an open Jackson network as described in Section 1.1. A typical customer enters the network, visits a random number of queues in a random order, and leaves the network. Correspondingly, we consider a Markov chain with state space $S = \{0, 1, 2, \ldots, k\}$, where states 1 to k correspond to the queues in the network and state 0 corresponds to "outside" the network. When the Markov chain is in state $i = 1, 2, \ldots, k$, it next jumps to state $j = 1, 2, \ldots, k$ with probability $q_{i,j}$ and to state $j = 0$ with probability q_i. When the Markov chain is in state $i = 0$, it next jumps to state $j = 1, 2, \ldots, k$ with probability $\frac{\nu_i}{\nu}$, where $\nu = \nu_1 + \nu_2 + \ldots + \nu_k$ (but it cannot jump back to state 0).

The irreducibility assumptions stated in Section 1.1 guarantee that this Markov chain is irreducible, which in turn implies that it has a unique equilibrium distribution. Let π_i denote the equilibrium probability that the Markov chain is in state i, for each $i = 0, 1, 2, \ldots, k$. We obtain the following set of equilibrium equations:

$$\pi_0 = \sum_{i=1}^{k} \pi_i q_i, \qquad \pi_i = \pi_0 \frac{\nu_i}{\nu} + \sum_{j=1}^{k} \pi_j q_{j,i}, \quad i = 1, 2, \ldots, k.$$

The equation for $i = 0$ is redundant, as it follows from the others by summation. The traffic equations (1) of the open Jackson network are equivalent to these equilibrium equations upon imposing that $\pi_0 = \nu$ (instead of the normalization equation $\pi_0 + \pi_1 + \ldots + \pi_k = 1$). This observation implies that the traffic equations have a solution and that this solution is unique.

3.2 Closed Jackson networks

Now consider a closed Jackson network as described in Section 2.1. The number of customers is given and each customer perpetually visits queues in a random order. The journey of each customer is described by a Markov chain with state space $S = \{1, 2, \ldots, k\}$. When the Markov chain is in state $i = 1, 2, \ldots k$, it next jumps to state j with probability $q_{i,j}$, for each $j = 1, 2, \ldots, k$.

The irreducibility assumptions stated in Section 2.1 guarantee that this Markov chain is irreducible and therefore has a unique equilibrium distribution. In other words, the following equilibrium equations have a unique solution $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ such that $\pi_1 + \pi_2 + \ldots + \pi_k = 1$:

$$\pi_i = \sum_{j=1}^{k} \pi_j q_{j,i}, \quad i = 1, 2, \ldots, k.$$

These are precisely the traffic equations (4) of the closed Jackson network. The uniqueness of the equilibrium distribution is guaranteed by imposing the normalization equation $\pi_1 + \pi_2 + \ldots + \pi_k = 1$. Without this additional equation, we obtain that the traffic equations have a unique solution up to a positive multiplicative constant.

4 Exercises

1. Consider the following open Jackson network of three queues. Customer enter the network according to a Poisson process at rate 10 per hour, and each new customer joins queue 1 with probability $\frac{2}{3}$ and queue 2 with probability $\frac{1}{3}$. The service times at all queues are independent and exponentially distributed, with mean 2 minutes at queue 1, 4 minutes at queue 2, and 3 minutes at queue 3. When a customer completes service at queue 1, this customer enters back queue 1 with probability $\frac{1}{5}$ and joins queue 3 otherwise. When a customer completes service at queue 2, this customer joins queue 3 (with probability 1). When a customer completes service at queue 3, this customer joints queue 2 with probability $\frac{1}{4}$ and leaves the network otherwise.

   (a) Calculate the effective arrival rate at each queue and verify that the network is stable.

   (b) Calculate the mean number of customers in the network.

   (c) Calculate the mean sojourn time of customers in the network.

2. Consider an open Jackson network like that described in Section 1.1. Let W denote the mean sojourn time of customers in the network and W_i the mean sojourn time of customers at queue i, for each $i = 1, 2, \ldots, k$. Verify that

   $$W = \sum_{i=1}^{k} \frac{\lambda_i}{\nu} W_i,$$

   where λ_i is the effective arrival rate at queue i, for each $i = 1, 2, \ldots, k$, and $\nu = \nu_1 + \nu_2 + \ldots + \nu_k$ is the total external arrival rate. How can you interpret the quantity $\frac{\lambda_i}{\nu}$ for each $i = 1, 2, \ldots, k$?

3. Tasks arrive at the processor of a computer according to a Poisson process with rate λ per time unit. The service time of each task is an independent random variable exponentially distributed with mean $\frac{1}{\mu}$ time units. We assume that $\lambda < \mu$. The processor applies the following preemptive-resume

scheduling policy. When the processor starts processing a task, it samples a random variable T that is exponentially distributed with parameter θ. If the task is not completed after T time units, the service of the task is preempted, the task is moved to the end of the queue, and the processor immediately starts processing the first task in the queue. (If there is a single task in the queue, the server resumes the service of this task immediately after interrupting it.) If the service of a task is resumed after a service interruption, the processor continues to process the task where it left off.

(a) Using the properties of the exponential distribution, explain how this system can be modeled as a single-server queue with a Markov routing process.

(b) Calculate the equilibrium distribution of the queue state. What do you observe?

(c) Can you think of an advantage of this scheduling policy compared to first-come-first-served?

Acknowledgements

Many thanks to Sem Borst and Rik Timmerman for their feedback on an earlier draft of these lecture notes.

References

[1] Thomas Bonald and Mathieu Feuillet. *Network Performance Analysis.* John Wiley & Sons.