6 Queueing Networks

So far, the queueing systems we studied were only *single resource* systems: that is, there was one service facility, possibly with multiple servers. Actual computer systems and communication networks are *multiple resource systems*. Thus, we may have online terminals or workstations, communication lines, etc., as well as the computer itself. The computer, even the simplest personal computer, has multiple resources, too, including main memory, virtual memory, coprocessors, I/O devices, etc. There may be a queue associated with each of these resources. Thus, a computer system or a communication network is a *network of queues*.

Queueing networks may be *single-class* or *multiclass* according to whether customers are not distinguishable or they have distinct characteristics.

A queueing network is *open* if customers enter from outside the network, circulate among the service centers (or queues or nodes) for service, and depart from the network. A queueing network is *closed* if a *fixed* number of customers circulate indefinitely among the queues. A queueing network is *mixed* if some customers enter from outside the network and eventually leave, and if some customers always remain in the network.

6.1 Open Jackson network

Consider an *open single-class* network consisting of K single-server queues with infinite buffer capacity. Jobs arriving from outside the system and joining queue i according to a Poisson process with rate λ_i^0. After service at queue i, which is *exponentially* distributed with parameter μ_i, the job either leaves the system with probability p_{i0}, or goes to queue j, with probability p_{ij}. Clearly, $\sum_{j=0}^{K} p_{ij} = 1$, since each job must go somewhere.

As usual, the arrival times and the service times are assumed to be mutually independent rvs.

Let $X_i(t)$ be the number of customers in queue (or node) i at time t and define

$$\mathbf{X}(t) = (X_1(t), \ldots, X_K(t)) \quad \forall t \geq 0.$$

As usual we will be interested in the computation of

$$\pi_{\underline{n}} = \lim_{t\to\infty} P(\mathbf{X}(t) = \underline{n})$$

with $\underline{n} := (n_1, \ldots, n_K) \in \mathbb{N}^K$.

Because of the Poisson and exponential assumptions the continuous-time, discrete-space stochastic process $\{\mathbf{X}(t),\, t \geq 0\}$ is seen to be a continuous-time Markov chain on the state-space $I = \mathbb{N}^K$.

Using $\mathbb{1}_{k>0} = 1$ if $k > 0$ and 0 otherwise and letting $\underline{e}_i$ be the vector with all components zero, except the i-th one which is one, the balance equations for this CTMC are (cf. (29))

$$\begin{aligned}\pi_{\underline{n}}\left(\sum_{i=1}^{K}\lambda_i^0+\sum_{i=1}^{K}\mathbb{1}_{n_i>0}\,(1-p_{ii})\mu_i\right) &= \sum_{i=1}^{K}\mathbb{1}_{n_i>0}\,\lambda_i^0\,\pi_{\underline{n}-\underline{e}_i}\\ &\quad+\sum_{i=1}^{K}p_{i0}\,\mu_i\,\pi_{\underline{n}+\underline{e}_i}\\ &\quad+\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\mathbb{1}_{n_j>0}p_{ij}\mu_i\,\pi_{\underline{n}+\underline{e}_i-\underline{e}_j}.\end{aligned} \tag{90}$$

Proposition 18 (Open Jackson network, 1957)**.** *Define* $\rho_i=\lambda_i/\mu_i$ *for all* $i=1,2,\ldots,K$*. If* $\lambda_i<\mu_i$ *for all* $i=1,2,\ldots,K$ *then*

$$\pi_{\underline{n}}=\prod_{i=1}^{K}(1-\rho_i)\rho_i^{n_i}\quad\forall\underline{n}=(n_1,\ldots,n_K)\in\mathbb{N}^K \tag{91}$$

where $(\lambda_1,\lambda_2\ldots,\lambda_K)$ *is the unique nonnegative solution of the system of linear equations*

$$\lambda_i=\lambda_i^0+\sum_{j=1}^{K}p_{ji}\,\lambda_j,\qquad i=1,2,\ldots,K. \tag{92}$$

■

Letting $\underline{\lambda}=(\lambda_1,\ldots,\lambda_K)$ and $\underline{\lambda}^0=(\lambda_1^0,\ldots,\lambda_K^0)$, (92) can be rewritten

$$\underline{\lambda}=\underline{\lambda}^0+\underline{\lambda}\mathbf{P} \tag{93}$$

with $\mathbf{P}=[p_{ij}]_{1\le i,j\le K}$ the routing matrix. If the matrix $(\mathbf{I}-\mathbf{P})^{-1}$ exists, the unique solution of (93) is given by

$$\underline{\lambda}=\underline{\lambda}^0(\mathbf{I}-\mathbf{P})^{-1}. \tag{94}$$

The traffic equations have a unique solution (i.e. matrix $\mathbf{I}-\mathbf{P}$ is invertible) if and only if all nodes of the network are open (the Jackson network is said to be *completely open*). A queue/node in the network is said to be open if a customer/job visiting it is certain to leave the network at some point. In practice, it is enough to have one queue i_0 such that $p_{i_00}>0$ and paths from each other queue to queue i_0 for the network to be completely open.

The existence of matrix $(\mathbf{I}-\mathbf{P})^{-1}$ is equivalent to the irreducibility of the Markov process $\{\mathbf{X}(t),t\ge 0\}$ on $\mathbb{N}^K$. Therefore, the statement of Proposition 18 can be reformulated as

Proposition 19 (Open Jackson network - alternative statement)**.** *If the matrix* $\mathbf{I}-\mathbf{P}$ *is invertible and if* $\lambda_i<\mu_i$ *for* $i=1,\ldots,K$ *then* (91) *holds, where* $(\lambda_1,\ldots,\lambda_K)$ *is given in* (94). ■

Let us comment this fundamental result of queueing network theory obtained by J. R. Jackson in 1957.

Equations (92) are referred to as the *traffic equations.* Let us show that λ_i is the total arrival rate at node i when the system is in steady-state.

To do so, let us first determine the total throughput of a node. The total throughput of node i consists of the customers who arrive from outside the network with rate λ_i^0, plus all the customers who are transferred to node i after completing service at node j for all nodes in the network. If λ_i is the total throughput of node i, then the rate at which customers arrive at node i from node j is $p_{ji}\lambda_j$. Hence, the throughput of node i, λ_i must satisfy (92).

Since, in steady-state, the throughput of every node is equal to the arrival rate at this node, we see that λ_i is also the total arrival rate in node i.

Hence, the conditions $\lambda_i < \mu_i$ for $i = 1, 2, \ldots, K$, are the *stability conditions* of an open Jackson network.

Let us now discuss the form of the limiting distribution (91). We see that (91) is a *product* of terms, where the i-th term $(1 - \rho_i)\,\rho_i^{n_i}$ is actually the limiting distribution function of the number of customers in an isolated M/M/1 queue with arrival rate λ_i and service rate μ_i. This property is usually referred to as the *product-form* property.

Therefore, the network state probability (i.e., $\pi_{\underline{n}}$) is the product of the state probabilities of the individual queues.

It is important to note that the steady-state probabilities behave as if the total arrival process at every node (usually referred to as the *flow*) were Poisson (with rate λ_i for node i), but the flows are not Poisson in general! The flows are Poisson if and only if there are *no cycles* in the network of queues. This is a consequence of the fact that (*i*) sampling a Poisson process yields another Poisson process, (*ii*) merging independent Poisson processes yield another Poisson process, and (*iii*) (Burke's Theorem) the departure process from an M/M/1 queue at equilibrium is a Poisson process.

Proof of Proposition 18. Because of the markovian assumptions (Poisson external arrivals and exponential service times, all of them independent), it is seen by construction rule $\sharp$ 2 that the stochastic process $\{\mathbf{X}(t),\, t \geq 0\}$ is a CTMC. Regarding the irreducibility property, we know that the traffic equations (92) have a unique nonnegative solution $(\lambda_1, \ldots, \lambda_K)$ when the matrix $\mathbf{I} - \mathbf{P}$ is invertible. The necessary and sufficient condition for this is that all queues are open (there is a way out, wherever we are inside the network). On the other hand, because every node may receive and serve infinitely many customers, one must show that any state of the CTMC is reachable from any other state. It is enough to show that the probability of going from state $(n_1, \ldots, n_K)$ to state $(m_1, \ldots, m_K)$ in exactly t units of time is strictly positive for all $t > 0$, $n_i \geq 0$, $m_i \geq 0$, $i = 1, 2, \ldots, K$. Let $s < t$. Since the probability of having 0 external arrival in $[0, s)$ is strictly positive we see that the probability that the system is empty at time s is also strictly positive. On the other hand, starting from an empty system, it should be clear that we can reach any state in exactly $t - s$ units of time). The only condition for having the irreducibility property is that matrix $\mathbf{I} - \mathbf{P}$ is invertible (this ensures that the state $(0, \ldots, 0)$ is reachable from any other state, since all queues can become empty).

Hence, Proposition 5 applies to the irreducible CTMC. Thus, it suffices to check that (91) satisfies the balance equations together with the condition $\sum_{\underline{n} \in \mathbb{N}^K} \pi_{\underline{n}} = 1$.

Observe from (91) that the latter condition is trivially satisfied as long as $\lambda_i < \mu_i$ for all $i = 1, 2, \ldots, K$ (stability condition). It remains to check that (91) satisfies the balance equations (90).

Observe from (91) that

$$\begin{aligned}
\pi_{\underline{n}-\underline{e}_i} &= \frac{\pi_{\underline{n}}}{\rho_i} \\
\pi_{\underline{n}+\underline{e}_i} &= \pi_{\underline{n}}\rho_i \\
\pi_{\underline{n}+\underline{e}_i-\underline{e}_j} &= \frac{\pi_{\underline{n}}\rho_i}{\rho_j}.
\end{aligned}$$

Plugging these values in the r.h.s. of (90) then dividing both sides of the resulting equation by $\pi_{\underline{n}}$ yields

$$\begin{aligned}
\sum_{i=1}^{K} \lambda_i^0 + \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,(1-p_{ii})\mu_i &= \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,\frac{\lambda_i^0\mu_i}{\lambda_i} + \sum_{i=1}^{K} p_{i0}\,\lambda_i \\
&\quad + \sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K} \mathbb{1}_{n_j>0}\, p_{ij}\frac{\lambda_i\mu_j}{\lambda_j}
\end{aligned}$$

that we can rewrite as

$$\begin{aligned}
\sum_{i=1}^{K} \lambda_i^0 - \sum_{i=1}^{K} p_{i0}\,\lambda_i &= -\sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,(1-p_{ii})\mu_i + \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,\frac{\lambda_i^0\mu_i}{\lambda_i} \\
&\quad + \sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K} \mathbb{1}_{n_j>0}\, p_{ij}\frac{\lambda_i\mu_j}{\lambda_j}. \qquad (95)
\end{aligned}$$

The l.h.s. of (95) is equal to 0. Indeed, by summing up the traffic equations (92) over all values of $i = 1, 2, \ldots, K$, we find that

$$\begin{aligned}
\sum_{i=1}^{K} \lambda_i &= \sum_{i=1}^{K} \lambda_i^0 + \sum_{i=1}^{K}\sum_{j=1}^{K} p_{ji}\lambda_j \\
&= \sum_{i=1}^{K} \lambda_i^0 + \sum_{j=1}^{K} \lambda_j \sum_{i=1}^{K} p_{ji} \\
&= \sum_{i=1}^{K} \lambda_i^0 + \sum_{j=1}^{K} \lambda_j (1-p_{j0})
\end{aligned}$$

so that $\sum_{i=1}^{K} \lambda_i^0 - \sum_{i=1}^{K} p_{i0}\,\lambda_i = 0$ as announced earlier.

Let us now show that the r.h.s. of (95) is also equal to 0. We have

$$
\begin{aligned}
&-\sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,(1-p_{ii})\mu_i + \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,\frac{\lambda_i^0 \mu_i}{\lambda_i} + \sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K} \mathbb{1}_{n_j>0} p_{ij}\frac{\lambda_i \mu_j}{\lambda_j} \\
&\quad = -\sum_{i=1}^{K} \mathbb{1}_{n_i>0}\;\mu_i + \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,\frac{\lambda_i^0 \mu_i}{\lambda_i} + \sum_{i=1}^{K}\sum_{j=1}^{K} \mathbb{1}_{n_j>0} p_{ij}\frac{\lambda_i \mu_j}{\lambda_j} \\
&\quad = -\sum_{i=1}^{K} \mathbb{1}_{n_i>0}\;\mu_i + \sum_{i=1}^{K} \mathbb{1}_{n_i>0}\,\frac{\lambda_i^0 \mu_i}{\lambda_i} + \sum_{j=1}^{K}\sum_{i=1}^{K} \mathbb{1}_{n_i>0} p_{ji}\frac{\lambda_j \mu_i}{\lambda_i} \\
&\quad = -\sum_{i=1}^{K} \mathbb{1}_{n_i>0}\frac{\mu_i}{\lambda_i}\left(\lambda_i - \lambda_i^0 - \sum_{j=1}^{K} p_{ji}\lambda_j\right) \\
&\quad = 0
\end{aligned}
$$

from (92). This concludes the proof. ★

This result actually extends to the case when each queue in the network is a multi-server queue. Assume that node i has c_i servers. The following result holds:

Proposition 20 (Open Jackson network of multi-server queues)**.** *Define $\mu_i(r) = \mu_i \min(r, c_i)$ for $r \geq 0$, $i = 1, 2, \ldots, K$ and let $\rho_i = \lambda_i / c_i\mu_i$ for $i = 1, 2, \ldots, K$.*

If $\lambda_i < c_i\, \mu_i$ for all $i = 1, 2, \ldots, K$, then

$$
\pi_{\underline{n}} = \prod_{i=1}^{K} C_i \left(\frac{\lambda_i^{n_i}}{\prod_{r=1}^{n_i} \mu_i(r)} \right) \quad \forall \underline{n} = (n_1, \ldots, n_K) \in \mathbb{N}^K, \tag{96}
$$

where $(\lambda_1, \lambda_2 \ldots, \lambda_K)$ is the unique nonnegative solution of the system of linear equations (92), and where C_i is given by

$$
C_i = \left[\sum_{r=0}^{c_i-1} \left(\frac{\lambda_i}{\mu_i}\right)^r \frac{1}{r!} + \left(\frac{\lambda_i}{\mu_i}\right)^{c_i} \frac{1}{c_i!} \left(\frac{1}{1-\rho_i}\right) \right]^{-1}. \tag{97}
$$

■

Example 9. Consider a switching facility that transmits messages to a required destination. A NACK (Negative ACKnowledgment) is sent by the destination when a packet has not been properly transmitted. If so, the packet in error is retransmitted as soon as the NACK is received.

We assume that packets arrive at the switch according to a Poisson process with rate λ^0. We assume that the time to send a message correctly is exponentially distributed with parameter μ. The time to send a message incorrectly and receive a NACK is also exponentially distributed with same parameter μ. Let p, $0 < p \leq 1$, be the probability that a message is transmitted properly.

Thus, we can model this switching facility as a Jackson network with one node, where $c_1 = 1$ (one server), $p_{10} = p$ and $p_{11} = 1 - p$. By Jackson's theorem we have that, π_n, the number of packets in the service facility in steady-state, is given by

$$\pi_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n, \quad n \geq 0$$

provided that $\lambda < \mu$, where λ is the solution of the traffic equation

$$\lambda = \lambda^0 + (1-p)\lambda.$$

Therefore, $\lambda = \lambda^0/p$, and

$$\pi_n = \left(1 - \frac{\lambda^0}{p\mu}\right)\left(\frac{\lambda^0}{p\mu}\right)^n, \quad n \geq 0$$

provided that $\lambda^0 < p\mu$.

The mean number of packets (denoted as $\overline{X}$) in the switching facility is then given by (see (64))

$$\overline{X} = \frac{\lambda^0}{p\mu - \lambda^0}$$

and by, Little's formula, the mean response time (denoted as $\overline{T}$) is

$$\overline{T} = \frac{1}{p\mu - \lambda^0}.$$

♦

Example 10. We consider the model in Example 9 but we now assume that the switching facility is composed of K nodes in series, each modelled as a single-server infinite queue with common service rate μ. In other words, we now have a Jackson network with K queues where $\lambda_i^0 = 0$ for $i = 2, 3, \ldots, K$ (no external arrivals at nodes $2, \ldots, K$), $\mu_i = \mu$ for $i = 1, 2, \ldots, K$, $p_{ii+1} = 1$ for $i = 1, 2, \ldots, K-1$, $p_{K,0} = p$ and $p_{K,1} = 1 - p$.

For this model, the traffic equations read

$$\lambda_i = \lambda_{i-1}$$

for $i = 2, 3, \ldots, K$, and

$$\lambda_1 = \lambda^0 + (1-p)\,\lambda_K.$$

It is easy to see that the solution to this system of equations is

$$\lambda_i = \frac{\lambda^0}{p} \quad \forall i = 1, 2, \ldots, K.$$

Hence, by Jackson's theorem, the joint distribution function $\pi_{\underline{n}}$ of the number of packets in the system is given by

$$\pi_{\underline{n}} = \left(\frac{p\mu - \lambda^0}{p\mu}\right)^K \left(\frac{\lambda^0}{p\mu}\right)^{n_1 + \cdots + n_K} \qquad \forall \underline{n} = (n_1, n_2, \ldots, n_K) \in \mathbb{N}^K$$

provided that $\lambda^0 < p\mu$. In particular, the probability $q_{ij}(r,s)$ of having r packets in node i and s packets in node $j > i$ is given by

$$\begin{aligned} q_{ij}(r,s) &= \sum_{n_l \geq 0,\ l \notin \{i,j\}} \pi_{n_1,\ldots,n_{i-1},r,n_{i+1},\ldots,n_{j-1},s,n_{j+1},\ldots,n_K} \\ &= \left(\frac{p\mu - \lambda^0}{p\mu}\right)^2 \left(\frac{\lambda^0}{p\mu}\right)^{r+s}. \end{aligned}$$

Let us now determine for this model the expected sojourn time of a packet. Since queue i has the same characteristics as an M/M/1 queue with arrival rate λ^0/p and mean service time $1/\mu$, the mean number of packets (denoted as $\overline{X_i}$) is given by

$$\overline{X_i} = \frac{\lambda^0}{p\mu - \lambda^0}$$

for every $i = 1, 2, \ldots, K$. Therefore, the total expected number of packets in the network is

$$\sum_{i=1}^{K} E[X_i] = \frac{K\,\lambda^0}{p\mu - \lambda^0}.$$

Hence, by Little's formula, the expected sojourn time in the network is given by

$$\overline{T} = \frac{1}{\lambda_0} \sum_{i=1}^{K} E[X_i] = \frac{K}{p\mu - \lambda^0}.$$

♦

Example 11 (The open central server network)**.** Consider a computer system with one CPU and several I/O devices. A job enters the system from the outside and then waits until its execution begins. During its execution by the CPU, I/O requests may be needed. When an I/O request has been fulfilled the job then returns to the CPU for additional treatment. If the latter is available then the service begins at once; otherwise the job must wait. Eventually, the job is completed (no more I/O requests are requested) and it leaves the system.

We are going to model this system as an open Jackson network with 3 nodes: one node (node 1) for the CPU and two nodes (nodes 2 and 3) for the I/O devices. In other words, we assume that $K = 3$, $\lambda_i^0 = 0$ for $i = 2, 3$ (jobs cannot access the I/O devices directly from the outside) and $p_{21} = p_{31} = 1$, $p_{10} > 0$.

For this system the traffic equations are:

$$\begin{aligned} \lambda_1 &= \lambda_1^0 + \lambda_2 + \lambda_3 \\ \lambda_2 &= \lambda_1\, p_{12} \\ \lambda_3 &= \lambda_1\, p_{13}. \end{aligned}$$

The solution of the traffic equations is $\lambda_1 = \lambda_1^0/p_{10}$, $\lambda_i = \lambda_1^0\, p_{1i}/p_{10}$ for $i = 2, 3$. Thus,

$$\pi_{\underline{n}} = \left(1 - \frac{\lambda_1^0}{\mu_1\, p_{10}}\right) \left(\frac{\lambda_1^0}{\mu_1\, p_{10}}\right)^{n_1} \prod_{i=2}^{3} \left(1 - \frac{\lambda_1^0\, p_{1i}}{\mu_i\, p_{10}}\right) \left(\frac{\lambda_1^0\, p_{1i}}{\mu_i\, p_{10}}\right)^{n_i} \qquad \forall \underline{n} = (n_1, n_2, n_3) \in \mathbb{N}^3$$

and

$$\overline{T} = \frac{1}{\mu_1\, p_{10} - \lambda_1^0} + \sum_{i=2}^{3} \frac{p_{1i}}{\mu_i\, p_{10} - \lambda_1^0\, p_{1i}}.$$

♦

We are now going to consider more general queueing networks that still enjoy the product-form property.

6.2 Kelly networks

In Jackson networks customers follow random routes. There are however many systems in which paths or routes are deterministic. Typical examples are flexible manufacturing systems and connection-oriented communication networks (e.g. ATM networks). For such systems one can use *Kelly networks*. In a Kelly network customers belong to different classes where each class corresponds to a deterministic route. More specifically, to each customer of class k is associated a deterministic route $r_k = (r_k^1, \ldots, r_k^{n_k})$ where r_k^j is the identity of the jth node to be visited and n_k the total number of visits. A customer may visit several times the same node.

There are K nodes and R classes of customers. Each node is equipped with a single server and an infinite capacity waiting room. Customers of class k arrive to the system according to a Poisson process (rate λ_k) and require exponential service times with node dependent rates. Let μ_i be the service rate of customers at node i.

Once routes and external arrival rates on each route have been specified one can calculate the *global* rate of customers of class k entering node i, denoted by $\hat{\lambda}_{ik}$. It is given by

$$\hat{\lambda}_{ik} = \lambda_k \sum_{j=1}^{n_k} \mathbb{1}_{r_k^j = i} = \begin{cases} 0 & \text{if node } i \text{ is not in route } r_k, \\ \ell\lambda_k & \text{if node } i \text{ appears } \ell \text{ times in route } r_k. \end{cases}.$$

In particular, if class k customers enter node i exactly once then $\hat{\lambda}_{ik} = \lambda_k$. The total arrival rate in node i, denoted by $\hat{\lambda}_i$, is given by

$$\hat{\lambda}_i = \sum_{k=1}^{R} \hat{\lambda}_{ik}.$$

Let $X_{ik}(t)$ be the number of customers in node i that are from class k at time t. The state of a Kelly network at time t is then the K-by-R matrix $\mathbf{X}(t) = [X_{ik}(t)]_{\substack{1\le i\le K \\ 1\le k\le R}}$. The state-space of a Kelly network is represented by the set $\mathcal{E}_{KR}$ of all K-by-R matrices with entries in $\mathbb{N}$. Entry n_{ik} gives the number of customers of class k in node i. Observe that, unless customers visit a node at most once, the stochastic process $\{\mathbf{X}(t), t > 0\}$ is not a Markov process (can you see why?).

The objective is to find the steady-state distribution of $\mathbf{X}(t)$, namely

$$\pi_{\mathbf{N}} = \lim_{t\to\infty} P(\mathbf{X}(t) = \mathbf{N}) = \lim_{t\to\infty} P(X_{ik}(t) = n_{ik}; 1 \le i \le K, 1 \le k \le R).$$

The following result holds:

Theorem 1 (Kelly network, 1975)**.** *If $\hat{\lambda}_i < \mu_i$ (stability condition) for all $i = 1, \ldots, K$ the stationary probability $\pi_{\mathbf{N}}$ that the network is in state $\mathbf{N} = [n_{ik}] \in \mathcal{E}_{KR}$ is*

$$\pi_{\mathbf{N}} = \prod_{i=1}^{K} \left(1 - \frac{\hat{\lambda}_i}{\mu_i}\right) \binom{\sum_{k=1}^{R} n_{ik}}{n_{i1}, n_{i2}, \ldots, n_{iR}} \prod_{k=1}^{R} \left(\frac{\hat{\lambda}_{ik}}{\mu_i}\right)^{n_{ik}}, \tag{98}$$

where $\binom{\sum_{k=1}^{R} n_{ik}}{n_{i1}, n_{i2}, \ldots, n_{iR}} = \frac{(\sum_{k=1}^{R} n_{ik})!}{\prod_{k=1}^{R} n_{ik}!}$ is the multinomial coefficient. ■

This is a product-form result where the ith term of the product is the steady-state of a node i taken in isolation. Observe that (98) reduces to (91) if there is only one class ($R = 1$).

From this result one can calculate various metrics of interest such as the expected number of customers of class k in node i

$$\overline{N}_{ik} = E[X_{ik}] = \frac{\hat{\lambda}_{ik}}{\mu_i - \hat{\lambda}_i}, \tag{99}$$

the expected number of customers in node i

$$\overline{N}_i = \frac{\hat{\lambda}_i}{\mu_i - \hat{\lambda}_i}, \tag{100}$$

the expected number of customers of class k in the network

$$\overline{N}^{(k)} = \sum_{i=1}^{K} \overline{N}_{ik}, \tag{101}$$

the expected sojourn time of customers of class k in the network (Hint: use Little's formula)

$$\overline{T}_k = \frac{1}{\lambda_k} \sum_{i=1}^{K} \frac{\hat{\lambda}_{ik}}{\mu_i - \hat{\lambda}_i} = \frac{1}{\lambda_k} \sum_{j=1}^{n_k} \frac{\lambda_k}{\mu_{r_k^j} - \hat{\lambda}_{r_k^j}} = \sum_{j=1}^{n_k} \frac{1}{\mu_{r_k^j} - \hat{\lambda}_{r_k^j}}. \tag{102}$$

In the first expression, some terms in the summation could be null (terms corresponding to nodes not in route k). In the second and third expressions, some terms may appear multiple times (case when a node appears more than once in route k).

The sojourn time in the network of an arbitrary customer is (use Little's formula)

$$\overline{T} = \frac{\sum_{i=1}^{K} \overline{N}_i}{\sum_{k=1}^{R} \lambda_k} = \frac{1}{\sum_{k=1}^{R} \lambda_k} \sum_{i=1}^{K} \frac{\hat{\lambda}_i}{\mu_i - \hat{\lambda}_i}. \tag{103}$$

We have the following relation

$$\overline{T} = \sum_{k=1}^{R} \frac{\lambda_k}{\sum_{l=1}^{R} \lambda_l} \overline{T}_k.$$

The sojourn time in the network of an arbitrary customer is the weighted average of the sojourn time in the network of customers pertaining to specific classes; the weight is the probability that a customer pertains to a given class.

6.2.1 Example

Consider a Kelly network with 2 nodes A and B and 3 routes $r_1 = (A, B, A)$, $r_2 = (A)$ and $r_3 = (B)$. Customers arrival rates on the 3 routes are λ_1, λ_2, and λ_3, respectively. The service rates are μ_A and μ_B, respectively. We can compute the following rates:

$$\left.\begin{array}{l}\hat{\lambda}_{A1} = 2\lambda_1\\ \hat{\lambda}_{A2} = \lambda_2\\ \hat{\lambda}_{A3} = 0\end{array}\right\} \Rightarrow \hat{\lambda}_A = 2\lambda_1 + \lambda_2 \ ; \qquad \left.\begin{array}{l}\hat{\lambda}_{B1} = \lambda_1\\ \hat{\lambda}_{B2} = 0\\ \hat{\lambda}_{B3} = \lambda_3\end{array}\right\} \Rightarrow \hat{\lambda}_B = \lambda_1 + \lambda_3 \ ,$$

and the following expected number of customers:

	class 1	class 2	class 3	random
node A	$\overline{N}_{A1} = \frac{2\lambda_1}{\mu_A - 2\lambda_1 - \lambda_2}$	$\overline{N}_{A2} = \frac{\lambda_2}{\mu_A - 2\lambda_1 - \lambda_2}$	$\overline{N}_{A3} = 0$	$\overline{N}_A = \frac{2\lambda_1 + \lambda_2}{\mu_A - 2\lambda_1 - \lambda_2}$
node B	$\overline{N}_{B1} = \frac{\lambda_1}{\mu_B - \lambda_1 - \lambda_3}$	$\overline{N}_{B2} = 0$	$\overline{N}_{B3} = \frac{\lambda_3}{\mu_B - \lambda_1 - \lambda_3}$	$\overline{N}_B = \frac{\lambda_1 + \lambda_3}{\mu_B - \lambda_1 - \lambda_3}$
network	$\overline{N}^{(1)} = \overline{N}_{A1} + \overline{N}_{B1}$	$\overline{N}^{(2)} = \overline{N}_{A2} + \overline{N}_{B2}$	$\overline{N}^{(3)} = \overline{N}_{A3} + \overline{N}_{B3}$	$\overline{N}_A + \overline{N}_B$

The expected sojourn time of a customer from each class and from a random customer are

$$\begin{aligned}
\overline{T}_1 &= \frac{2}{\mu_A - 2\lambda_1 - \lambda_2} + \frac{1}{\mu_B - \lambda_1 - \lambda_3}\\
\overline{T}_2 &= \frac{1}{\mu_A - 2\lambda_1 - \lambda_2}\\
\overline{T}_3 &= \frac{1}{\mu_B - \lambda_1 - \lambda_3}\\
\overline{T} &= \frac{\overline{N}_A + \overline{N}_B}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{\lambda_1\overline{T}_1 + \lambda_2\overline{T}_2 + \lambda_3\overline{T}_3}{\lambda_1 + \lambda_2 + \lambda_3} \ .
\end{aligned}$$