15 Closed production networks

In the previous chapter we developed and analyzed stochastic models for production networks with a *free inflow of jobs*. In this chapter we will study production networks for which the nature of the input process is different: there is only new input when a finished job leaves the network. For example, one may think of a production system where jobs (e.g., parts) are transported through the system on *pallets*. Typically, the number of pallets available is limited, since pallets are expensive. When processing is completely finished, a job is transported to the Load/Unload (LU) station, where the job is removed from the pallet, and a new job (e.g., raw part) is immediately attached to the pallet and released in the network. In doing so, the number of circulating jobs (or pallets) in the network remains constant over time. Networks with a fixed population are called *closed networks*. One of the important design issues in closed production networks is to determine the population size (e.g., number of pallets) required to meet a certain target throughput.

In the following section we start with a simple closed queueing network model, with only single-server exponential stations and one job type (see also [2]).

15.1 Exponential closed single-server queueing network model

We consider a production system consisting of M work stations, numbered $1, 2, \dots, M$; see figure 1. Each work station has a single machine. The production system is processing one type of jobs. The number of circulating jobs is N.

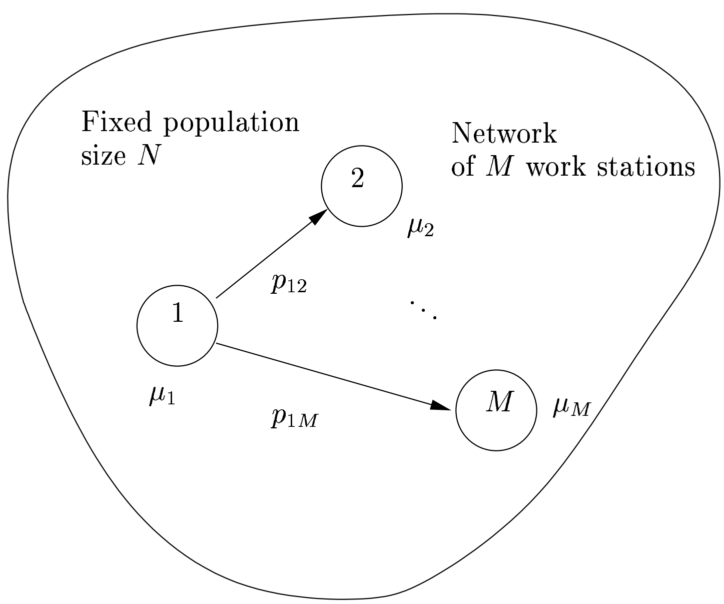

Figure 1: Exponential closed single-server queueing network model with M work stations and N circulating jobs

The processing times at work station m are exponentially distributed with mean $1/\mu_m$, and the processing order is FCFS. The routing of jobs through is system is Markovian: after visiting work station m, a job moves to station n with probability p_{mn} (so $\sum_{n=1}^{M} p_{mn} = 1$). Let P denote the matrix of routing probabilities p_{mn}. We assume that P is irreducible (so a job can reach from each station any other station in one or more transitions).

This model is known as a *closed Jackson network*; see, e.g., [3]. It is called closed, because the population size remains constant over time. Note that here stability is not an issue; we do not have to worry about the population size possibly growing to infinity.

Since the processing times are assumed to be exponential and the routing is Markovian, this network can be described by a Markov process with states $\underline{k} = (k_1, k_2, \ldots, k_M)$ where k_m denotes the number of jobs in work station m. The possible states are the ones for which $\underline{k} \geq 0$ and

$$\sum_{m=1}^{M} k_m = N.$$

Hence the state space is finite; however, the number of states is equal to

$$\binom{N+M-1}{M-1},$$

so it may be very big for already moderate values of N and M. Let $p(\underline{k})$ denote the equilibrium probability of state $\underline{k}$. Below we will derive an explicit form for these probabilities.

By equating the flow out and into state $\underline{k}$ we get

$$p(\underline{k}) \sum_{m=1}^{M} \mu_m \epsilon(k_m) = \sum_{n=1}^{M} \sum_{m=1}^{M} p(\underline{k} + \underline{e}_n - \underline{e}_m) \mu_n p_{nm} \epsilon(k_m). \tag{1}$$

As solution we are going to try the form

$$p(\underline{k}) = C x_1^{k_1} x_2^{k_2} \cdots x_M^{k_M}.$$

Substitution of this form into the balance equation (1) and dividing by common powers yields (after rearranging terms)

$$\sum_{m=1}^{M} \left(\mu_m - \sum_{n=1}^{M} \frac{x_n}{x_m} \mu_n p_{nm} \right) \epsilon(k_m) = 0.$$

The left-hand side is a sum of functions $\epsilon(k_m)$. This sum only vanishes for all $\underline{k}$ if the coefficients of all $\epsilon(k_m)$ vanish, so the x_m's should satisfy

$$x_m \mu_m = \sum_{n=1}^{M} x_n \mu_n p_{nm}, \qquad m = 1, 2, \ldots, M.$$

If we set $v_m = x_m \mu_m$, then

$$v_m = \sum_{n=1}^{M} v_n p_{nm}, \qquad m = 1, 2, \ldots, M.$$

Clearly v_m can be interpreted as the *relative visiting frequency* or *relative arrival rate* to work station m. The above set of equations does not have a unique solution, and therefore we have to add a normalization equation, such as $v_1 = 1$. This equation is natural if the network has a LU station, numbered station 1; then v_m denotes the mean number of times a job has to visit work station m before returning to the LU station.

So x_m is given by

$$x_m = \frac{v_m}{\mu_m}, \qquad m = 1, \ldots, M,$$

and thus we find that

$$p(\underline{k}) = C \left(\frac{v_1}{\mu_1}\right)^{k_1} \left(\frac{v_2}{\mu_2}\right)^{k_2} \cdots \left(\frac{v_M}{\mu_M}\right)^{k_M},$$

where C follows from normalization. Summarizing, the conclusion is that

$$p(\underline{k}) = C p_1(k_1) p_2(k_2) \cdots p_M(k_M), \qquad \underline{k} \geq 0, \quad \sum_{m=1}^{M} k_m = N, \tag{2}$$

where

$$p_m(k_m) = \left(\frac{v_m}{\mu_m}\right)^{k_m}.$$

Solution (2) is a *product form solution*. Note that $p_m(\cdot)$ is closely related to the queue length distribution of the $M/M/1$ system with (absolute) arrival rate v_m and service rate μ_m. An important difference with the solution for open exponential networks is that in a closed network the queue lengths at the work stations are *dependent*. Another difference is that it is not so easy to compute the normalizing constant C. Simply adding the products

$$\left(\frac{v_1}{\mu_1}\right)^{k_1} \left(\frac{v_2}{\mu_2}\right)^{k_2} \cdots \left(\frac{v_M}{\mu_M}\right)^{k_M}$$

over all possible states $\underline{k}$ will lead to numerical complications when the state space is large (such as overflow or underflow problems). However, efficient and numerically stable algorithms for the computation of the normalizing constant have been developed; see, e.g., Buzen's convolution algorithm [1, 5]. This algorithm is briefly explained below.

Define $C(m, n)$ as

$$C(m, n) = \sum_{\substack{k_1, \ldots, k_m \geq 0 \\ \sum_{i=1}^{m} k_i = n}} \left(\frac{v_1}{\mu_1}\right)^{k_1} \left(\frac{v_2}{\mu_2}\right)^{k_2} \cdots \left(\frac{v_m}{\mu_m}\right)^{k_m}. \tag{3}$$

Clearly, the normalizing constant C for the network with M stations and N circulating jobs is equal to $1/C(M, N)$. To compute $C(M, N)$, note that, by distinguishing the cases $k_m = 0$ and $k_m > 0$ in the sum of (3), the following relation immediately follows

$$C(m, n) = C(m-1, n) + \frac{v_m}{\mu_m} C(m, n-1). \tag{4}$$

Hence, together with the initial conditions

$$C(0,n)=0, \quad n=1,\ldots,N, \quad C(m,0)=1, \quad m=1,\ldots,M,$$

the constant $C(M,N)$ can be recursively computed from (4).

15.2 Exponential closed multi-server queueing network model

We are now going to extend the results of the previous section for single-server stations to multi- and infinite-server stations. Let us assume that work station m has c_m identical parallel machines (possibly $c_m=\infty$). By direct substitution into the balance equations it may be verified that the simultaneous queue length probabilities $p(\underline{k})$ again have a product form solution, i.e.,

$$p(\underline{k}) = Cp_1(k_1)p_2(k_2)\cdots p_M(k_M), \qquad \underline{k}\geq 0, \quad \sum_{m=1}^{M} k_m = N, \tag{5}$$

where

$$p_m(k_m) = \begin{cases} \dfrac{1}{k_m!}\left(\dfrac{v_m}{\mu_m}\right)^{k_m} & k_m \leq c_m-1; \\ \dfrac{1}{c_m!c_m^{k_m-c_m}}\left(\dfrac{v_m}{\mu_m}\right)^{k_m} & k_m\geq c_m. \end{cases}$$

Note that $p_m(\cdot)$ is closely related to the queue length distribution of the $M/M/c_m$ system with arrival rate v_m and service rate μ_m (in fact, the only difference is the normalizing constant).

So far we only considered the detailed queue length probabilities $p(\underline{k})$. In principle, these probabilities can be used to compute mean performance characteristics such as mean queue lengths and mean production lead times. In the next section we will develop an efficient recursive scheme for the computation of mean performance characteristics. It is not (directly) based on the state probabilities $p(\underline{k})$, but it is uses Little's law and an extension of the PASTA property to closed queueing networks; this approach is usually referred to as Mean Value Analysis (MVA), see, e.g., [6].

15.3 Mean value analysis for exponential closed networks

Mean value analysis is based on the *Arrival Theorem* for exponential closed networks. For networks with one job type this theorem states that an arbitrary job moving from one station to another sees the system *in equilibrium* corresponding to a population with *one job less* (later we will also formulate this result for networks with multiple job types). Below we will prove the Arrival Theorem for the exponential single-server network ($c_m=1$ for all stations m).

Let $S(N)$ denote the state space of a network with N circulating jobs, i.e.,

$$S(N) = \{\underline{k}=(k_1,\ldots,k_M)|k_1,\ldots,k_M\geq 0, \sum_{i=1}^{M} k_i = N\}.$$

Then, by (2), the number of jumps per time unit that see the network in state $\underline{k} \in S(N-1)$ (a jumping job does not see himself) is equal to

$$\sum_{m=1}^{M} p(\underline{k}+\underline{e}_m)\mu_m = Cp_1(k_1)\cdots p_M(k_M)\sum_{m=1}^{M} v_m,$$

where the normalizing constant C is given by

$$C = \frac{1}{C(M,N)}.$$

Similarly, the total number of jumps per time unit is equal to

$$\sum_{\underline{k}\in S(N-1)} \sum_{m=1}^{M} p(\underline{k}+\underline{e}_m)\mu_m = \sum_{\underline{k}\in S(N-1)} Cp_1(k_1)\cdots p_M(k_M)\sum_{m=1}^{M} v_m.$$

The fraction of jumps that see the network in state $\underline{k} \in S(N-1)$ is the ratio of the above two rates, and thus, it is equal to

$$\frac{p_1(k_1)\cdots p_M(k_M)\sum_{m=1}^{M} v_m}{\sum_{\underline{k}\in S(N-1)} p_1(k_1)\cdots p_M(k_M)\sum_{m=1}^{M} v_m} = \frac{p_1(k_1)\cdots p_M(k_M)}{C(M,N-1)},$$

which is exactly the probability that the network with $N-1$ circulating jobs is in state $\underline{k}$.

Now we will first explain the mean value approach for a single-server network. Let us introduce some notation; all quantities below depend on the population size k.

$$\begin{aligned} E(S_m(k)) &= \text{mean production lead time at work station } m; \\ E(L_m(k)) &= \text{mean number of jobs in work station } m; \\ \Lambda_m(k) &= \text{throughput of work station } m, \text{ in a network with population } k. \end{aligned}$$

Based on the Arrival Theorem, a job arriving at station m in a network with population k, will find on average $E(L_m(k-1))$ jobs in station m. The mean (residual) processing time of each job in station m is $1/\mu_m$ (also for the one in process). Hence, the mean production lead time is

$$E(S_m(k)) = (E(L_m(k-1))+1)\frac{1}{\mu_m}, \qquad m = 1,2,\ldots,M. \tag{6}$$

The mean cycle time $E(C(k))$ of a job is

$$E(C(k)) = \sum_{n=1}^{M} v_n E(S_n(k)).$$

Hence, the mean number of times per time unit that a job passes station m is equal to $v_m/E(C(k))$, and thus the throughput of station m is $k \cdot v_m/E(C(k))$, since there are k circulating jobs. So we have

$$\Lambda_m(k) = \frac{kv_m}{\sum_{n=1}^{M} v_n E(S_n(k))}, \qquad m = 1,2,\ldots,M. \tag{7}$$

Finally, by Little's law

$$E(L_m(k)) = \Lambda_m(k)E(S_m(k)), \qquad m = 1, 2, \ldots, M. \tag{8}$$

The relations (6)-(8) can be used to recursively compute the means $E(S_m(k))$, $\Lambda_m(k)$ and $E(L_m(k))$ for populations from $k = 0$ up to $k = N$. Initially we have

$$E(L_m(0)) = 0, \qquad m = 1, 2, \ldots, M.$$

Now we consider the multi-server case. It is also possible to develop an exact mean value analysis for multi-server networks, although the resulting algorithm is more complicated than the one above. The reason is that marginal queue length distributions are needed for the computation of the probabilities of waiting (cf. [7]). Instead we develop an efficient and accurate *approximative* mean value approach.

Recall that for an $M/M/c$ system with arrival rate λ and service rate μ we have

$$E(S) = \Pi_W \cdot \frac{1}{c\mu} + \left(E(L) - \frac{\lambda}{\mu}\right) \cdot \frac{1}{c\mu} + \frac{1}{\mu},$$

where Π_W is the probability of waiting, or by PASTA, the probability that all servers are busy. So

$$\Pi_W = \frac{\frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c}{\left(1 - \frac{\lambda}{c\mu}\right)\sum_{i=0}^{c-1}\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c}.$$

For workstation m we may write

$$E(S_m(k)) = \Pi_m(k-1) \cdot \frac{1}{c_m\mu_m} + \left(E(L_m(k-1)) - \frac{\Lambda_m(k-1)}{\mu_m}\right) \cdot \frac{1}{c_m\mu_m} + \frac{1}{\mu_m}, \tag{9}$$

where $\Pi_m(k-1)$ is the probability that all machines in workstation m are busy, in a network with population $k-1$. Instead of trying to compute $\Pi_m(k-1)$ exactly, we approximate this probability by the probability Π_W that all servers are busy in an $M/M/c_m$ system with arrival rate $\Lambda_m(k-1)$ and service rate μ_m. Hence,

$$\Pi_m(k-1) \approx \frac{\frac{1}{c_m!}\left(\frac{\Lambda_m(k-1)}{\mu_m}\right)^{c_m}}{\left(1 - \frac{\Lambda_m(k-1)}{c_m\mu_m}\right)\sum_{i=0}^{c_m-1}\frac{1}{i!}\left(\frac{\Lambda_m(k-1)}{\mu_m}\right)^i + \frac{1}{c_m!}\left(\frac{\Lambda_m(k-1)}{\mu_m}\right)^{c_m}}. \tag{10}$$

For $c_m = \infty$, relation (9) simplifies to

$$E(S_m(k)) = \frac{1}{\mu_m}. \tag{11}$$

Summarizing, in an exponential multi-server network we can compute approximations for $E(S_m(k))$, $\Lambda_m(k)$, $\Pi_m(k)$ and $E(L_m(k))$ by using the recursive relations (7)-(11). The computational complexity of this scheme is (roughly) the same as the one for the single-server network.

15.4 Incorporation of material handling

As we discussed in the previous chapter transportation delays can be modelled by infinite (or ample) server stations. Here the situation is even a bit simpler. Let T_{mn} denote the mean transportation time from station m to station n. The only thing we have to change in the mean value scheme is the computation of the mean cycle time $E(C(k))$. The mean total transportation delay in a cycle has to be added to the mean cycle time. So we get

$$E(C(k)) = \sum_{n=1}^{M} v_n E(S_n(k)) + \sum_{n=1}^{M} \sum_{l=1}^{M} v_n p_{nl} T_{nl},$$

and thus

$$\Lambda_m(k) = \frac{k v_m}{E(C(k))} = \frac{k v_m}{\sum_{n=1}^{M} v_n E(S_n(k)) + \sum_{n=1}^{M} \sum_{l=1}^{M} v_n p_{nl} T_{nl}}.$$

15.5 General closed multi-server queueing network model

In this section we consider the situation where the processing times have a general distribution. Let $E(B_m)$ and $E(R_m)$ denote the mean processing time and mean residual processing time in work station m. Also the routing may be non-Markovian (but for example, fixed); only the relative visiting frequencies v_m matter. Below we will present an approximative mean value analysis for this general network.

The only relations we have to modify are relation (6) for a single-server station and relation (9) for a multi-server station. Let us first consider a single-server work station. Then we can mimic the arrival relation for the $M/G/1$ queue, yielding

$$E(S_m(k)) = \rho_m(k-1)E(R_m) + (L_m(k-1) - \rho_m(k-1)) \cdot E(B_m) + E(B_m),$$

where $\rho_m(k-1)$ is the occupation rate of station m in a network with population $k-1$, i.e.,

$$\rho_m(k-1) = \Lambda_m(k-1) \cdot E(B_m).$$

In case of a multi-server work station we modify (9) into

$$E(S_m(k)) = \Pi_m(k-1) \cdot \frac{E(R_m)}{c_m} + (E(L_m(k-1)) - \Lambda_m(k-1)E(B_m)) \cdot \frac{E(B_m)}{c_m} + E(B_m).$$

Example 15.1 Let us consider the production system in figure 2.

Station 1 is the Load/Unload (LU) station and there are 10 circulating pallets. The mean processing times and the squared coefficients of variation of the processing times are listed in table 1.

The throughput of this system is the number of parts released from the LU station per time unit, i.e., it is equal to $\Lambda_1(10)$. The approximate mean value analysis predicts a throughput of 0.736 parts per time unit; simulation yields a throughput of 0.743 ± 0.003

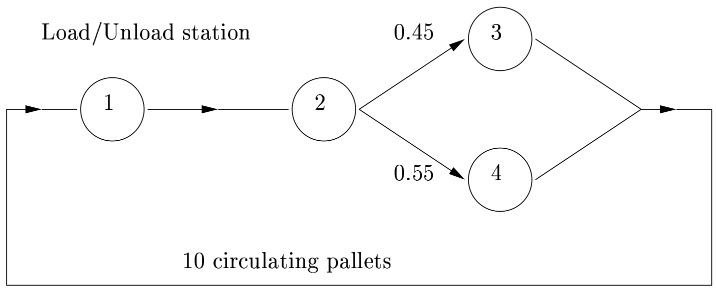

Figure 2: Closed production system with 4 stations and 10 circulating pallets

Station	$E(B_m)$	$c^2_{B_m}$
1	1.25	0.25
2	1.25	0.50
3	2.00	0.33
4	1.60	1.00

Table 1: Processing characteristics

parts per time unit (0.003 is the width of the 95% confidence interval). In table 2 we compare the results for the mean production lead times.

We may conclude that the results produced by the approximate mean value analysis are quite accurate, definitely from a practical point of view. Further observe that the throughput is predicted more accurately than the mean production lead times (typically due to cancellation of errors).

Example 15.2 Let us consider a production system with C machines and N pallets. For processing jobs they have to be equipped with tools. The number of operations to be performed is M; each operation requires a specific tool set. Label these tool sets $1, 2, \ldots, M$. Tool set m has r_m copies, and they can be assigned to c_m ($c_m \leq r_m$) machines, such that all these c_m machines are functionally identical. The relative workload to be handled by tool set m is $v_m E(B_m)$, which is known in advance. We want to assign tool sets to machines such that the throughput $TH(c_1, c_2, \ldots, c_m)$ is maximized. Hence, the optimization problem is:

$$\begin{aligned} &\max TH(c_1, c_2, \ldots, c_m) \\ &\text{subject to} \\ &\sum_{m=1}^{M} c_m \leq C, \\ &1 \leq c_m \leq r_m, \quad m = 1, 2, \ldots, M. \end{aligned}$$

A heuristic solution for this optimization problem may be found by subsequently allocating tool sets to machines; the tool set allocated to the next machine is the one yielding

Station	$E(S_m(10))$	
	amva	sim
1	4.417	4.890 $\pm$ 0.106
2	5.050	4.760 $\pm$ 0.169
3	4.181	3.860 $\pm$ 0.068
4	4.086	3.790 $\pm$ 0.118

Table 2: Comparison of results produced by the approximate mean value analysis (amva) and simulation (sim)

maximum increase in throughput.

15.6 Closed network model with multiple visits to work stations

Now we look at the situation where jobs can make several visits to the same work station, each visit involving a different type of operation, and thus a different processing time.

Let n_m be the number of distinct types of operations at work station m and let v_{mr} be the mean number of visits to work station m for a type r operation, $1 \leq r \leq n_m$. The mean processing time for a type r operation at work station m is denoted by $E(B_{mr})$; the mean residual processing time is $E(R_{mr})$. In each work station there is exactly one machine.

To formulate approximate mean value relations we first introduce some notation.

$$
\begin{aligned}
E(S_{mr}(k)) &= \text{mean production lead time at work station } m \\
&\quad\ \text{for a job receiving a type } r \text{ operation;} \\
\Lambda_{mr}(k) &= \text{arrival rate at station } m \text{ of jobs for} \\
&\quad\ \text{their type } r \text{ operation;} \\
E(L_{mr}(k)) &= \text{mean number of jobs at station } m \text{ (waiting or in service)} \\
&\quad\ \text{for their type } r \text{ operation, in a network with population } k.
\end{aligned}
$$

For the mean production lead time we get

$$E(S_{mr}(k)) = \sum_{s=1}^{n_m} \rho_{ms}(k-1)E(R_{ms}) + \sum_{s=1}^{n_m}(E(L_{ms}(k-1)) - \rho_{ms}(k-1))E(B_{ms}) + E(B_{mr}), \quad (12)$$

where $\rho_{ms}(k-1)$ denotes the occupation rate for type s operations in work station m, so

$$\rho_{ms}(k-1) = \Lambda_{ms}(k-1)E(B_{ms}).$$

For the throughput and mean number of jobs we have

$$
\begin{aligned}
\Lambda_{mr}(k) &= \frac{k v_{mr}}{\sum_{n=1}^{M}\sum_{s=1}^{n_m} v_{ns}E(S_{ns}(k))}, & (13) \\
E(L_{mr}(k)) &= \Lambda_{mr}(k)E(S_{mr}(k)), & (14)
\end{aligned}
$$

for all m and r. Hence, the means $E(S_{mr}(k))$, $\Lambda_{mr}(k)$ and $E(L_{mr}(k))$ can be computed recursively from (12)-(14) for populations starting from $k = 0$ up to $k = N$.

15.7 Closed queueing network model with multiple job types

We consider a production system consisting of M work stations, numbered $1, 2, \ldots, M$. The number of machines in workstation m is c_m. The production system is processing R types of jobs, labeled $1, 2, \ldots, R$; typically, the number of job types may be very large. The number of circulating type r jobs is fixed and equal to N_r, $r = 1, 2, \ldots, R$. The processing times at work station m are exponentially distributed with mean $1/\mu_m$ (so they are *job-type independent*), and the processing order is FCFS. The routing of jobs through is system is Markovian: after visiting work station m, a type r job moves to station n with probability p^r_{mn} (so $\sum_{n=1}^{M} p^r_{mn} = 1$). Thus each job type has its own routing. Let P^r denote the matrix of routing probabilities p^r_{mn}; we assume that P^r is irreducible (so a type r job can reach from each station any other station in one or more transitions).

The state description of the multiple job-type system is more complicated than the single-job type system. The state vector is $\underline{k} = (\underline{k}_1, \underline{k}_2, \ldots, \underline{k}_M)$ where subvector $\underline{k}_m$ describes the (aggregate) situation at work station m; that is, $\underline{k}_m = (k_{m1}, k_{m2}, \ldots, k_{mR})$ with k_{mr} indicating the number of type r jobs in work station m. Note that the stochastic process with states $\underline{k}$ is *not* a Markov process; to predict the future at time t we actually have to know the *exact order of jobs* at each work station (not only their number), since the routing is *job-type dependent*.

Let v_{mr} be the relative visiting frequency of type r jobs to work station m. For each job type r, the frequencies $v_{1r}, v_{2r}, \ldots, v_{Mr}$ satisfy the set of equations

$$v_{mr} = \sum_{n=1}^{M} v_{nr} p^r_{nm}, \qquad m = 1, 2, \ldots, M.$$

Together with a normalization equation such as $v_{1r} = 1$, this set of equations has a unique solution. It can be shown that the equilibrium probabilities $p(\underline{k})$ have a product form solution. That is,

$$p(\underline{k}) = C p_1(\underline{k}_1) p_2(\underline{k}_2) \cdots p_M(\underline{k}_M), \qquad \underline{k}_r \geq 0, \quad \sum_{m=1}^{M} k_{mr} = N_r, \quad r = 1, \ldots, R,$$

where, if $c_m = 1$,

$$p_m(\underline{k}_m) = \frac{(k_{m1} + k_{m2} + \cdots + k_{mR})!}{k_{m1}! k_{m2}! \cdots k_{mR}!} \left(\frac{v_{m1}}{\mu_m}\right)^{k_{m1}} \left(\frac{v_{m2}}{\mu_m}\right)^{k_{m2}} \cdots \left(\frac{v_{mR}}{\mu_m}\right)^{k_{mR}}$$

and if $c_m > 1$, we have to multiply this product with the extra factor

$$\frac{1}{v_m(1) v_m(2) \cdots v_m(k_{m1} + k_{m2} + \cdots + k_{mR})},$$

where $v_m(i) = \min(i, c_m)$.

Although mathematically elegant, this product form solution is not practical for the computation of mean performance characteristics. But based on this solution it can be shown that the following generalization of the Arrival Theorem holds: an arbitrary type r job moving from one station to another sees the system *in equilibrium* corresponding to a population with *one job of his own type less*. Using this theorem and Little's law we can formulate a recursive scheme to efficiently compute mean performance characteristics.

15.8 Mean value analysis for closed networks with multiple job types

We present an exact mean value analysis for single-server exponential networks with multiple job types. The derivation of approximate mean value schemes for multi-server general networks proceeds along the same lines as for networks with one job type.

Let $\underline{N} = (N_1, N_2, \ldots, N_R)$ denote the population vector and define

$$\begin{aligned} E(S_{mr}(\underline{N})) &= \text{mean production lead time at work station } m \text{ for a type } r \text{ job;} \\ \Lambda_{mr}(\underline{N}) &= \text{arrival rate at station } m \text{ of type } r \text{ jobs;} \\ E(L_{mr}(\underline{N})) &= \text{mean number of type } r \text{ jobs at station } m \text{ (waiting or in service,)} \\ & \quad\ \text{in a network with population vector } \underline{N}. \end{aligned}$$

By virtue of the Arrival Theorem we have

$$E(S_{mr}(\underline{N})) = \left(\sum_{s=1}^{r} E(L_{ms}(\underline{N} - \underline{e}_r)) + 1 \right) \frac{1}{\mu_m}, \qquad m = 1, 2, \ldots, M, \quad r = 1, 2, \ldots, R, \tag{15}$$

and application of Little's law yields

$$\Lambda_{mr}(\underline{N}) = \frac{N_r v_{mr}}{\sum_{n=1}^{M} v_{nr} E(S_{nr}(\underline{N}))}, \tag{16}$$

$$E(L_{mr}(\underline{N})) = \Lambda_{mr}(\underline{N}) E(S_{mr}(\underline{N})). \tag{17}$$

The relations (15)-(17) can be used to recursively compute the means, starting with a population vector $\underline{k} = \underline{0}$ up to $\underline{k} = \underline{N}$. The number of recursion steps, however, is equal to

$$\prod_{r=1}^{R} (1 + N_r),$$

and this number explodes when the number of job types becomes large.

The recursion in (15), which is due to the Arrival Theorem, can be avoided by formulating approximate *fixed point equations*; see [8].

As approximation we assume that an arriving type r job sees the system in equilibrium with a population $\underline{N}$ (instead of $\underline{N} - \underline{e}_r$). Thus the mean number of jobs seen on arrival is the mean number in a system *including himself*. But of course, the job does not have to wait for himself. Therefore, to avoid self queueing, the mean number $E(L_{mr}(\underline{N}))$ is multiplied with $(N_r - 1)/N_r$. This results in

$$E(S_{mr}(\underline{N})) = \sum_{s \neq r} E(L_{ms}(\underline{N}))\frac{1}{\mu_m} + \frac{N_r - 1}{N_r} E(L_{mr}(\underline{N}))\frac{1}{\mu_m} + \frac{1}{\mu_m}, \quad m = 1, 2, \ldots, M, \quad r = 1, 2, \ldots, R. \tag{18}$$

The relations (16)-(18) form a set of $3MR$ fixed point equations for $3MR$ unknowns, namely the means $E(S_{mr}(\underline{N}))$, $\Lambda_{mr}(\underline{N})$ and $E(L_{mr}(\underline{N}))$. The solution may be found by successive substitutions. In practice successive substitutions converges fast. In theory, however, convergence and uniqueness of the solution of (16)-(18) is still an open problem. The accuracy of the approximate mean value analysis is typically $5 - 10\%$ for the throughputs $\Lambda_{mr}(\underline{N})$, and $15 - 30\%$ for the means $E(S_{mr}(\underline{N}))$ and $E(L_{mr}(\underline{N}))$.

References

[1] J.P. Buzen, *Computational algorithms for closed queueing networks with exponential servers*, CACM, 16 (1973), pp. 527–531.

[2] W.T. Gordon, G.F. Newell, *Closed queueing systems with exponential servers*, Operations Res., 15 (1967), pp. 252–265.

[3] J.R. Jackson, *Networks of waiting lines*, Mgmt. Sci., 10 (1963), pp. 131–142.

[4] S.S. Lavenberg, M. Reiser, *Stationary state probabilities of arrival instants for closed queueing networks with multiple customer types*, J. Appl. Prob., 17 (1980), pp. 1048–1061.

[5] M. Reiser, *Numerical methods in separable queueing networks*, Studies in Manage. Sci., 7 (1977), pp. 113–142.

[6] M. Reiser, S.S. Lavenberg, *Mean value analysis of closed multichain queueing networks*, J. ACM, 22 (1980), pp. 313–322.

[7] M. Reiser, *Mean value analysis and convolution method for queue-dependent servers in closed queueing networks*, Perf. Eval., 1 (1981), pp. 7–18.

[8] P.J. Schweitzer, A. Seidmann and S. Shalev-Oren, *The correction terms in Approximate Mean Value Analysis*, Opns. Res. Letters, 4 (1986), pp. 197–200.