

# Closed Queuing Networks

Paul A. Jensen

Copyright August 8, 2003

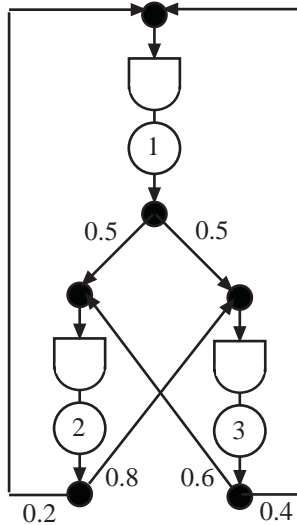


Figure 1. A closed network

A closed queuing network consists of several stations as illustrated in Fig. 1. For this structure the flow through each station is passed to other stations. It is a closed network because no flow enters from outside the network. The number of items present in the system is  $N$ .

Each component of the network is a single server queue with the service activity represented by a circle and the queue represented by the delay symbol (rectangular on one end and curved on the other). The processing rate of station  $i$  is  $\mu_i$ . The processing time is assumed to have an exponential distribution with the mean time of  $t_i = 1/\mu_i$ . Items are served with a first-come-first-served (FCFS) discipline. In general, we use  $m$  as the number of stations in the system.

We identify  $p_{ij}$  as the probability that an item leaving station  $i$  is passed to station  $j$ . The sum of the probabilities leaving a station must equal 1.  $p_{ij}$  is the transition probability from  $i$  to  $j$ . The example has three stations with the numbers on the arcs showing the transition probabilities.

We show the transition probabilities with a square matrix called the transition matrix  $\mathbf{P}$ .

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

For the closed network we define the following notation. We use  $i$  for the general station index. These quantities all depend on  $N$ , the number of items in the closed system.

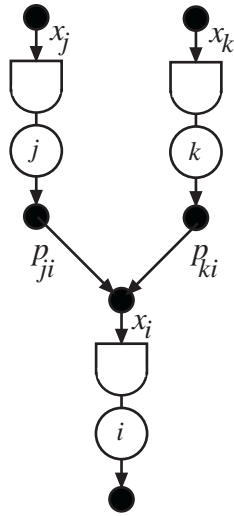
$$x_i(N) = \text{the flow through station } i. \text{ (arrival rate)}$$

$n_i(N)$  = the expected number of items at station  $i$ .

$w_i(N)$  = the expected time at station  $i$ .

The results of this section are based on *mean value analysis* (1, 2). Although the method handles more general problems than considered here, it provides a simple procedure to compute the values of flow, the expected number using each station and the expected time spent in each station. The results of this section are limited to single server queues.

## Flow



To compute the flows for a general network, we construct and solve a system of linear equations. To illustrate, consider the example in Fig. 2 representing part of a closed system. The value of  $x_i$ , the flow through station  $i$ , depends on the amounts provided by the preceding stations,  $j$  and  $k$ .

$$x_i = p_{ji}x_j + p_{ki}x_k$$

Placing all variables on the left we have:

$$x_i - p_{ji}x_j - p_{ki}x_k = 0$$

This generalizes to the expression that must hold for all stations:

$$x_i - \sum_{j=1}^m p_{ji}x_j = 0 \text{ for } i = 1 \dots m$$

Figure 2. Stations in a network

We cannot immediately solve this system of equations because the right sides are all 0. In the following we replace  $x_i$  by  $v_i$  and replace the first equation by

$$v_1 = 1$$

The values of  $v_i$  are the station flows relative to the flow in station 1. We call  $v_i$  the relative arrival rate. The set of linear equations that provides the values of the relative arrival rates is:

$$v_1 = 1$$

$$v_i - \sum_{j=1}^m p_{ji} v_j = 0 \text{ for } i = 2 \dots m$$

We construct an augmented matrix from the transition matrix.

$$\mathbf{P}_a = \begin{bmatrix} 1 & -p_{12} & -p_{13} & \cdots & -p_{1m} \\ 0 & 1 - p_{22} & -p_{23} & \cdots & -p_{2m} \\ 0 & -p_{32} & 1 - p_{33} & \cdots & -p_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -p_{m2} & -p_{m3} & \cdots & 1 - p_{mm} \end{bmatrix}$$

Define  $\mathbf{v}$  as the column vector of relative arrival rates and  $\boldsymbol{\eta}$  as a column vector with 1 as the first element and 0 as the remaining elements. Then the relative arrival rates are determined by the solution to the set of linear equations.

$$\mathbf{v}^T \mathbf{P}_a = \boldsymbol{\eta}^T \text{ or}$$

$$\mathbf{v}^T = \boldsymbol{\eta} \mathbf{P}_a^{-1} \quad (1)$$

The T superscript indicates the matrix transpose operation.

When the residence times the stations are known, the expected number in the system with a flow of 1 in station 1 is:

$$\sum_{i=1}^m v_i w_i(N) = \bar{W}(N)$$

With  $N$  items in the system we define the *system throughput* as

$$\lambda(N) = x_1(N) = \frac{N}{\bar{W}(N)} = \frac{N}{\sum_{i=1}^m v_i w_i(N)} \quad (2)$$

The flows or arrival rates at the other stations are:

$$x_i(N) = \lambda(N) v_i \text{ for } i = 2 \dots m \quad (3)$$

## Expected Number and Time

From Little's Law we compute the expected numbers at the stations:

$$n_i(N) = x_i(N) w_i(N) \text{ for } i = 1 \dots m \quad (4)$$

We have not described how to compute  $w_i(N)$ , the expected time an item will spend at station  $i$ . This is the key to the analysis. We

observe that when an item arrives at station  $i$ , the time required to pass through the station is the expected time for processing the item,  $t_i$ , plus the time required to process the items that are at the station when the item arrives. The expected value for the latter time is the expected number at station  $i$  if there were  $N - 1$  items in the system multiplied by the time required to process them.

$$\begin{aligned} w_i(N) &= t_i + n_i(N-1)t_i \\ w_i(N) &= t_i(1 + x_i(N-1)w_i(N-1)) \end{aligned} \quad (5)$$

We see that the station times for  $N$  items depend on the solution for  $N - 1$  items. To solve the equations for any  $N$ , we first observe that

$$x_i(0) = 0, n_i(0) = 0, w_i(1) = t_i \text{ for } i = 1 \dots m$$

We compute the values of  $v_i$  from Eq. 1.

The general recursive step for  $N \geq 1$  is:

- With Eq. 5 compute:

$$w_i(N) = t_i(1 + x_i(N-1)w_i(N-1)) \text{ for } i = 1 \dots m$$

- With Eq. 2 compute:  $\lambda(N)$
- With Eq. 3 compute:

$$x_i(N) = \lambda(N)v_i \text{ for } i = 1 \dots m$$

- With Eq. 4 compute:

$$n_i(N) = x_i(N)w_i(N) \text{ for } i = 1 \dots m$$

Starting from  $N = 1$ , we perform these steps sequentially for  $N = 2, 3$ , etc. until a specified value of  $N$  is reached.

At each value of  $N$  other interesting quantities describing individual stations and the system can be computed. Using the system throughput,  $\lambda(N)$ , and the other quantities already computed, we can find the following expected values (we are dropping  $(N)$  for clarity):

- Mean time in system:  $W = \frac{N}{\lambda}$
- Time in queue:  $w_{qi} = w_i - t_i$  for  $i = 1 \dots m$
- Number in the queue:  $n_{qi} = x_i w_{qi}$  for  $i = 1 \dots m$
- Number in the system queues:  $N_q = \sum_{i=1}^m n_{qi}$
- Mean time in system queues:  $W_q = \frac{N_q}{\lambda}$

- Number in service:  $n_{si} = x_i t_i$  for  $i = 1 \dots m$
- Total number in service:  $N_s = \sum_{i=1}^m n_{si}$
- Time in service:  $W_s = \frac{N_s}{\lambda}$
- Station efficiency:  $\rho_i = n_{si}$  for  $i = 1 \dots m$

## Example

To illustrate a specific case, consider the network shown in Fig. 1. We assume the processing rates:

$$\mu_1 = 1, \mu_2 = 0.5, \mu_3 = 0.2$$

The associated processing times are:

$$t_1 = 1, t_2 = 2, t_3 = 5$$

The network of Fig.1 has the  $\mathbf{P}$  matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.2 & 0 & 0.8 \\ 0.4 & 0.6 & 0 \end{bmatrix}$$

We first compute the relative arrival rates using Eq. 1.

$$\mathbf{P}_a = \begin{bmatrix} 1 & -0.5 & -0.5 \\ 0 & 1 & -0.8 \\ 0 & -0.6 & 1 \end{bmatrix}$$

$$\mathbf{v}^T \mathbf{P}_a = \boldsymbol{\eta}^T$$

$$v_1 = 1, v_2 = 1.538, v_3 = 1.731$$

With  $N = 0$ , all stations are empty.

$$n_1(0) = 0, n_2(0) = 0, n_3(0) = 0$$

$N = 1$ .

- With Eq. 5 compute the expected station times:

$$w_1(1) = 1, w_2(1) = 2, w_3(1) = 5$$

- With Eq. 2 compute the system throughput:  $\lambda(1) = 0.079$

- With Eq. 3 compute the station flows;

$$x_1(1) = 0.079, x_2(1) = 0.121, x_3(1) = 0.136$$

- With Eq. 4 compute the expected numbers at each station:  
 $n_1(1) = 0.079, n_2(1) = 0.242, n_3(1) = 0.680$

$N = 2.$

- With Eq. 5 compute the expected station times:  
 $w_1(2) = 1.079, w_2(2) = 2.483, w_3(2) = 8.399$
- With Eq. 2 compute the system throughput:  $\lambda(2) = 0.103$
- With Eq. 3 compute the station flows;  
 $x_1(2) = 0.103, x_2(2) = 0.158, x_3(2) = 0.178$
- With Eq. 4 compute the expected numbers at each station:  
 $n_1(2) = 0.111, n_2(2) = 0.393, n_3(2) = 1.496$

The table below shows the results for the example with the value of  $N$  ranging from 1 to 10. As the number in the system increases, the throughput also increases, but at a slower and slower rate. For the higher values, the congestion in the system causes the system time to increase as more and more items spend time in the queues.

$N$	$n_1(N)$	$n_2(N)$	$n_3(N)$	Throughput	System Time
1	0.079	0.242	0.680	0.07855	12.731
2	0.111	0.393	1.496	0.10290	19.436
3	0.123	0.476	2.400	0.11113	26.996
4	0.128	0.518	3.354	0.11399	35.091
5	0.130	0.537	4.333	0.11500	43.478
6	0.130	0.546	5.324	0.11536	52.012
7	0.131	0.549	6.320	0.11549	60.614
8	0.131	0.551	7.319	0.11553	69.246
9	0.131	0.551	8.318	0.11555	77.891
10	0.131	0.552	9.318	0.11555	86.541

In practice one might want to select the number in the system that provides a given throughput. For the example if the desired throughput were 0.115, we would provide 5 items in the closed system.

## Queue Add-in

The *Queue Add-in* creates models of closed queuing systems with the *Closed Network* menu command. The mean value analysis is performed with the *Solve Closed* option on the menu.

## References

1. Bolch, G., S. Greiner, H. deMeer, K.S Trivedi, *Queueing Networks and Markov Chains*, John Wiley and Sons, 1998, Section 8.2
2. Reiser, M., S. Lavenberg, “Mean Value Analysis of Closed Multichain Queuing Networks”, J of ACM, 27(2), 3113 – 322, April 1980