Mathematica Balkanica

Mathematical Society of South-Eastern Europe
A quarterly published by
the Bulgarian Academy of Sciences – National Committee for Mathematics

---

Mathematica
Balkanica

New Series Vol. 2, 1988, Fasc. 1

Resource Sharing among N Conflicting Queues – Queue Size Distribution and Response Measures

Theodore K. Apostolopoulos

Presented by St. Negrepontis

In this paper a queueing model where N queues compete for the control of a common server is examined. A very successful approximate method for this contention scheme which reduces significantly the enormous state space required for the complete description is presented. The steady-state queue size distribution using the simple iterative scheme is obtained. Besides, based on this distribution various response measures are evaluated.

1. Introduction

Analyses and modelling of computer systems, computer networks and data communication networks, by using a network of queues have received increasing attention in the last years. In such situations it is important to know system utilization, queue size distributions, delay, etc. Open and closed queueing network models have been usually used in order to analyze computer networks and computer systems, respectively.

Many analytical models admit closed form solutions or efficient computational procedures, which provide insight into the mechanisms underlying a system. So, performance evaluation is not just determining whether or not a system meets certain objectives; it is also clarifying if and how system performance can be improved. Queueing network models, in particular, are well suited as computer system and computer network models because they are able to capture the interaction between the workload and the resources of the system.

A significant breakthrough occurred when J. R. Jackson developed a solution technique for open queueing networks [3]. In addition, W. J. Gordon and G. J. Newell obtained the equilibrium state probabilities for closed queueing networks [2]. The above results have been used extensively in order to analyze computer systems and computer networks.

The present work was motivated by the problem of investigating the behaviour of data communication and computer network systems. In such situations, whenever a given station attempts transmission, the attempt may be unsuccessful because of interference from another station. In this case, a retransmission procedure occurs. The fact that the activity at one queue affects the behaviour of other queues gives rise to statistical dependence among the queues.

Models for the analysis of computer networks of this type have been presented in [1,5].

In this paper, we examine the case where a common server is shared among N queues in such a way that at any time at most one queue is served (see Fig. 1). The service rights are obtained according to specific rules. These rules for the control of the server may be considered to be distributed. That is, there is not any central scheduler to share the server among the queues, but any non-empty queue attempts to control the server according to predefined rules regardless of the actions of the rest queues. The interaction among the conflicting queues results in strong interdependence among the queues. In this paper, we propose a simple way to decompowe the interrelated queues to single queues expressing the interdependence through a certain number of parameters. The number of parameters equals the number of queues. We conclude to a simple iterative scheme capable to handle the above-described situation. Our approach has checked by analytical results when the number of conflicting queues is equal to 2 and using simulation results when the number of queues is greater than 2.

In the next section we give the formulation of the problem, and we include all the necessary assumptions. In section 3 we evaluate the queue size distribution, while in section 4 various response measures are computed. Finally, in section 5 a discussion concerning our approach and some numerical results are presented.

2. Formulation of the problem

We consider N queues, which are serviced by a common server. We will use a discrete time approach so the time is supposed to be slotted. The slot length is the time required by any job of each particular queue to receive service, if this queue has the control of the server. Each individual job needs service of one unit time (one slot) regardless of which queue it belongs. The i-th queue has a waiting room for at most M_i jobs, and receives jobs with a probability λ_i per slot, that is, the interarrival time is supposed to be geometrically distributed with a mean $1/\lambda_i$.

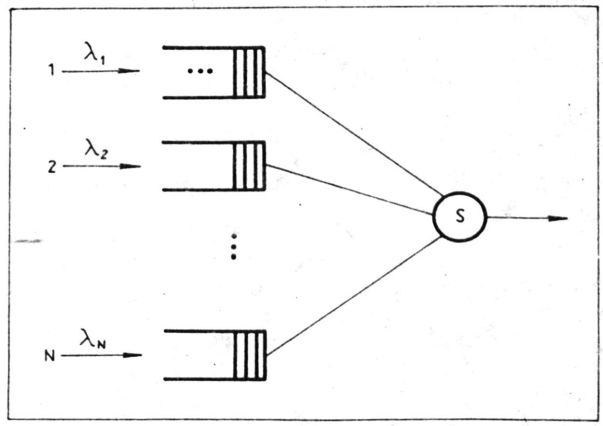

Figure 1. N queues with a common server

We assume that any non-empty queue which in the sequel we will call active queue, may acquire the control of the server at the beginning of any slot with a probability $p(i)$, where i is the number of active queues at the beginning of the corresponding slot. So, the total probability that a job belonging to anyone of the i active queues receives service is $ip(i)$, and the probability that the server is wasted when there are i active queues is $1 - ip(i)$. The dynamic definition of the probability that a queue acquire the control of the server permits the system to adjust to instantaneous load conditions. Obviously, a good policy concerning the probabilities $p(i)$ must obey the following rule: As the number of active queues decreases, the probability $p(i)$ increases, and as the number of active queues increases, the probability $p(i)$ decreases in such a way that $ip(i)$ approaches unity. Such a choice for the probabilities $p(i)$ insures low delay in the case of light load conditions and high departure rate in the case of heavy load conditions.

In general, we can describe the state of the above defined system by a vector $\mathbf{n}=(n_1, n_2, \ldots, n_i \ldots, n_N)$, where n_i is the number of jobs waiting in the room of the i-th queue, $i=1, 2, 3, \ldots, N$. Using this vector as state description vector, we can construct and solve a discrete-time Markov chain in order to have the queue size distribution of the system. But the number of possible states is too large to permit a tractable numerical analysis, except for very small values of the number of queues N and the waiting room sizes M_i. Therefore, approximative methods are mandatory.

We can simplify the above situation, if we impose a rather weak constraint, which is typical in practical cases, that all queues have equal waiting rooms of size M, that is, $M_i=M$ for $i=1, 2, \ldots, N$, and that the arrival probability of a new job at any queue is independent of each individual queue and its value is λ, that is, $\lambda_i = \lambda$ for $i=1, 2, \ldots, N$.

Under the above considerations, every queue is statistically identical with the rest queues as far as the input and the output process is concerned. The probability that an active queue has the control of the server is independent of the number of the jobs waiting in the queue and depends only on the number of active queues. It is obvious that the number of jobs waiting in a queue does describe the queueing behaviour of every other queue correctly. The rationale for this property lies in the fact that all queues have the same chance to receive a new job or to service one. So, if we can analyze an isolated queue, we can have estimates for the behaviour of the whole system.

We observe that each individual queue is affected upon only by the number i of active queues and not by the number of jobs waiting in each active queue. So, we can choose the vector $\mathbf{s}=(i, j)$ as state description vector, where i is the number of active queues and j is the number of jobs waiting at a tagged queue, say queue 1, at the beginning of a slot. With this state description the state space is significantly decreased. So, if we can construct a Markov chain describing the dynamic behaviour of the system, we can calculate the joint probability distribution of the number of active queues and the number of jobs waiting in a queue. In order to do so only one problem remains: the description of the evolution of the number of active queues if another queue, except the tagged one, has the control of the common server. In the next section we propose a simple way to overcome this difficulty and we evaluate the queue size distribution of the tagged queue.

3. The queue size distribution

The stochastic process $\{s^t, t=0, 1, 2, \ldots\}$ describes the dynamic behaviour of our system. In order to construct a Markov chain out of this stochastic process we examine the state of the system at the beginning of a slot, and we make a further assumption which will be justified under the assumption that all queues have statistically the same behaviour.

We must differentiate among the following events, concerning the control of the server.

Event 1: the tagged queue acquires the control of the server.

Event 2: another queue, except the tagged one, acquires the control of the server.

Event 3: none queue acquires the control of the server.

Similarly, we must differentiate between the following events, concerning the input process.

Event 4: the tagget queue receives or not a new job.

Event 5: k empty queues among the rest become active.

Finally, in the case when another queue except the tagged one acquires the control of the server we distinguish between the following events.

Event 6: after service completion the queue becomes empty.

Event 7: after service completion the queue remains active.

Given the state (i, j) of the system we have

$$\begin{aligned}
&\Pr\{\text{Event } 1\} = p(i)(1-\delta_{j0}) \\
&\Pr\{\text{Event } 2\} = ip(i)\delta_{j0} + (i-1)p(i)(1-\delta_{j0}) \\
&\Pr\{\text{Event } 3\} = 1 - ip(i) \\
&\Pr\{\text{Event } 4\} = \lambda\delta_{jj+1} + (1-\lambda)\delta_{jj} \\
&\Pr\{\text{Event } 5\} = b(N-i,\ k,\ \lambda)(1-\delta_{j0}) + b(N-1-i,\ k,\ \lambda)\delta_{j0} \\
&\Pr\{\text{Event } 6\} = t_0(i) \\
&\Pr\{\text{Event } 7\} = 1 - t_0(i),
\end{aligned} \tag{1}$$

where δ_{ij} is the Kronecker delta

$$\delta_{ij} = \begin{cases} 1 & \text{for } i=j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

and $b(N, i, \lambda)$ is the binomial distribution density function, that is,

$$b(N, i, \lambda) = \begin{cases} \binom{N}{i} \lambda^i (1-\lambda)^{N-i} & \text{for } i=0, 1, 2, \ldots, N \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The probability $t_0(i)$ is the probability that a job departure from an active queue, when there are i active queues at the beginning of the previous slot, leaves the queue empty. We introduce these probabilities (assuming that they exist) in order to reduce the general stochastic process $\{\mathbf{s}', t=0, 1, 2,\ldots\}$ into a Markov chain since the state of the system at the beginning of the $(t+1)$-th slot depends only on the state of the system at the beginning of the t-th slot and on the events occuring during the t-th slot. These probabilities are the same for all queues because the behaviour of all queues is statistically the same. So, they can be computed from our tagged queue. In the sequel, we will present the transition matrix P for the Markov chain, and we will compute the probabilities $t_0(i)$.

Using the above-defined probabilities for the events 1-7 given by (1), the transition probabilities $p_{ss'}$ from state $\mathbf{s}=(i, j)$ to state $\mathbf{s}'=(i', j')$ can be derived by the relations

$$p_{(i,j)(i',j')}=p(i)b(N-i,\ i'-i,\ \lambda)(\lambda\delta_{j',j}+(1-\lambda)\delta_{j',j-1})$$

$$(4)\quad +(i-1)p(i)[t_0(i)b(N-i,\ i'-i+1,\ \lambda)+(1-t_0(i))b(N-i,\ i'-i,\ \lambda)](\lambda\delta_{j,j+1}$$

$$+(1-\lambda)\delta_{j',j})+(1-ip(i))b(N-i,\ i'-i,\ \lambda)(\lambda\delta_{j',j+1}+(1-\lambda)\delta_{j',j}),$$

$$\text{for } i=1,\ 2,\ldots,\ N \text{ and } j=1,\ 2,\ldots,\ M-1;$$

$$p_{(i,1)(i',j')}=p(i)[\lambda\delta_{j',1}b(N-i,\ i'-i,\ \lambda)+(1-\lambda)\delta_{j',0}b(N-i,\ i'-i+1),\ \lambda)]$$

$$(5)\quad +(i-1)p(i)[t_0(i)b(N-i,\ i'-i+1,\ \lambda)+(1-t_0(i))b(N-i,\ i'-i,\ \lambda)](\lambda\delta_{j',j+1}$$

$$+(1-\lambda)\delta_{j',j})+(1-ip(i))b(N-i,\ i'-i,\ \lambda)(\lambda\delta_{j',j+1}+(1-\lambda)\delta_{j,j}),$$

$$\text{for } i=1,\ 2,\ldots,\ N;$$

$$p_{(i,m)(i',j')}=p(i,j)b(N-i,\ i'-i,\ \lambda)\delta_{j',M-1}+(i-1)p(i)[t_0(i)b(N-i,\ i'-i+1,\ \lambda)$$

$$(6)\quad +(1-t_0(i)b(N-i,\ i'-i,\ \lambda)]\delta_{j',M}+(1-ip(i))b(N-i,\ i'-i,\ \lambda)\delta_{j',M},$$

$$\text{for } i=1,\ 2,\ldots,\ N;$$

$$p_{(i,0)(i',j')}=ip(i)[t_0(i)[b(N-i-1,\ i'-i,\ \lambda)\lambda\delta_{j',1}$$

$$+b(N-i-1,i'-i+1,\lambda)(1-\lambda)\delta_{j',0}]+(1-t_0(i))[b(N(N-i-1,\ i'-i-1,\lambda)\lambda\delta_{j,1}$$

$$(7)\ +b(N-i-1,i'-i,\lambda)(1-\lambda)\delta_{j',0}]]+(1-ip(i))[b(N-i-1,i'-i-1,\lambda)\lambda\delta_{j',1}$$

$$+b(N-i-1,\ i'-i,\ \lambda)(1-\lambda)\delta_{j',0},$$

$$\text{for } i=1,\ 2,\ldots,\ N;$$

$$(8)\qquad p_{(0,0)(i',j')}=b(N-1,\ i'-1,\ \lambda)\lambda\delta_{j',1}+b(N-1,\ i',\ \lambda)(1-\lambda)\delta_{j',0}.$$

We will evaluate the probabilities $t_0(i)$ from our tagged queue for which we have the evolution of the number of jobs in its waiting room. This probability is given by

$$t_0(i)=(1-\lambda)\pi_{(i,1)}/\sum_{j=1}^{M}\pi_{(i,j)}, \tag{9}$$

where $\pi_{(i,j)}$ is the steady-state probability that the system is in state $\mathbf{s}=(i, j)$.

We observe that the transition matrix P is a function of the steady-state probabilities $\pi_{(i,j)}$. Thus, the steady-state probabilities $\pi_{(i,j)}$, which are components of a row vector π, can be computed as the solution of the non-linear system:

$$\pi=\pi \,.\, P(\pi)$$
$$\Sigma\pi_{(i,j)}=1, \tag{10}$$

using the following simple iterative scheme.

a) If $M=1$, set $t_0(i)=1$, for $i=1, 2, \ldots, N$. Otherwise, if $M>1$, assign arbitrary initial values to the probabilities $t_0(i)$, $0<t_0(i)<1$.

b) Solve the system of equations (10), which in view of step a) is a system of linear equations, with respect to the steady-state probabilities $\pi_{(i,j)}$.

c) Compute new estimates for the values of the probabilities $t_0(i)$ from (9).

d) Repeat steps b) and c) until a convergence criterion is satisfied.

Thus, the satedy-state probabilities $\pi_{(i,j)}$ may be computed by the solution of a series of systems of linear equations. The choice of the initial values for the probabilities $t_0(i)$ does not seem to be crucial although a few iterations can be saved giving reasonable initial values. Obviously, if $M=1$, only one iteration is necessary.

4. Response measures

In this section we will evaluate some system response measures, namely, the mean departure rate, the average number of jobs waiting in a queue, the average number of active queues, the mean delay for a job and the blocking probability.

Mean departure rate r_{out}. The conditional probability that a job of the tagged queue leaves the system is by definition the probability $p(i)$, when the tagged queue is active and there are i active queues including the tagged one. Thus, the mean departure rate of the tagged queue r^q_{out} is given by

$$r^q_{out}=\sum_{i=1}^{N}\sum_{j=1}^{M}\mathrm{p(i)}\pi_{(i,j)}. \tag{11}$$

So, the total departure rate r_{out} is given by

$$r_{out}=N \,.\, r^q_{out}. \tag{12}$$

Average number of jobs waiting in the queue J. In order to evaluate the average queue length of any queue we choose the tagged queue. Obviously, in equilibrium, the result holds for all queues because they behave identically. Thus,

$$J = \sum_{i=1}^{N} \sum_{j=0}^{M} j\pi_{(i,j)}. \tag{13}$$

Average number of active queues Q. Obviously, the average number of active queues can be obtained from the relation

$$q = \sum_{i=0}^{N} \sum_{j=0}^{M} i\pi_{(i,j)}. \tag{14}$$

The average number of active queues can also be obtained as follows. We can consider that Q is the expected value of N "independent" processes when each process is active with probability p. In this case, we have

$$Q = N \cdot \rho, \tag{15}$$

where

$$\rho = 1 - \sum_{i=0}^{N} \pi_{(i,0)}. \tag{16}$$

Mean delay D. Using Little's theorem [4], we conclude

$$D = J/r^q_{\text{out}}. \tag{17}$$

Blocking probability P_B. The blocking probability P_B is defined as the probability of a job rejection because the waiting room is full. So, we have

$$P_B = \sum_{i=1}^{N} \pi_{(i,M)}. \tag{18}$$

Because of the finite waiting room the real input rate of a queue r^q_{in} is less than λ and is given by

$$r^q_{\text{in}} = \lambda(1 - P_B). \tag{19}$$

Besides, in equilibrium the real input rate must be equal to the output rate, that is,

$$r^q_{\text{in}} = r^q_{\text{out}}. \tag{20}$$

So, we can also use (19) and (20) in order to evaluate the mean output rate.

5. Discussion

In this section we will compare the results obtained using our model with analytical results when the number of queues which share the common server is

equal to 2, and we will dliscuss some simulation results when the number of queues is greater than 2.

We will examine the case when the number of queues is supposed to be $N=2$ with waiting rooms for $M=3$ jobs. In this case we can construct and solve a two dimensional Markov chain with state vector (j_1, j_2), where $j_i=1, 2$, is the number of jobs waiting in the i-th queue. So, we can have the exact solution and we can compare this solution with the solution obtained using our approach. The following cases are examined:

1. $\lambda=0.05$ jobs/slot
   $p(1)=1,\ p(2)=0.5$

2. $\lambda=0.05$ jobs/slot
   $p(1)=0.1,\ p(2)=0.1$

3. $\lambda=0.5$ jobs/slot
   $p(1)=1,\ p(2)=0.5$

4. $\lambda=0.5$ jobs/slot
   $p(1)=0.1,\ p(2)=0.1$

In cases 1 and 3 the probability that any queue has the control of the server is equal to 1, while in cases 2 and 4 is equal to 0.2. The results are shown and compared in Tables 1 and 2 for the queue size distribution and for the response measures, respectively.

Our results have been compared with simulation results. The simulation model which is used was of fixed time increment and the simulation run length was for 50000 time units. The simulation results show a striking agreement with our results for various values of λ and for various values of the probabilities $p(i)$.

In particular, the following cases are examined. The number of queues is supposed to be $N=10$ with waiting rooms of size $M=3$ jobs. The probabilities $p(i)$ are given by

$$p(1)=1,\ p(i)=(1-1/i)^{i-1} \qquad \text{for } i=2, 3, \ldots, N \qquad \text{(Table 3)}$$

Table 1

Cases	Steady state probabilities for $N=2$ $M=3$			
a) *, b) **	state 0	state 1	state 2	state 3
1a.	0.948682	0.051246	0.000071	$9.1.10^{-8}$
1b.	0.948682	0.051246	0.000071	$9.5.10^{-8}$
2a.	0.529713	0.278796	0.132061	0.59427
2b.	0.529713	0.278796	0.132061	0.059427
3a.	0.262820	0.416666	0.247863	0.072649
3b.	0.263145	0.416672	0.247114	0.073067
4a.	0.001976	0.019762	0.177865	0.800395
4b.	0.001976	0.019762	0.177865	0.800395

(*) results obtained using the exact analysis
(**) results obtained using our approach

Table 2

Cases	Response measures for $N=2$ $M=3$			
a) *, b) **	r_{out}	J	Q	D
1a.	0.099999	0.051388	0.102635	1.027777
1b.	0.099999	0.051388	0.102635	1.027777
2a.	0.094057	0.721203	0.940572	15.335417
2b.	0.094057	0.721203	0.940572	15.335412
3a.	0.927350	1.130341	1.474358	2.437788
3b.	0.926932	1.130100	1.473701	2.438375
4a.	0.199604	2.776679	1.996047	27.821784
4b.	0.199604	2.776679	1.996047	27.821781

(*) results obtained using the exact analysis
(**) results obtained using our approach

Table 3

$N=10$, $M=3$, $p(1)=1$, $p(i)=(1-1/i)^{i-1}$ for $i=2,\ldots,10$.

Cases		Response measures			
a) *, b) **		r_{out}	J	Q	D
$\lambda=0.01$	a.	0.099999	0.011847	0.117978	1.184757
$\lambda=0.01$	b.	0.099919	0.011801	0.117971	1.181057
$\lambda=0.025$	a.	0.249928	0.049577	0.466195	1.983682
$\lambda=0.025$	b.	0.249999	0.049575	0.466190	1.983008
$\lambda=0.05$	a.	0.401593	1.294923	6.770601	32.244664
$\lambda=0.05$	b.	0.401623	1.294982	6.770563	32.243720
$\lambda=0.1$	a.	0.388274	2.444812	9.649301	62.966010
$\lambda=0.1$	b.	0.388183	2.444892	9.649353	62.982974
$\lambda=0.25$	a.	0.387465	2.824623	9.980742	72.900043
$\lambda=0.25$	b.	0.387465	2.824623	9.980748	72.900043

Table 4
$N=10$, $M=3$, $p(i)=1/i$ for $i=1, 2,\ldots, 10$

Cases		Response measures			
a) *, b) **		r_{out}	J	Q	D
$\lambda=0.01$	a.	0.099999	0.010499	0.104946	1.049999
$\lambda=0.01$	b.	0.099953	0.010510	0.104951	1.051494
$\lambda=0.025$	a.	0.249999	0.028749	0.286340	1.149984
$\lambda=0.025$	b.	0.249793	0.028752	0.286343	1.151033
$\lambda=0.05$	a.	0.499969	0.072433	0.705876	1.448758
$\lambda=0.05$	b.	0.499999	0.072425	0.705843	1.448503
$\lambda=0.1$	a.	0.959749	0.590610	4.070083	6.153792
$\lambda=0.1$	b.	0.959740	0.590612	4.070085	6.153875
$\lambda=0.25$	a.	1.000000	2.446678	9.707716	24.466772
$\lambda=0.25$	b.	0.999999	2.446701	9.707752	24.467043

(*) results obtained using our approach
(**) simulation results

(in this case we have $ip(i)<1$ for $i=2, 3, \ldots, N$)

and

$$p(i)=1/i \qquad \text{(Table 4)}$$

(in this case we have $ip(i)=1$ for $i=1, 2, \ldots, N$).

The arrival probability λ assumes the following values: 0.01, 0.025, 0.05, 0.1, 0.25.

We observe that all the results show excellent agreement with the results obtained using our approach. This fact corroborates the accuracy of our approach, which is based on the reduction of the state space.

References

1. T. K. Apostolopoulos, E. N. Protonotarios. Queueing analysis of buffered CSMA/CD protocols. *IEEE Trans. Commun.*, Sept. 1986.
2. W. J. Gordon, G. J. Newell. Closed queueing systems with exponential servers. *Oper. Res.*, **15**, 1967.
3. J. R. Jackson. Networks of waiting lines. *Oper. Res.*, **5**, 1957.
4. L. Kleinrock. Queueing Systems: vol. I, Theory. N. Y., 1975.
5. E. Sykas, D. Karvelas, E. Protonotarios. Queueing analysis of a buffered URN scheme. *IEEE Trans. Commun.*, August 1986.

Department of Statistics and Information Science,
Athens School of Economics and Business Science,
Athens, Greece

Received 15.06.1987