1992

O.J. Boxma, J.A.C. Resing

Tandem queues with deterministic service times

Department of Operations Research, Statistics, and System Theory Report BS-R9238 December

CWI is het Centrum voor Wiskunde en Informatica van de Stichting Mathematisch Centrum
CWI is the Centre for Mathematics and Computer Science of the Mathematical Centre Foundation

Tandem Queues With Deterministic Service Times

O.J. Boxma
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands;
Faculty of Economics, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

J.A.C. Resing
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract
In this paper we consider a tandem queueing model for a sequence of multiplexers at the edge of an ATM network. All queues of the tandem queueing model have unit service times. Each successive queue receives the output of the previous queue plus some external arrivals. For the case of two queues in series, we study the end-to-end delay of a cell (customer) arriving at the first queue, and the covariance of its delay at both queues. The joint queue length process at all queues is studied in detail for the 2-queue and 3-queue cases, and we outline an approach to the case of an arbitrary number of queues in series.

1991 Mathematics Subject Classification: 60K25, 68M20.
Keywords and Phrases: ATM, tandem queue, deterministic service times, delay.
Note: Part of the first author's research has been supported by the European Grant BRA-QMIPS of CEC DG XIII. The research of the second author was done during the time that he was affiliated with CWI, in a joint project with PTT Research.

Report BS-R9238
ISSN 0924-0659
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1 Introduction

In this paper we consider a queueing model for a sequence of multiplexers at the edge of an ATM network. Our main goal is to study the influence of the interference of traffic streams on the cell delay in the multiplexers. In particular, the end-to-end delay of a cell and the covariance between the delays of a cell in successive queues are derived.

The basic model is a discrete-time model and consists of two queues in tandem, both with unit (deterministic) service time. The output of the first queue enters the second queue. In both queues there are external batch arrivals. So the input process of the second queue consists of the interference of an internal sequence of cells and an external sequence of cells (see Fig. 1).

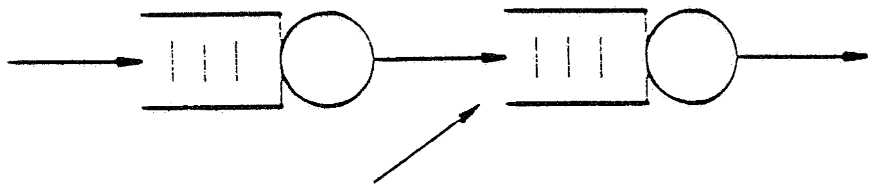


Fig. 1

External arrivals at the two queues in successive time slots are assumed to be independent and identically distributed in each time slot. However, within a time slot the arrivals at the two queues may be correlated. Cells are served in first come first served order. Furthermore we assume that, when cells arrive simultaneously at the second queue, external arrivals are served before the cell coming from the first queue.

We look at the system at discrete time points n, $n = 0, 1, 2, \ldots$. Cells arriving in the time slot $(n-1, n)$ are taken into account only at time n. The cell which is served in the time slot $(n-1, n)$ is not taken into account at time n.

We are interested in the following three quantities:

- The joint queue length process in the two queues.
- The end-to-end delay of a cell arriving at the first queue.
- The covariance between the delay in the first queue and the delay in the second queue of a cell arriving at the first queue.

In section 2 we present an expression, due to Morrison [4], for the generating function $\phi(s_1, s_2)$ of the stationary distribution of the joint queue length process. The expression for $\phi(s_1, s_2)$ is in terms of $\psi(s_1, s_2)$, the generating function of the joint arrival process, and σ, the solution for fixed s_2 of the equation $s_1 = \psi(s_1, s_2)$.

In section 3 we study the end-to-end delay of an arbitrary cell arriving at the first queue. In section 4 we use the results of section 3 to calculate the covariance between the delay in queue 1 and the delay in queue 2 of a cell arriving at the first queue. A numerical example which provides some insight into the covariance between the delays in successive queues is given. In the example, the input of the first queue consists of two independent Bernoulli streams and in the second queue there is interference of another independent Bernoulli stream.

In section 5 we compare the mean delay in the model with two queues in series with the mean delay in a corresponding model with only one queue. Because of the bursty arrival processes occurring in ATM networks, we drop the assumption of independence of arrivals in successive slots in section 6. The dependence is modelled by Markov modulation. Finally, in section 7 we extend the analysis of section 2 to a model with more than two queues in series.

We finish the introduction by giving a short review of related literature. In Morrison [4] the joint queue length process of the basic model mentioned before is analysed. In [4] only the model with two queues in tandem and independent arrivals in successive time slots is considered. In Morrison [3] so-called concentrating rooted tree networks of discrete-time single server queues with unit service time are considered. It is shown that the network of queues may be replaced by a single queue, which has the same output as the queue at the root of the tree. The result is applied to the case of several queues in tandem. This leads to an expression for the generating function of the queue length distribution in individual queues (not of the joint queue length distribution) in the case that arrivals to different queues are independent.

In Kaplan [1] and Shalmon and Kaplan [5] a related queueing system in continuous time is considered. In [1] only two queues in tandem are considered, whereas in [5] more than two queues in series are considered. The external arrivals to the different queues are given by independent Poisson processes. An expression is given for the steady-state joint Laplace Stieltjes transform of delays in the successive queues that a certain cell stream passes. As a special case this result contains the steady-state Laplace Stieltjes transform of the end-to-end delay of a cell arriving at a certain queue.

2 The joint queue length process

Let $(Y_1^{(n)}, Y_2^{(n)})$ be the number of external arrivals to the two queues in the n-th time slot with, for all n, generating function

$$\psi(s_1, s_2) = E(s_1^{Y_1^{(n)}} s_2^{Y_2^{(n)}}). \tag{1}$$

Let furthermore $(X_1^{(n)}, X_2^{(n)})$ denote the contents of the two queues at time n, with generating function

$$\phi^{(n)}(s_1, s_2) = E(s_1^{X_1^{(n)}} s_2^{X_2^{(n)}}). \tag{2}$$

If the mean total input rate is less than one, i.e. $E(Y_1^{(n)} + Y_2^{(n)}) < 1$, then one can show (see Morrison [4]) that a stationary distribution of the joint queue length process exists. Let $1[\cdot]$ denote the indicator function. From the equations

$$X_1^{(n+1)} = X_1^{(n)} - 1[X_1^{(n)} > 0] + Y_1^{(n)} \tag{3}$$

and

$$X_2^{(n+1)} = X_2^{(n)} - 1[X_2^{(n)} > 0] + 1[X_1^{(n)} > 0] + Y_2^{(n)} \tag{4}$$

one can derive that the generating function of the stationary distribution of the joint queue length process, $\phi(s_1, s_2) = \lim_{n\to\infty} \phi^{(n)}(s_1, s_2)$, satisfies

$$\begin{aligned} \phi(s_1, s_2) &= \psi(s_1, s_2)\{\frac{1}{s_1}[\phi(s_1, s_2) - \phi(s_1, 0) - \phi(0, s_2) + \phi(0, 0)] \\ &+ \frac{1}{s_2}[\phi(0, s_2) - \phi(0, 0)] + \frac{s_2}{s_1}[\phi(s_1, 0) - \phi(0, 0)] + \phi(0, 0)\}. \end{aligned} \tag{5}$$

This is the key equation from which an expression for the generating function $\phi(s_1, s_2)$ can be derived. In (5) the term $\phi(s_1, s_2) - \phi(s_1, 0) - \phi(0, s_2) + \phi(0, 0)$ corresponds to the situation where both queues are nonempty. In this case one customer goes from queue 1 to queue 2 in a time slot and one customer from queue 2 leaves the system. Hence the total result is that one customer from queue 1 leaves the system in a time slot and so we get the factor $1/s_1$. The term $\phi(0, s_2) - \phi(0, 0)$ corresponds to the situation where only the second queue is nonempty. In this case one customer from queue 2 leaves the system and so we get the factor $1/s_2$. The term $\phi(s_1, 0) - \phi(0, 0)$ corresponds to the situation where only the first queue is nonempty. In this case one customer goes from queue 1 to queue 2 and so we get the factor s_2/s_1. Finally the term $\phi(0, 0)$ corresponds to an empty sytem.

An explicit expression for the generating function $\phi(s_1, s_2)$ is given by (see Morrison [4])

$$\begin{aligned} \phi(s_1, s_2) &= [1 - \psi_1(1, 1) - \psi_2(1, 1)]\frac{(1 - s_2)\psi(s_1, s_2)}{\psi(s_1, s_2) - s_1} \cdot \\ &\quad [\frac{\psi(s_1, 0)}{\psi(1, 0)} - \frac{s_1 - s_2}{\sigma - s_2} \cdot \frac{\psi(\sigma, 0)}{\psi(1, 0)}], \end{aligned} \tag{6}$$

where $\sigma = \sigma(s_2)$ is, for fixed $|s_2| < 1$, the unique root within the unit circle of the equation $s_1 - \psi(s_1, s_2) = 0$, and the subscript i in $\psi_i(\cdot, \cdot)$ denotes differentiation with respect to the i-th variable.

In particular, for the generating function of the stationary queue length distribution in the first queue we find the well-known single queue result (see Morrison [4], formula (2.18)), and for that in the second queue we have

$$\begin{aligned}\phi(1,s_2) &= [1-\psi_1(1,1)-\psi_2(1,1)]\frac{(1-s_2)\psi(1,s_2)}{\psi(1,s_2)-1}\cdot \\ &\quad [1-\frac{1-s_2}{\sigma-s_2}\cdot\frac{\psi(\sigma,0)}{\psi(1,0)}]. \end{aligned}\tag{7}$$

For the mean contents EX_1 and EX_2 in queue 1 and queue 2, respectively, we find

$$EX_1 = EY_1 + \frac{E[Y_1(Y_1-1)]}{2(1-EY_1)},\tag{8}$$

and

$$\begin{aligned}EX_2 &= \frac{\psi_1(1,0)}{\psi(1,0)} + EY_2 + \frac{E[Y_2(Y_2-1)]}{2(1-EY_1-EY_2)} \\ &+ \frac{E[Y_1Y_2]}{1-EY_1-EY_2} + \frac{E[Y_1(Y_1-1)]}{2(1-EY_1)}\cdot\frac{EY_2}{1-EY_1-EY_2}. \end{aligned}\tag{9}$$

So far we summarized the results of Morrison [4]. In the rest of this paper we shall concentrate on the following four questions:

- What is the end-to-end delay of a cell arriving at the first queue?
- What is the covariance between the delay in queue 1 and the delay in queue 2 of this cell?
- Can one extend the result to systems where the arrivals in successive time slots are dependent?
- Can one extend the results to systems with more than two queues in series?

3 End-to-end delay

Let K be an arbitrary cell arriving at the first queue. In this section we shall concentrate on the end-to-end delay of K. As preparation, we first study the joint distribution of U_1, the number of cells arriving before K in the same time slot as K at the first queue, and U_2, the number of external cells arriving in the same time slot as K at the second queue. Let $\gamma(s_1,s_2) = E(s_1^{U_1}s_2^{U_2})$ be the generating function of the joint distribution of (U_1,U_2).

Lemma 1

$$\gamma(s_1,s_2) = \frac{\psi(1,s_2)-\psi(s_1,s_2)}{E(Y_1)(1-s_1)}.\tag{10}$$

Proof: Let U_0 be the total number of cells (including K) arriving at the first queue in the same time slot as K. Then

$$\begin{aligned} P(U_0 = k_0, U_2 = k_2) &= \frac{k_0 \cdot P(Y_1 = k_0, Y_2 = k_2)}{E(Y_1)}, \\ & \qquad k_0 = 1, 2, \ldots, \quad k_2 = 0, 1, \ldots, \end{aligned} \tag{11}$$

and

$$P(U_1 = k_1 | U_0 = k_0, U_2 = k_2) = \frac{1}{k_0}, \quad k_1 = 0, 1, \ldots, k_0 - 1. \tag{12}$$

The lemma follows by a straightforward calculation from (11) and (12).

In particular, for the first moment of U_1 and U_2 we find

$$E(U_1) = \frac{E(Y_1[Y_1 - 1])}{2E(Y_1)}, \tag{13}$$

$$E(U_2) = \frac{E(Y_1 Y_2)}{E(Y_1)}. \tag{14}$$

Next, define X_i as the number of cells present in queue $i, i = 1, 2$, at the beginning of the time slot in which K arrives. Of course, the joint generating function of (X_1, X_2) equals $\phi(s_1, s_2)$. Furthermore, define $\tilde{X}_1$ as the number of cells in queue 1 before K at the end of the time slot in which K arrives, and $\tilde{X}_2$ as the number of cells in queue 2 at the end of the time slot in which K arrives. Let $\tilde{\phi}(s_1, s_2)$ be the joint generating function of $(\tilde{X}_1, \tilde{X}_2)$.

Lemma 2

$$\begin{aligned} \tilde{\phi}(s_1, s_2) &= \gamma(s_1, s_2)\{\frac{1}{s_1}[\phi(s_1, s_2) - \phi(s_1, 0) - \phi(0, s_2) + \phi(0, 0)] \\ &+ \frac{1}{s_2}[\phi(0, s_2) - \phi(0, 0)] + \frac{s_2}{s_1}[\phi(s_1, 0) - \phi(0, 0)] + \phi(0, 0)\}. \end{aligned} \tag{15}$$

Proof: We have

$$\tilde{X}_1 = X_1 - 1[X_1 > 0] + U_1, \tag{16}$$

and

$$\tilde{X}_2 = X_2 - 1[X_2 > 0] + 1[X_1 > 0] + U_2. \tag{17}$$

Now the results follows analogous to formula (5). (Remark the resemblance between (3), (4) and (16),(17)).

Finally, define Z_i, $i = 1, 2$, as the delay of K in queue i, with joint generating function $\chi(s_1, s_2) = E(s_1^{Z_1} s_2^{Z_2})$.

Lemma 3

$$\begin{aligned}\chi(s_1,s_2) &= \frac{\psi(1,s_2)}{s_2}\tilde{\phi}(s_1\psi(1,s_2),s_2) \\ &+ (1-\frac{1}{s_2})\psi(1,s_2)\tilde{\phi}(s_1\psi(1,s_2),0).\end{aligned} \tag{18}$$

Proof: We have $Z_1 = \tilde{X}_1$ and

$$Z_2 = \tilde{X}_2 - 1[\tilde{X}_2 > 0] + \sum_{i=1}^{\tilde{X}_1+1} Y_2^{(i)}, \tag{19}$$

where $Y_2^{(i)}$, $i = 1,2,\dots$, are i.i.d. random variables with generating function $\psi(1,s)$. Equation (19) follows by changing the order of service in the second queue. In the first time slot after the time slot in which K arrives, one of the cells present in the second queue, if any, is served. In subsequent time slots cells arriving from queue 1 at queue 2 are immediately served in queue 2. Hence the only cells present in queue 2 at the arrival moment of K at queue 2, are those already present at the end of the time slot in which K arrived at queue 1, possibly -1, plus the external arrivals to queue 2 during the stay of K in queue 1. (Remember that we assume that, when cells arrive simultaneously at the second queue, external arrivals are served before the cell coming from queue 1.)

From (19) we have

$$\begin{aligned}\chi(s_1,s_2) &= \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} P(\tilde{X}_1 = x_1, \tilde{X}_2 = x_2)E(s_1^{Z_1}s_2^{Z_2}|\tilde{X}_1 = x_1, \tilde{X}_2 = x_2) \\ &= \sum_{x_1=0}^{\infty} P(\tilde{X}_1 = x_1, \tilde{X}_2 = 0)s_1^{x_1}\psi(1,s_2)^{x_1+1} \\ &+ \sum_{x_1=0}^{\infty}\sum_{x_2=1}^{\infty} P(\tilde{X}_1 = x_1, \tilde{X}_2 = x_2)s_1^{x_1}s_2^{x_2-1}\psi(1,s_2)^{x_1+1} \\ &= \frac{\psi(1,s_2)}{s_2}\tilde{\phi}(s_1\psi(1,s_2),s_2) + (1-\frac{1}{s_2})\psi(1,s_2)\tilde{\phi}(s_1\psi(1,s_2),0).\end{aligned}$$

Combination of the Lemma's 1,2 and 3 gives an expression for $\chi(s_1,s_2)$ in terms of $\phi(s_1,s_2)$ and $\psi(s_1,s_2)$. Furthermore, for the end-to-end delay $Z = Z_1 + Z_2$ with generating function $\tau(s) := E(s^Z)$, we have $\tau(s) = \chi(s,s)$.

4 Covariance between delays

The result in the previous section enables us to find an expression for the covariance between the delay in queue 1 and the delay in queue 2. We have

$$Cov(Z_1,Z_2) = Cov(\tilde{X}_1,\tilde{X}_2) + Cov(\tilde{X}_1, 1[\tilde{X}_2 = 0]) + E(Y_2)Var(\tilde{X}_1). \tag{20}$$

Of course $E(Y_2) = \psi_2(1,1)$ is a known quantity. For the other three quantities we find after tedious calculations

$$\begin{aligned} Var(\tilde{X}_1) &= \gamma_1(1,1)(1-\gamma_1(1,1)) + \frac{\frac{1}{2}\psi_{11}(1,1)}{1-\psi_1(1,1)} \\ &+ \gamma_{11}(1,1) + \frac{\frac{1}{3}\psi_{111}(1,1)}{1-\psi_1(1,1)} + \frac{\frac{1}{4}\psi_{11}(1,1)^2}{(1-\psi_1(1,1))^2}, \end{aligned}$$

$$\begin{aligned} Cov(\tilde{X}_1, 1[\tilde{X}_2 = 0]) &= \frac{1-\psi_1(1,1)-\psi_2(1,1)}{\psi(1,0)}[\gamma_1(1,0) \\ &\quad -\gamma(1,0)(\gamma_1(1,1) + \frac{\psi_{11}(1,1)}{2(1-\psi_1(1,1))})], \end{aligned}$$

and

$$\begin{aligned} Cov(\tilde{X}_1, \tilde{X}_2) &= \gamma_{12}(1,1) - \gamma_1(1,1)\gamma_2(1,1) + \frac{1}{2}\frac{\psi_{112}(1,1)}{1-\psi_1(1,1)} \\ &+ \frac{1}{6}\frac{\psi_2(1,1)\psi_{111}(1,1)}{(1-\psi_1(1,1))^2} + \frac{1}{4}\frac{\psi_2(1,1)\psi_{11}(1,1)^2}{(1-\psi_1(1,1))^3} \\ &+ \frac{1}{2}\frac{\psi_{11}(1,0)}{\psi(1,0)}\frac{1-\psi_1(1,1)-\psi_2(1,1)}{1-\psi_1(1,1)} + \frac{1}{2}\frac{\psi_{12}(1,1)\psi_{11}(1,1)}{(1-\psi_1(1,1))^2}. \end{aligned}$$

Remark that the formulas become much simpler in the case that the arrival processes to the two queues are independent.

Next we use a simple example to provide some insight into the correlation between delays in successive queues. We assume that a Bernoulli stream with parameter p passes two queues. In both queues there is interference of one other stream. These interfering streams are also Bernoulli streams with parameter q_1 and q_2 respectively, see Fig. 1 with an additional q_1 stream at the first queue.

All Bernoulli streams are independent of each other. Hence the generating function of the external arrivals is given by

$$\psi(s_1, s_2) = (1-p+ps_1)(1-q_1+q_1s_1)(1-q_2+q_2s_2).$$

We are interested in the correlation of the delays in the two queues of an arbitrary cell arriving at the first queue. For $Cov(Z_1, Z_2)$ and $Var(Z_1)$ $(= Var\tilde{X}_1)$ expressions were found before. Similarly we can find an expression for $Var(Z_2)$. For this example some results are shown in Fig. 2 and Fig. 3, see page 18. In both figures p, i.e. the load of the main stream, equals 0.25. In Fig. 2 we vary the load of the interfering stream in queue 1 for some fixed values of the load of the interfering stream in queue 2. In Fig. 3 we vary the load of the interfering stream in queue 2 for some fixed values of the load of the interfering stream in queue 1. Remark that in the figures the range over which the parameters q_1 (in

Fig. 2) and q_2 (in Fig. 3) vary, is not fixed. For example in Fig. 3 if $q_1 = 0.25$, q_2 varies from 0 to 0.5, whereas if $q_1 = 0.74$, q_2 varies from 0 to 0.01. From the figures we conclude the following:

- As expected the correlation between the delays is always positive.
- Both when q_2 is varied and when q_1 is varied the correlation first increases, then reaches a maximum and finally decreases again.
- In moderate traffic the correlation is not very large (≤ 0.25).
- When the load in queue 1 is high and the load of the interfering stream in queue 2 is small the correlation can be large (> 0.4).
- When the load in queue 2 tends to 1 the correlation tends to 0. This can be explained by the fact that the load in queue 1 stays away from 1. Therefore the delay in queue 1 remains finite, whereas the delay in queue 2 tends to infinity. Hence the delay in queue 1 has hardly any influence on the delay in queue 2.
- When the load of the interfering stream in queue 1 (resp. queue 2) equals 0 there is no delay at all in the first (resp. second) queue and hence the correlation equals 0.

5 Comparison of mean delays

In this section we compare the mean delays in the following two models. In the first model a certain stream (called a-stream) passes two queues in tandem, with an interfering stream (called b-stream) in the second queue. In the second model both the a-stream and the b-stream are input streams of a single queue. Let EZ denote the mean delay of an arbitrary cell of the a-stream in the single queueing model and let EZ_1, resp. EZ_2, denote the mean delay of an arbitrary cell of the a-stream in the first, resp. second, queue of the tandem queueing model. To make the comparison fair, we assume in both models that in the case of simultaneous arrivals, cells of the b-stream have priority above cells of the a-stream.

Lemma 4 *For the quantities EZ, EZ_1 and EZ_2, we have*

$$EZ = \frac{E[(Y_1+Y_2)(Y_1+Y_2-1)]}{2(1-EY_1-EY_2)} + \frac{E[Y_1(Y_1-1)]}{2EY_1} + \frac{E[Y_1Y_2]}{EY_1}, \tag{21}$$

$$EZ_1 = \frac{E[Y_1(Y_1-1)]}{2EY_1(1-EY_1)}, \tag{22}$$

$$EZ_2 = \frac{1-EY_2}{EY_1}[E[Y_1|Y_2=0]-EY_1] + \frac{E[Y_2(Y_2-1)]}{2(1-EY_1-EY_2)}$$
$$+\frac{E[Y_1Y_2]}{1-EY_1-EY_2}\frac{1-EY_2}{EY_1} + \frac{E[Y_1(Y_1-1)]}{2(1-EY_1-EY_2)}\frac{EY_2(1-EY_2)}{EY_1(1-EY_1)} \quad (23)$$

Proof: In the single queue model the delay Z of an arbitrary cell K equals the number of cells before K at the end of the time slot in which K arrives. For this number of cells, a formula similar to (16) can be derived with U_1 replaced by $U_1 + U_2$. Now, formula (21) follows from (8), (13) and (14). Formula (22) follows directly from (21) by putting $Y_2 = 0$. From (17) and (19) one derives

$$EZ_2 = EX_2 + EU_2 - 1 + \gamma(1,0)[\phi(0,0) + \phi_2(0,0)] + EZ_1EY_2. \quad (24)$$

Now (23) follows from (9), (10), (14) and (22) after straightforward calculations.

From Lemma 4 we conclude that

$$EZ_2 - EZ = \frac{1-EY_2}{EY_1}[E[Y_1|Y_2=0] - EY_1 - \frac{E[Y_1(Y_1-1)]}{2(1-EY_1)}] \quad (25)$$

and

$$EZ_1 + EZ_2 - EZ = \frac{1-EY_2}{EY_1}[E[Y_1|Y_2=0]-EY_1] + \frac{EY_2}{EY_1}\frac{E[Y_1(Y_1-1)]}{2(1-EY_1)}. \quad (26)$$

We conclude that in the case of *independent* arrivals (i.e. $E[Y_1|Y_2=0] = EY_1$), we have

$$EZ_2 \le EZ \le EZ_1 + EZ_2. \quad (27)$$

In the case of *dependent* arrivals, we dot not have these inequalities. For example, if $Y_1 = 0$ or 1 with probability 1 and $E[Y_1|Y_2 = 0] > EY_1$, then $EZ_2 > EZ$. On the other hand, if $Y_1 = 0$ or 1 with probability 1 and $E[Y_1|Y_2 = 0] < EY_1$, then $EZ_1 + EZ_2 < EZ$.

6 Dependent arrivals

In this section we extend the results of section 2 to a model in which the arrivals in successive time intervals may be dependent. Remember that in ATM networks we are dealing with bursty arrival processes and hence we are interested in dependence between arrivals. The dependence is modelled by Markov modulation, i.e. we assume that the number of arrivals in a time interval depends on the state of an underlying Markov chain. Let $T_n, n = 1, 2, \ldots$ be the Markov chain with state space $\{1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1-r_1 & r_1 \\ r_2 & 1-r_2 \end{pmatrix}. \quad (28)$$

When the Markov chain is in state i, the number of external arrivals to the two queues has generating function $\Psi^{(i)}(s_1, s_2)$. We are interested in the steady-state joint generating function of the queue lengths distribution when the underlying Markov chain is in state j, $j = 1, 2$, i.e.

$$\Phi^{(j)}(s_1, s_2) = \lim_{n \to \infty} E(s_1^{X_n^{(1)}} s_2^{X_n^{(2)}} 1[T_n = j]). \tag{29}$$

Define

$$\Psi(s_1, s_2) = \begin{pmatrix} (1-r_1)\Psi^{(1)}(s_1, s_2) & r_2\Psi^{(2)}(s_1, s_2) \\ r_1\Psi^{(1)}(s_1, s_2) & (1-r_2)\Psi^{(2)}(s_1, s_2) \end{pmatrix} \tag{30}$$

and

$$\Phi(s_1, s_2) = \begin{pmatrix} \Phi^{(1)}(s_1, s_2) \\ \Phi^{(2)}(s_1, s_2) \end{pmatrix}. \tag{31}$$

Then we can prove the following analog of (5) in the case of Markov modulated arrivals:

$$\begin{aligned} \Phi(s_1, s_2) &= \Psi(s_1, s_2)\{\frac{1}{s_1}[\Phi(s_1, s_2) - \Phi(s_1, 0) - \Phi(0, s_2) + \Phi(0, 0)] \\ &+ \frac{1}{s_2}[\Phi(0, s_2) - \Phi(0, 0)] + \frac{s_2}{s_1}[\Phi(s_1, 0) - \Phi(0, 0)] + \Phi(0, 0)\}. \end{aligned} \tag{32}$$

This enables us to find an expression for the vector $\Phi(s_1, s_2)$ if we can determine the boundary vectors $\Phi(s_1, 0)$, $\Phi(0, s_2)$ and $\Phi(0, 0)$. Letting $s_2 \to 0$ in (32) we obtain

$$\begin{aligned} \Phi(s_1, 0) &= \Psi(s_1, 0)[\Phi(0, 0) + \lim_{s_2 \to 0} \frac{1}{s_2}(\Phi(0, s_2) - \Phi(0, 0))] \\ &= \Psi(s_1, 0)[\Psi(0, 0)]^{-1}\Phi(0, 0). \end{aligned} \tag{33}$$

Remark that the inverse $[\Psi(0, 0)]^{-1}$ only exists when $\Psi^{(1)}(0, 0) \neq 0$, $\Psi^{(2)}(0, 0) \neq 0$ and $r_1 \neq 1 - r_2$. From now on we assume that these three conditions are fulfilled.

In order to determine the boundary vector $\Phi(0, s_2)$ we need the following lemma (see Khamisy and Sidi [2]). In [2] a single server queue with two priority classes is considered. The analysis of this model leads to a similar equation as (32) and hence several lemmas proved in [2] are also useful for the analysis of our model.

Lemma 5 *For a given $|s_2| < 1$ the following equation in s_1,*

$$\begin{aligned} s_1^2 &- s_1[(1-r_1)\Psi^{(1)}(s_1, s_2) + (1-r_2)\Psi^{(2)}(s_1, s_2)] \\ &+ (1-r_1-r_2)\Psi^{(1)}(s_1, s_2)\Psi^{(2)}(s_1, s_2) = 0 \end{aligned} \tag{34}$$

has exactly two solutions $\sigma_k = \sigma_k(s_2)$, $k = 1, 2$, in the unit circle $|s_1| < 1$.

We can rewrite (32) in the form

$$\begin{pmatrix} \Phi^{(1)}(s_1,s_2) \\ \Phi^{(2)}(s_1,s_2) \end{pmatrix} = K(s_1,s_2)^{-1} \begin{pmatrix} c_1(s_1,s_2) & r_2\Psi^{(2)}(s_1,s_2) \\ r_1\Psi^{(1)}(s_1,s_2) & c_2(s_1,s_2) \end{pmatrix} \cdot$$

$$\cdot \begin{pmatrix} \Phi^{(1)}(s_1,0) & \Phi^{(1)}(0,s_2) & \Phi^{(1)}(0,0) \\ \Phi^{(2)}(s_1,0) & \Phi^{(2)}(0,s_2) & \Phi^{(2)}(0,0) \end{pmatrix} \begin{pmatrix} \frac{s_2-1}{s_1} \\ \frac{s_1-s_2}{s_1 s_2} \\ \frac{(s_2-1)(s_1-s_2)}{s_1 s_2} \end{pmatrix}$$

where

$$\begin{aligned} K(s_1,s_2) &= det(I - \frac{\Psi(s_1,s_2)}{s_1}) \\ &= 1 - \frac{1}{s_1}[(1-r_1)\Psi^{(1)}(s_1,s_2) + (1-r_2)\Psi^{(2)}(s_1,s_2)] \\ &+ \frac{1}{s_1^2}(1-r_1-r_2)\Psi^{(1)}(s_1,s_2)\Psi^{(2)}(s_1,s_2) \end{aligned}$$

and

$$c_i(s_1,s_2) = (1-r_i)\Psi^{(i)}(s_1,s_2) - (1-r_1-r_2)\frac{1}{s_1}\Psi^{(1)}(s_1,s_2)\Psi^{(2)}(s_1,s_2). \quad (35)$$

Since $\Phi^{(1)}(s_1,s_2)$ is analytic in the disk $|s_i| < 1$, $i = 1,2$, and $K(\sigma_k, s_2) = 0$ for $k = 1,2$, we have

$$\begin{aligned} & c_1(\sigma_k,s_2)\Big\{(s_2-1)\Phi^{(1)}(\sigma_k,0) + \frac{\sigma_k - s_2}{s_2}\Phi^{(1)}(0,s_2) \\ & + \frac{(s_2-1)(\sigma_k-s_2)}{s_2}\Phi^{(1)}(0,0)\Big\} + r_2\Psi^{(2)}(\sigma_k,s_2)\Big\{(s_2-1)\Phi^{(2)}(\sigma_k,0) \\ & + \frac{\sigma_k-s_2}{s_2}\Phi^{(2)}(0,s_2) + \frac{(s_2-1)(\sigma_k-s_2)}{s_2}\Phi^{(2)}(0,0)\Big\} = 0 \end{aligned}$$

from which the boundary functions $\Phi^{(j)}(0,s_2)$, $j = 1,2$ can be expressed as functions of $\Phi^{(j)}(0,0)$, $j = 1,2$ (see also (33)). We could also use the equation for $\Phi^{(2)}(s_1,s_2)$ instead of $\Phi^{(1)}(s_1,s_2)$ to obtain the boundary functions $\Phi^{(j)}(0,s_2)$, $j = 1,2$, but this leads to the same results (i.e. the equations are linear dependent).

Finally we have to determine the constants $\Phi^{(j)}(0,0)$, $j = 1,2$. Let π_j be the limiting probability of the underlying Markov chain in state j, $j = 1,2$, i.e.

$$\pi_j = \lim_{n\to\infty} Pr(T_n = j) = \frac{r_{3-j}}{r_1 + r_2}. \quad (36)$$

Then we have the normalization condition

$$\Phi^{(1)}(1,0)+\Phi^{(2)}(1,0) = 1-\pi_1(\Psi_1^{(1)}(1,1)+\Psi_2^{(1)}(1,1))-\pi_2(\Psi_1^{(2)}(1,1)+\Psi_2^{(2)}(1,1)). \quad (37)$$

Furthermore we need the following lemma (see Khamisy and Sidi [2])

Lemma 6 *Under the assumption of steady-state, i.e.*

$$\pi_1(\Psi_1^{(1)}(1,1)+\Psi_2^{(1)}(1,1))+\pi_2(\Psi_1^{(2)}(1,1)+\Psi_2^{(2)}(1,1))<1 \tag{38}$$

the following equation in s,

$$s^2-s[(1-r_1)\Psi^{(1)}(s,s)+(1-r_2)\Psi^{(2)}(s,s)]+(1-r_1-r_2)\Psi^{(1)}(s,s)\Psi^{(2)}(s,s)=0 \tag{39}$$

has exactly two solutions 1, $\sigma=\sigma(s)$, *in the unit circle* $|s|\leq 1$.

Since $\Phi^{(1)}(s,s)$ is analytic in the disk $|s|\leq 1$ we have

$$c_1(\sigma,\sigma)\Phi^{(1)}(\sigma,0)+r_2\Psi^{(2)}(\sigma,\sigma)\Phi^{(2)}(\sigma,0)=0. \tag{40}$$

Using (33) we can rewrite (37) and (40) in two independent linear equations in the unknowns $\Phi^{(j)}(0,0)$, $j=1,2$ and hence we can determine $\Phi^{(j)}(0,0)$, $j=1,2$.

7 The model with more than two queues

In this section we show that the analysis of section 2 can be generalized to models with more than two queues in series. We mainly restrict our attention to a model with three queues. The analog of formula (5) in the case of three queues in series is

$$\begin{aligned}\phi(s_1,s_2,s_3) = \psi(s_1,s_2,s_3)\Big\{&\frac{1}{s_1}[\phi(s_1,s_2,s_3)-\phi(s_1,s_2,0)-\phi(0,s_2,s_3)\\ &-\phi(s_1,0,s_3)+\phi(s_1,0,0)+\phi(0,s_2,0)+\phi(0,0,s_3)-\phi(0,0,0)]\\ &+\frac{s_3}{s_1}[\phi(s_1,s_2,0)-\phi(s_1,0,0)-\phi(0,s_2,0)+\phi(0,0,0)]\\ &+\frac{s_2}{s_1s_3}[\phi(s_1,0,s_3)-\phi(s_1,0,0)-\phi(0,0,s_3)+\phi(0,0,0)]\\ &+\frac{1}{s_2}[\phi(0,s_2,s_3)-\phi(0,s_2,0)-\phi(0,0,s_3)+\phi(0,0,0)]\\ &+\frac{s_2}{s_1}[\phi(s_1,0,0)-\phi(0,0,0)]+\frac{s_3}{s_2}[\phi(0,s_2,0)-\phi(0,0,0)]\\ &+\frac{1}{s_3}[\phi(0,0,s_3)-\phi(0,0,0)]+\phi(0,0,0)\Big\}.\end{aligned} \tag{41}$$

Letting $s_2\to 0$ in (41) gives

$$\begin{aligned}\phi(s_1,0,s_3) &= \psi(s_1,0,s_3)\{\frac{1}{s_3}[\phi(0,0,s_3)-\phi(0,0,0)]+\phi(0,0,0)\\ &+ (s_3-1)\phi_2(0,0,0)+\phi_2(0,0,s_3)\}.\end{aligned} \tag{42}$$

This implies

$$\phi(s_1,0,s_3) = \frac{\psi(s_1,0,s_3)}{\psi(0,0,s_3)}\phi(0,0,s_3), \tag{43}$$

and hence

$$\phi(s_1,0,0) = \frac{\psi(s_1,0,0)}{\psi(0,0,0)}\phi(0,0,0). \tag{44}$$

Letting $s_3 \to 0$ in (41) gives

$$\begin{aligned}\phi(s_1,s_2,0) = \psi(s_1,s_2,0)\left\{\frac{s_2}{s_1}[\phi(s_1,0,0)-\phi(0,0,0)]+\phi(0,0,0)\right.\\ \left.+\frac{s_2}{s_1}\phi_3(s_1,0,0)+(1-\frac{s_2}{s_1})\phi_3(0,0,0)\right\} = \psi(s_1,s_2,0)\cdot \\ \left\{\frac{s_2}{s_1}[\phi_3(s_1,0,0)+\phi(s_1,0,0)]+(1-\frac{s_2}{s_1})[\phi_3(0,0,0)+\phi(0,0,0)]\right\}.\end{aligned} \tag{45}$$

Note that the term within large brackets in (45) only contains ϕ functions with zeros at both the 2nd and 3rd positions. The explanation is that, if queue 3 is empty at the $(n+1)$th slot, then queue 2 must have been empty at the nth slot while queue 3 must have had at most one customer at the nth slot. This leads to the terms $\phi(s_1,0,0)$ and $\phi_3(s_1,0,0)$ and similar terms with s_1 replaced by zero. The same argument explains why those terms can subsequently be expressed in ϕ functions with zeros at all three positions, and why $\phi(s_1,0,s_3)$ in (43) could be expressed in $\phi(0,0,s_3)$.

Substituting $s_1=0$ in (42) and letting $s_3 \to 0$ gives

$$\phi(0,0,0) = \psi(0,0,0)\{\phi_3(0,0,0)+\phi(0,0,0)\} \tag{46}$$

and differentiating (43) with respect to s_3 gives

$$\phi_3(s_1,0,0) = \frac{\psi(s_1,0,0)}{\psi(0,0,0)}\phi_3(0,0,0) + \frac{d}{ds_3}\left(\frac{\psi(s_1,0,s_3)}{\psi(0,0,s_3)}\right)_{s_3=0}\phi(0,0,0). \tag{47}$$

Hence we have expressed $\phi(s_1,s_2,0)$ in $\phi(0,0,0)$ and the known function $\psi(.,.,.)$:

$$\begin{aligned}\phi(s_1,s_2,0) = \psi(s_1,s_2,0)\{\frac{s_2}{s_1}\frac{\psi(s_1,0,0)}{\psi(0,0,0)}\frac{\phi(0,0,0)}{\psi(0,0,0)}\\ +\frac{s_2}{s_1}\frac{d}{ds_3}\left(\frac{\psi(s_1,0,s_3)}{\psi(0,0,s_3)}\right)_{s_3=0}\phi(0,0,0)+(1-\frac{s_2}{s_1})\frac{\phi(0,0,0)}{\psi(0,0,0)}\}.\end{aligned} \tag{48}$$

Letting $s_1 \to 0$ in (45) gives

$$\begin{aligned}\phi(0,s_2,0) &= \psi(0,s_2,0)\{\phi_3(0,0,0)+\phi(0,0,0)\\ &+ s_2(\phi_{13}(0,0,0)+\phi_1(0,0,0))\}\end{aligned} \tag{49}$$

and differentiation of (43) and (44) with respect to s_1 gives

$$\phi_1(0,0,0) = \frac{\phi(0,0,0)}{\psi(0,0,0)}\psi_1(0,0,0), \tag{50}$$

and

$$\phi_{13}(0,0,0) = \frac{d}{ds_3}\left(\frac{\psi_1(0,0,s_3)}{\psi(0,0,s_3)}\right)_{s_3=0}\phi(0,0,0) + \frac{\psi_1(0,0,0)}{\psi(0,0,0)}\phi_3(0,0,0). \tag{51}$$

Hence we have expressed $\phi(0,s_2,0)$ in $\phi(0,0,0)$. We have so far succeeded in expressing $\phi(s_1,s_2,s_3)$ in $\phi(0,s_2,s_3)$ and its special terms $\phi(0,0,s_3)$ and $\phi(0,0,0)$. A relation between $\phi(0,s_2,s_3)$, $\phi(0,0,s_3)$ and $\phi(0,0,0)$ is obtained in the following way. Rewrite (41) by multiplying both sides with $s_1s_2s_3$. This gives

$$\begin{aligned}&\phi(s_1,s_2,s_3)s_2s_3[s_1-\psi(s_1,s_2,s_3)] = \psi(s_1,s_2,s_3)\{s_2s_3(s_3-1)\phi(s_1,s_2,0)\\ &+s_2(s_2-s_3)\phi(s_1,0,s_3)+s_3(s_1-s_2)\phi(0,s_2,s_3)\\ &+s_2(s_2-s_3)(s_3-1)\phi(s_1,0,0)+s_3(s_1-s_2)(s_3-1)\phi(0,s_2,0)\\ &+(s_1-s_2)(s_2-s_3)\phi(0,0,s_3)+(s_1-s_2)(s_2-s_3)(s_3-1)\phi(0,0,0)\}. \end{aligned} \tag{52}$$

Now let $\sigma_1 = \sigma_1(s_2,s_3)$ for fixed s_2 and s_3 be the unique solution within the unit disk of $s_1 = \psi(s_1,s_2,s_3)$. Then

$$\begin{aligned}&\phi(0,s_2,s_3) = (1-s_3)\phi(0,s_2,0)+(1-\frac{s_2}{s_3})\phi(0,0,s_3)\\ &+(s_3-1)(1-\frac{s_2}{s_3})\phi(0,0,0)+\frac{s_2}{s_2-\sigma_1}\{(s_3-1)\phi(\sigma_1,s_2,0)\\ &+(\frac{s_2}{s_3}-1)\phi(\sigma_1,0,s_3)+(1-\frac{s_2}{s_3})(1-s_3)\phi(\sigma_1,0,0)\}.\end{aligned}$$

Hence we have expressed $\phi(0,s_2,s_3)$ in terms of $\phi(0,0,s_3)$ and $\phi(0,0,0)$.

We next express $\phi(0,0,s_3)$ in $\phi(0,0,0)$. Substitution of $s_1 = s_2 = s$ in (52) gives

$$\begin{aligned}&\phi(s,s,s_3)ss_3[s-\psi(s,s,s_3)] = \psi(s,s,s_3)\{ss_3(s_3-1)\phi(s,s,0)\\ &+s(s-s_3)\phi(s,0,s_3)+s(s-s_3)(s_3-1)\phi(s,0,0).\end{aligned}$$

Now let $\sigma_2 = \sigma_2(s_3)$ for fixed s_3 denote the unique solution within the unit disk of $s = \psi(s,s,s_3)$. Then we have

$$\phi(\sigma_2,0,s_3) = (1-s_3)\phi(\sigma_2,0,0)+\frac{s_3(1-s_3)}{\sigma_2-s_3}\phi(\sigma_2,\sigma_2,0) \tag{53}$$

and so we have expressed $\phi(0,0,s_3)$ via (43), (44), (48) and (53) in $\phi(0,0,0)$. Finally we get $\phi(0,0,0)$ from (48) and the normalization condition $\phi(1,1,0) = 1-\psi_1(1,1,1)-\psi_2(1,1,1)-\psi_3(1,1,1)$.

For four or more queues, a similar approach should be followed. It is easy to write down the analog of (5) and (41) in the case of four queues, a functional equation for $\phi(s_1, s_2, s_3, s_4)$. The approach to subsequently determine this function is as follows.

1. Taking $s_4 \to 0$ leads to an expression for $\phi(s_1, s_2, s_3, 0)$ in $\phi(s_1, s_2, 0, 0)$ and $\phi_4(s_1, s_2, 0, 0)$, and eventually only in $\phi(0, 0, 0, 0)$ and known functions (cf. (45)-(48)).

2. Taking $s_3 \to 0$ eventually only leaves ϕ functions with zeros on the first three positions (cf. (42),(43)). Taking $s_2 \to 0$ eventually only leaves ϕ functions with zeros on the first two positions.

3. Hence it remains to determine $\phi(0, s_2, s_3, s_4)$, $\phi(0, 0, s_3, s_4)$, $\phi(0, 0, 0, s_4)$ and $\phi(0, 0, 0, 0)$. For this purpose, multiply the functional equation for $\phi(s_1, s_2, s_3, s_4)$ with $s_1 s_2 s_3 s_4$ (cf.(52)). Now consider the unique solution of $s_1 = \psi(s_1, s_2, s_3, s_4)$ in the unit s_1 disk for fixed s_2, s_3, s_4. This expresses $\phi(0, s_2, s_3, s_4)$ in $\phi(0, 0, s_3, s_4)$, $\phi(0, 0, 0, s_4)$ and $\phi(0, 0, 0, 0)$. Subsequently the same procedure is followed for $s = \psi(s, s, s_3, s_4)$ (in order to express $\phi(0, 0, s_3, s_4)$ in $\phi(0, 0, 0, s_4)$ and $\phi(0, 0, 0, 0)$) and for $t = \psi(t, t, t, s_4)$. Finally the remaining unknown $\phi(0, 0, 0, 0)$ is determined from a normalization condition.

Remark: It would be interesting to interpret taking $s_1 = s_2$, and $s_1 = s_2 = s_3$, (hence aggregating queue lengths in the first few queues) in terms of the queue reduction procedure of Morrison [3]. Morrison shows the following for the type of queueing network studied in the present paper: Upstream queues Q_1 and Q_2 may be replaced by a single equivalent queue $\tilde{Q}_2$ with a certain prescribed input and *the same output* as Q_2; similarly Q_1, Q_2 and Q_3 may be replaced by a single equivalent queue $\tilde{Q}_3$. While this reduction procedure does not yield the *joint* queue length distribution that has been derived in the present paper, it does contribute to one's understanding of the tractability of the analysis of this tandem model.

Note: After the submission of this paper, dr. J.A. Morrison has kindly sent us his unpublished manuscript 'Three discrete-time queues in tandem'; this report concerns the 3-queue case of section 7, but with independent arrivals at the different queues.

Acknowledgement: This research has been made possible by financial support of PTT Research, Leidschendam, The Netherlands. We like to thank J.L. van den Berg (PTT Research) for stimulating discussions.

References

1. Kaplan, M. A. (1980), A two-fold tandem net with deterministic links and source interference. Oper. Res., Vol. 28, No. 3, 512-526.

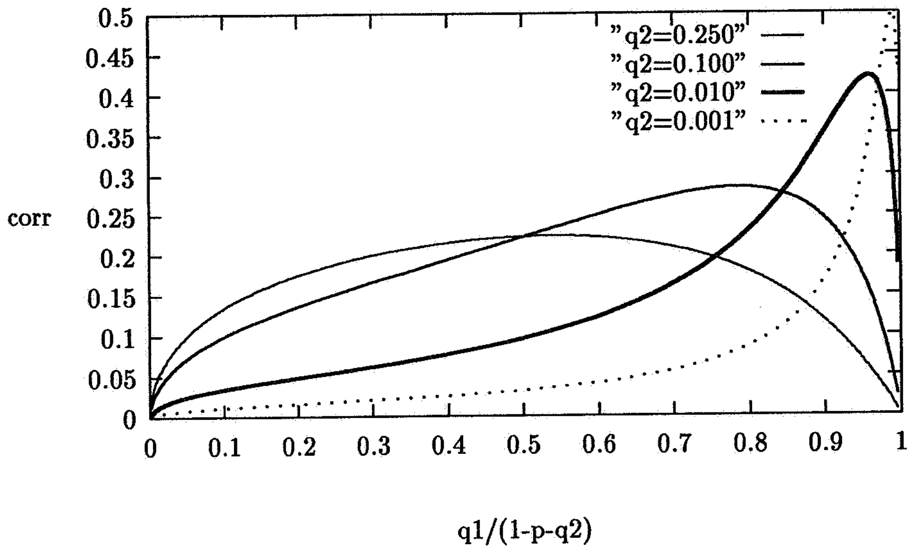
2. Khamisy, A. and Sidi, M. (1991), Discrete-time priority queueing systems with two-state Markov modulated arrival processes. In: Proceedings of IEEE Infocom '91, April 1991, Bal Harbour, Florida, USA, pp 1456-1463.

3. Morrison, J.A. (1978), A combinatorial lemma and its application to concentrating trees of discrete-time queues. Bell STJ, Vol. 57, No. 5, 1645-1652.

4. Morrison, J.A. (1979), Two discrete time queues in tandem. IEEE Trans. Communications, Vol. COM-27, No. 3, 563-573.

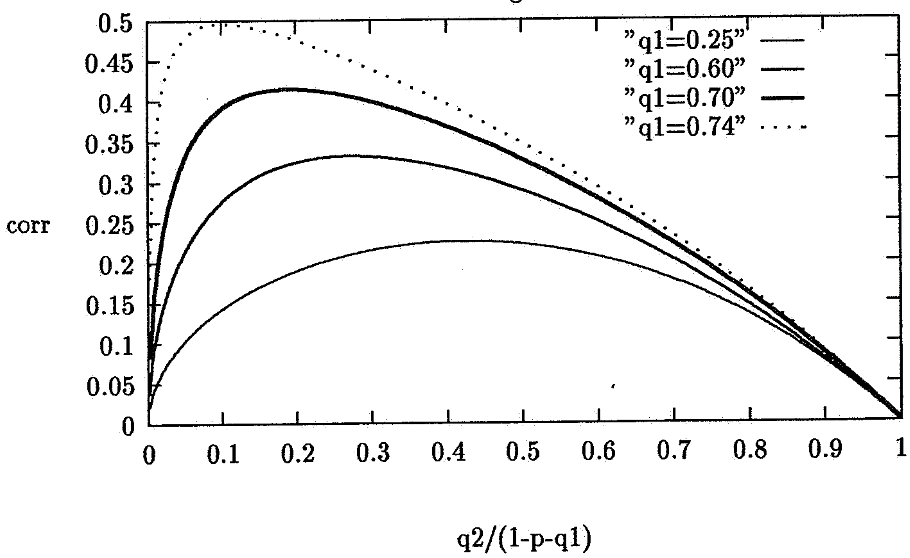
5. Shalmon, M. and Kaplan, M.A. (1984), A tandem network of queues with deterministic service and intermediate arrivals. Oper.Res. 32, Vol. 4, 753-773.

Fig. 2



corr
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
"q2=0.250"
"q2=0.100"
"q2=0.010"
"q2=0.001"
q1/(1-p-q2)

Fig. 3



corr
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
"q1=0.25"
"q1=0.60"
"q1=0.70"
"q1=0.74"
q2/(1-p-q1)