Performance bounds for early arrival discrete time queueing systems

Abdel-Karim Aboul-Hassan and Sherif I. Rabia
Math. & Physics Eng. Dept., Faculty of Engineering, Alexandria University, Alexandria (21544), Egypt

In an earlier paper we presented some results concerning discrete time queueing systems. These results were based on the late arrival assumption. The first result presented exact expressions for three performance measures namely utilization, occupation and delay in the case of single node queueing systems. The second was concerned with the network case where an approximate analysis based on linear programming was utilized to obtain upper and lower bounds for two performance measures namely occupation and delay. Moreover, we arrived at very tight estimates for the occupation and delay in a certain class of discrete time Jackson network assuming late arrival scheme. In the present work, we examine some discrete time queueing networks assuming the early arrival scheme. In fact, the early arrival scheme complicates the problem since some kind of slot dependency appears. This problem is resolved and upper and lower bounds for the occupation and delay in a discrete time Jackson network with early arrival are obtained. These results are needed as applications assume both types of arrival schemes. One of these applications is the Asynchronous Transfer Mode (ATM) technology which will be investigated.

في مقال سابق قدمنا بعض النتائج التي تختص بأنظمة طوابير الانتظار ذات الزمن المتقطع التي تعمل بخاصية الوصول المتأخر. كانت النتيجة الأولى التي توصلنا إليها هي الحصول على قيم مضبوطة لثلاثة من مقاييس الأداء وهي: معدل الاستخدام ومقدار الإشغال وزمن البقاء وذلك بالنسبة للأنظمة ذات العقدة الواحدة. أما النتيجة الثانية فكانت تختص بشبكات الطوابير حيث استطعنا عمل تحليل تقريبي لكل من مقدار الإشغال وزمن البقاء باستخدام البرمجة الخطية. وبالإضافة إلى ذلك فقد استطعنا تعيين مجموعة من شبكات جاكسون تتميز بحدود دنيا وعليا متقاربة جدا. في هذا البحث نقدم بعض النتائج التي تختص بأنظمة الطوابير ذات الزمن المتقطع التي تعمل بخاصية وصول مبكر. لاحظنا أن تطبيق خاصية الوصول المبكر تجعل العملية العشوائية التي تعبر عن النظام لا تتبع عمليات ماركوف. وقد قمنا بالتغلب على هذه المشكلة وتم الحصول على حدود دنيا وعليا لمقاييس الأداء. وتعد هذه النتائج من الأهمية بمكان لكثرة استخدام أنظمة الطوابير ذات الزمن المتقطع في التطبيقات الحديثة وخاصة نظم الاتصالات.

Keywords: Discrete time queueing systems, Eearly arrival scheme, Jackson network, Performance bounds.

1. Introduction

In this paper, an approximate analysis of discrete time queueing systems working under the early arrival scheme [1] is presented. Instead of obtaining exact expressions for the performance measures, we calculate upper and lower bounds on the performance measures of interest.

Discrete time queueing systems received great interest in recent years because they are used in the modeling and analysis of computer and communication networks working under the Asynchronous Transfer Mode (ATM) technology [2-4]. The basic feature of this technology (as opposed to other packet switching techniques) is that the basic cell (called ATM cell) containing data and routing information has a fixed size. Thus, the time required to transmit such cells within the same system is constant and can be considered as the basic time slot.

The proposed technique is applied to a discrete time Jackson network. However, similar steps can be followed to analyze other discrete time queueing systems. It has been pointed out in the literature [5] that the independence result of the continuous time Jackson network [6] does not hold in the discrete time case. To the best of our knowledge, discrete time Jackson network does not have a closed from solution. Therefore, obtaining upper and lower bounds on the performance measures of interest is a good alternative solution.

Our technique depends on constructing a linear program whose objective function is the required performance measure (mainly the total number of customers in the network). The constraints are obtained basically by assuming that the system reaches a steady state and examining the implication of this assumption on certain moments. The constraints variables are defined in a careful way to overcome the non-Markovian nature of the process. Solving the linear program for both minimization and maximization gives the required bounds. Mathematica [7] is used to automate the solution process and make it amenable for large systems.

Many numerical results are presented. In one example, a comparison is made between the present case and the late arrival case [8]. It is shown that the class with tight bounds in the late arrival case does not have the same property here. In another example, a comparison between three types of routing is made in terms of bounds tightness. An emphasis is put on the computation time. It is noted that the time required to generate bounds using our technique is much less than simulation time. Hence, when bounds tightness is acceptable quick performance measures estimates are better obtained using our approach.

This paper is organized as follows: Section 2 gives the specifications of the network under consideration. Section 3 represents the body of this work where the proposed technique is described in more details. Applying the steps of the technique leads to the main theorem (Theorem 1). It gives the linear program whose solution leads to the required bounds. Numerical results are presented in Section 4. Concluding remarks and some open problems are given in Section 5.

2. The model

We are interested here in analyzing a discrete time Jackson network. In the discrete time setting [1], we assume that the time axis is divided into intervals of equal length. These intervals are usually called time slots. The boundaries of these slots are called time points. System events, i.e., arrivals and departures, are assumed to occur only at these time points. More precisely, system events occur only *just after* or *just before* a time point. In order to be able to compute the system state at any time point, one must determine the order at which arrivals and departures occur. Hence, system evolution is assumed to be controlled by one of two basic schemes [1]: late arrival scheme and early arrival scheme. In the late arrival scheme (see fig. 1), arrivals are assumed to occur at the end of a time slot where departures occur at the beginning of the time slot. According to this scheme, an arriving customer to an idle server can not depart during the same time slot. He must wait until the next time slot (at least) in order to complete his service. The other possibility is to control the system with the early arrival scheme (see fig. 2). In this case, arrivals are assumed to occur at the beginning of a time slot where departures take place at the end of the time slot. In contrast to the late arrival scheme, an arriving customer to an idle server will get service in the same time slot and may depart before the beginning of the next time slot. In the present work, it is assumed that the system evolution is controlled by the early arrival scheme. For a parallel study regarding the other scheme, the reader is referred to [8].

The network consists of N nodes. Each node has an infinite queue and a single server. At the beginning of the time slot n, n = 1, 2,…, Arrivals (from outside the network) at node i, i = 1, 2,…, N occur in batches. The batch size is

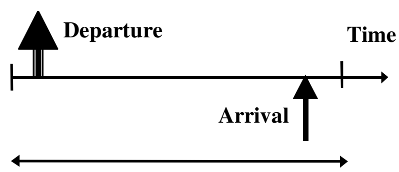


Fig. 1. Late arrival scheme.

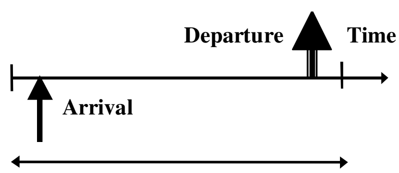


Fig. 2. Early arrival scheme.

denoted by $A_i(n)$. We assume that $\{A_i(n), n = 1, 2,...\}$ constitute a time stationary stochastic process [9]. Moreover, we assume that arrivals at different nodes are independent. Throughout this paper, we will use the notation $\overline{Y}$ to denote the expectation of the random variable Y. Hence, $\overline{A_i(n)} = E(A_i(n))$, $\overline{A_i^2(n)} = E(A_i^2(n))$ and so on. Since the process $\{A_i(n)\}$ is time stationary, we will write $\overline{A_i}$ to denote the expected number of arrivals during any time slot.

By the end of the time slot n, a single customer departs from node i with probability σ_i provided that node i is not empty. It is assumed that the service probability σ_i is independent of the number of time slots that the customer has stayed in this node. This assumption implies that the service time follows a geometric distribution. The departure process at node i is denoted by $\{R_i(n)\}$ where $R_i(n) \in \{0, 1\}$ is the number of departures from node i during the nth time slot. We will use R_i to represent the steady state of the process $\{R_i(n)\}$, i.e., $R_i = \lim_{n\to\infty} R_i(n)$ assuming that the system is stable.

After the customer completes his service at node i, he moves to node j with probability P_{ij} and leaves the network with probability $1 - \sum_{j=1}^{N} P_{ij}$. We assume for simplicity that $P_{ii} = 0$. According to the early arrival scheme employed here, a departure from a certain node is not counted as an arrival to his destination node until the beginning of the next time slot. The internal arrivals process at node i is denoted by $\{K_i(n)\}$ where $K_i(n) \in \{0, 1,..., N-1\}$ is the number of internal arrivals at node i during the nth time slot. Moreover, $K_i = \lim_{n\to\infty} K_i(n)$ represents the steady state of the process $\{K_i(n)\}$.

The system state is given by $X(n) = (X_1(n), X_2(n),..., X_N(n))$ where $X_i(n)$ is the number of customers at node i by the end of time slot n, $n = 1, 2,...$ and $X(0)$ is the initial state. Note that $X(n)$ is not a Markov chain because internal arrivals during one time slot depend on the departures during the previous time slot. Following our notation the vector $X = (X_1, X_2,..., X_N)$ represents the steady state of the network.

Finally, We assume that the policy applied at each node is non-idling, i.e., the server at node i will be idle during the time slot n if and only if $X_i(n-1) = A_i(n) = K_i(n) = 0$.

3. Linear programming setting

As far as we know, no closed form solution exist for the steady state distribution of $X(n)$ and it seems to be a hard problem to obtain such distribution especially because $X(n)$ is non-Markovian. From a practical point of view, it may be sufficient to obtain formulas for performance measures such as total number of customers in the network and total delay through the network. In the present work, we present a technique based on linear programming to obtain upper and lower bounds on these performance measures. This technique appeared originally in the analysis of continuous time queueing networks (see for example [10,11]). In [8,12], we adapted this technique to be used in the analysis of discrete time queueing systems working under the late arrival scheme. Here, we extend the application of this technique to deal with the early arrival case. In order to arrive at the required performance bounds we have the following steps:

First, we assume that the system has reached a steady state and examine the consequence of this assumption on the moments:

$E(X_i^2(n)),\ i = 1, 2,..., N,$

$E(X_i(n) X_j(n)),\quad i, j = 1, 2,..., N,\ i \neq j.$

We obtain a set of equality constraints in a new set of variables. We define these variables in a careful way to overcome the non-Markovian nature of the process. In fact some added variables have to be defined which were not required in the late arrival case [8].

Second, the non-idling nature of the applied policy leads to another set of inequality constraints in the same set of variables.

Third, we examine our definition of the decision variables and add another set of inequality constraints that gives upper and lower bounds on these variables.

Fourth, we formulate the objective function. We are interested mainly in the total number of customers in the network. Hence, this performance measure is formulated as a function of the decision variables.

Finally, a linear program is constructed and solved for both minimization and maximization to obtain the required bounds. The application of the first step leads to Lemmas 1 and 2, the second step leads to Lemma 3 and the third step leads to Lemmas 4, 5 and 6. These lemmas together with the application of the fourth and fifth steps lead to our basic result presented in Theorem 1.

Definition 1

The utilization stochastic process $\{U_i(n)\}$, $i = 1, 2, \ldots, N$ is defined as follows:

$U_i(n)$ = 0, if node i is idle during the nth time slot
= 1, otherwise.

Following our notation given in Section 2, $U_i = \lim_{n\to\infty} U_i(n)$ and $\overline{U_i}$ represents the steady state utilization of node i which can be computed using the usual traffic equations:

$$\sigma_i \overline{U_i} = \overline{A_i} + \sum_{j=1, j\neq i}^{N} \sigma_j \overline{U_j} P_{ji}. \quad (1)$$

Definition 2

Assuming that the system is stable, we define the following variables:

$$z_{ij} = \lim_{n\to\infty} E(U_i(n)X_j(n)),\ i, j = 1, 2, \ldots, N,$$

$$v_{ij} = \lim_{n\to\infty} E(U_i(n)U_j(n)),\ i, j = 1, 2, \ldots, N, \quad i \neq j,$$

$$s_{ij} = \lim_{n\to\infty} E(R_i(n)K_j(n)),\ i, j = 1, 2, \ldots, N, \quad i \neq j$$

$$t_{ij} = \lim_{n\to\infty} E(R_i(n)R_j(n)),\ i = 1, 2, \ldots, N-1,$$
$$j = i+1, i+2, \ldots, N,$$

Note that the non-idling assumption implies that $U_i(n)X_i(n) = X_i(n)$ for all i. However, $\{U_i(n)X_j(n)\}$ for $i \neq j$ is a new stochastic process that is bounded above by $\{X_j(n)\}$. This point will be made clearer in Lemma 3.

Lemma 1

$$2(\overline{A_i} - \sigma_i)\, z_{ii} + 2 \sum_{j=1, j\neq i}^{N} \sigma_j P_{ji} z_{ji} + \sum_{j=1, j\neq i}^{N} \sum_{q=1, q\neq i, q\neq j}^{N} \upsilon_{jq} \sigma_j \sigma_q P_{ji} P_{qi} + \overline{A_i^2} + \sigma_i \overline{U_i} - 2\sigma_i \overline{A_i} + (1 + 2\overline{A_i} - 2\sigma_i) \sum_{j=1, j\neq i}^{N} \overline{U_j} \sigma_j P_{ji} = 0,\ i = 1, 2, \ldots, N. \quad (2)$$

Proof

$$E(X_i^2(n) \mid X(n-1)) = \sum_{a=0}^{\infty} \sum_{k=0}^{N-1} \sum_{r=0}^{1} (X_i(n-1) + a + k - r)^2 \times P(A_i(n) = a, K_i(n) = k, R_i(n) = r).$$

After some simplifications, one arrives at the following result:

$$E(X_i^2(n) \mid X(n-1)) = X_i^2(n-1) + 2X_i(n-1)(\overline{A_i} + \overline{K_i(n)} - \overline{R_i(n)}) + \overline{A_i^2} + \overline{K_i^2(n)} + \overline{R_i^2(n)} + 2\overline{A_i K_i(n)} - 2\overline{A_i R_i(n)} - 2\overline{K_i(n) R_i(n)}. \quad (3)$$

The expectations $\overline{A_i(n)}$ and $\overline{A_i^2(n)}$ are system parameters and it is easy to show that:

$$\overline{R_i(n)} = \overline{R_i^2(n)} = U_i(n)\sigma_i. \quad (4)$$

The other expectations that appear in eq. (3) need further computation. To compute $\overline{K_i(n)}$, we define $K_{ij}(n) \in \{0, 1\}$ to be the number of customers that depart from node j during the $(n-1)th$ time slot and arrive at node i by the beginning of the nth time slot. Moreover, it can be easily proved that $\overline{K_{ij}(n)} = \overline{R_j(n-1)} P_{ji}$. Using eq. (4), then

$\overline{K_{ij}(n)} = U_j(n-1)\sigma_j P_{ji}$. Since $K_i(n) = \sum_{j=1, j\neq i}^{N} K_{ij}(n)$, then:

$$\overline{K_i(n)} = \sum_{j=1, j\neq i}^{N} U_j(n-1)\sigma_j P_{ji} . \quad (5)$$

To compute $\overline{A_i K_i(n)}$, one notes that arrivals to the network during any time slot are independent of the network state. Moreover, the number of internal arrivals at node i during the nth time slot does not depend of the number of arrivals to the network during the same time slot due to the early arrival assumption. Therefore, A_i and $K_i(n)$ are independent. Hence,

$$\overline{A_i K_i(n)} = \overline{A_i} \sum_{j=1, j\neq i}^{N} U_j(n-1)\sigma_j P_{ji} . \quad (6)$$

To compute $\overline{K_i^2(n)}$, we apply eq. (5) and use the independence between $K_{ij}(n)$'s for the same i to obtain:

$$\overline{K_i^2(n)} = \sum_{j=1, j\neq i}^{N} Var(K_{ij}(n)) + (\sum_{j=1, j\neq i}^{N} U_j(n-1)\sigma_j P_{ji})^2 .$$

Because $K_{ij}(n) \in \{0, 1\}$, one can check that:

$$Var(K_{ij}(n)) = U_j(n-1)\sigma_j P_{ji} - (U_j(n-1)\sigma_j P_{ji})^2 .$$

Hence,

$$\overline{K_i^2(n)} = \sum_{j=1, j\neq i}^{N} U_j(n-1)\sigma_j P_{ji} + \sum_{j=1, j\neq i}^{N} \sum_{q=1, q\neq i, q\neq j}^{N} (U_j(n-1)\sigma_j P_{ji})(U_q(n-1)\sigma_q P_{qi}) . \quad (7)$$

To compute $\overline{K_i(n)R_i(n)}$, we note that:

$$\overline{K_i(n)R_i(n)} = \sum_{k_i=1}^{N-1} E(R_i(n) \mid K_i(n) = k_i > 0) \times k_i \times P(K_i(n) = k_i) . \quad (8)$$

Due to the single service and early arrival assumptions, we have:

$$E(R_i(n) \mid K_i(n) = k_i > 0) = P(R_i(n) = 1 \mid K_i(n) = k_i > 0) = \sigma_i . \quad (9)$$

Substituting from eq. (9) into eq. (8) and applying eq. (5), then:

$$\overline{K_i(n)R_i(n)} = \sigma_i \sum_{j=1, j\neq i}^{N} U_j(n-1)\sigma_j P_{ji} . \quad (10)$$

The computation of $\overline{A_i R_i(n)}$ is very similar to that of $\overline{K_i(n)R_i(n)}$ and it is left to the reader to verify that:

$$\overline{A_i R_i(n)} = \sigma_i \overline{A_i} . \quad (11)$$

Now, substituting from eqs. (4), (5), (6), (7), (10) and (11) into eq. (3), taking the expectation of both sides and applying the steady state assumption, then:

$$2(\overline{A_i}\,\overline{X_i} + \sum_{j=1, j\neq i}^{N} \overline{X_i U_j}\,\sigma_j P_{ji} - \sigma_i \overline{X_i}) + \overline{A_i^2} + \sum_{j=1, j\neq i}^{N} \overline{U_j}\,\sigma_j P_{ji} + \sum_{j=1, j\neq i}^{N} \sum_{q=1, q\neq i, q\neq j}^{N} \overline{U_j U_q}\,\sigma_j \sigma_q P_{ji} P_{qi} + \overline{U_i}\,\sigma_i + 2\overline{A_i} \sum_{j=1, j\neq i}^{N} \overline{U_j}\,\sigma_j P_{ji} - 2\overline{A_i}\,\sigma_i - 2\sigma_i \sum_{j=1, j\neq i}^{N} \overline{U_j}\,\sigma_j P_{ji} = 0.$$

Applying definition 2 to the above equation gives eq. (2).

Lemma 2
For $i = 1, 2,..., N-1, j = i+1, i+2,..., N$:

$$\overline{A_j}\, z_{ii} + \sum_{q=1,q\neq j}^{N} \sigma_q P_{qj} z_{qi} - \sigma_j z_{ji} + \overline{A_i}\, z_{jj} + \sum_{q=1,q\neq i}^{N} \sigma_q P_{qi} z_{qj} -$$

$$\sigma_i z_{ij} + \overline{A_i}\;\overline{A_j} + \overline{A_i} \sum_{q=1,q\neq j}^{N} \overline{U_q}\, \sigma_q P_{qj} - \overline{A_i}\;\overline{U_j}\, \sigma_j +$$

$$\overline{A_j} \sum_{q=1,q\neq i}^{N} \overline{U_q}\, \sigma_q P_{qi} +$$

$$\sum_{q_1=1,q_1\neq i}^{N} \sum_{q_2=1,q_2\neq j,q_1\neq q_2}^{N} v_{q_1q_2} \sigma_{q_1} \sigma_{q_2} P_{q_1 i} P_{q_2 j} - s_{ji} -$$

$$\overline{A_j}\;\overline{U_i}\, \sigma_i - s_{ij} + t_{ij} = 0 . \qquad (12)$$

Proof

$E(X_i(n)\, X_j(n) \,|\, X(n-1)) =$

$$(X_i(n-1)+a_i+k_i-r_i)\times(X_j(n-1)+a_j+k_j-r_j)\times$$
$$\sum_{a_i=0}^{\infty}\sum_{k_i=0}^{N-1}\sum_{r_i=0}^{1}\sum_{a_j=0}^{\infty}\sum_{k_j=0}^{N-1}\sum_{r_j=0}^{1} P(A_i(n)=a_i, K_i(n)=k_i, R_i(n)=r_i, A_j(n)=a_j, K_j(n)=k_j, R_j(n)=r_j).$$

After some simplification, we obtain:

$$E(X_i(n)\, X_j(n) \mid X(n-1)) = X_i(n-1)\, X_j(n-1) + X_i(n-1)\, (\overline{A_j} + \overline{K_j(n)} - \overline{R_j(n)}) +$$
$$X_j(n-1)\, (\overline{A_i} + \overline{K_i(n)} - \overline{R_i(n)}) +$$
$$\overline{A_iA_j} + \overline{A_iK_j(n)} - \overline{A_iR_j(n)} + \overline{A_jK_i(n)} +$$
$$\overline{K_i(n)K_j(n)} -$$
$$\overline{K_i(n)R_j(n)} - \overline{A_jR_i(n)} - \overline{R_i(n)K_j(n)} +$$
$$\overline{R_i(n)R_j(n)} . \qquad (13)$$

The main task now is to compute the expectations in the right hand side of eq. (13). Since arrival at different nodes are independent, then:

$$\overline{A_iA_j} = \overline{A_i}\;\overline{A_j} . \qquad (14)$$

Number of internal arrivals at node j during the nth time slot depends only on the departures from the other nodes during the previous time slot and the routing probabilities $\{P_{ij}\}$. This implies that $A_i(n)$ and $K_j(n)$ are independent. Using eq. (5), then:

$$\overline{A_iK_j(n)} = \overline{A_i} \sum_{q=1,q\neq j}^{N} \overline{U_q(n-1)}\, \sigma_q P_{qj} . \qquad (15)$$

Number of departures from node j during the nth time slot depends only on the number of (internal or external) arrivals at this node during the same time slot as well as the service probability σ_i. This implies that $A_i(n)$ and $R_j(n)$ are independent. Using eq. (4), then:

$$\overline{A_iR_j(n)} = \overline{A_i}\, \overline{U_j(n)}\, \sigma_j . \qquad (16)$$

Recall that $K_i(n) = \sum_{q_1=1,q_1\neq i}^{N} K_{iq_1}(n)$ and

$K_j(n) = \sum_{q_2=1,q_2\neq j}^{N} K_{jq_2}(n)$. Hence,

$$\overline{K_i(n)K_j(n)} = \sum_{q_1=1,q_1\neq i}^{N} \sum_{q_2=1,q_2\neq j}^{N} \overline{K_{iq_1}(n)K_{jq_2}(n)} . \qquad (17)$$

Since $K_{iq_1}(n), K_{jq_2}(n) \in \{0, 1\}$, then
$\overline{K_{iq_1}(n)K_{jq_2}(n)} = Pr\left(K_{iq_1}(n)=1, K_{jq_2}(n)=1\right)$.

The two events $\{K_{iq_1}(n) = 1\}$ and $\{K_{jq_2}(n) = 1\}$ are disjoint if $q_1 = q_2$ and are independent if $q_1 \neq q_2$. Hence,

$$\overline{K_i(n)K_j(n)} = 0,\ q_1 = q_2$$
$$= Pr\left(K_{iq_1}(n)=1\right)\times Pr\left(K_{jq_2}(n)=1\right),\ q_1 \neq q_2 . \qquad (18)$$

From the definition of the stochastic process $\{K_{iq}(n), n \geq 1\}$, then:

$$Pr\left(K_{iq_1}(n)=1\right) = Pr\left(R_{q_1}(n-1)=1\right)\times P_{q_1 i}$$
$$= \overline{R_{q_1}(n-1)} \times P_{q_1 i} .$$

Substituting for $\overline{R_{q_1}(n-1)}$ from eq. (4), then:

$$Pr\left(K_{iq_1}(n)=1\right)=U_{q_1}(n-1)\sigma_{q_1}P_{q_1 i}\,. \quad (19)$$

Similarly,

$$Pr\left(K_{jq_2}(n)=1\right)=U_{q_2}(n-1)\sigma_{q_2}P_{q_2 j}\,. \quad (20)$$

Substituting from eqs. (19) and (20) into eq. (18), then:

$$\overline{K_{iq_1}(n)K_{jq_2}(n)} \\ =0, \qquad q_1=q_2 \\ =U_{q_1}(n-1)U_{q_2}(n-1)\sigma_{q_1}\sigma_{q_2}P_{q_1 i}P_{q_2 j}, \qquad q_1\neq q_2. \quad (21)$$

Substituting from eq. (21) into eq. (17), we obtain the required expression:

$$\overline{K_i(n)K_j(n)} = \sum_{q_1=1,q_1\neq i}^{N}\ \sum_{q_2=1,q_2\neq j,q_1\neq q_2}^{N} U_{q_1}(n-1)\times U_{q_2}(n-1)\sigma_{q_1}\sigma_{q_2}P_{q_1 i}P_{q_2 j}\,. \quad (22)$$

Now, substituting from eqs. (4), (5), (14), (15), (16) and (22) into eq. (13), taking the expectation of both sides and applying the steady state assumption, then:

$$\overline{A_jX_i} + \sum_{q=1,q\neq j}^{N}\overline{U_qX_i}\sigma_qP_{qj} - \overline{U_jX_i}\sigma_j + \overline{A_iX_j} + \sum_{q=1,q\neq i}^{N}\overline{U_qX_j}\sigma_qP_{qi} - \overline{U_iX_j}\sigma_i + \overline{A_iA_j} + \overline{A_i}\sum_{q=1,q\neq j}^{N}\overline{U_q}\sigma_qP_{qj} - \overline{A_i}\,\overline{U_j}\sigma_j + \overline{A_j}\sum_{q=1,q\neq i}^{N}\overline{U_q}\sigma_qP_{qi} + \overline{R_iR_j} + \sum_{q_1=1,q_1\neq i}^{N}\ \sum_{q_2=1,q_2\neq j,q_1\neq q_2}^{N}\overline{U_{q_1}U_{q_2}}\sigma_{q_1}\sigma_{q_2}P_{q_1 i}P_{q_2 j} - \overline{K_iR_j} - \overline{A_j}\,\overline{U_i}\sigma_i - \overline{R_iK_j} = 0.$$

Applying definition 2 to the above equation gives eq. (12).

Lemma 3

$$z_{ji}\leq z_{ii},\ i,j=1,2,\dots,N,\ i\neq j. \quad (23)$$

Proof

Recall from definition 2 that $z_{ij}=\lim_{n\to\infty}E(U_i(n)X_j(n))$. The non-idling nature of the policy applied at each node implies that $X_i(n)=0$ if $U_i(n)=0$. However, for $i\neq j$, $U_j(n)$ may be zero when $X_i(n)>0$. Hence, $U_i(n)X_i(n)\geq U_j(n)X_i(n)$ for all $n\geq 1$. Taking the expectation of both sides and assuming a steady state exists give eq. (23).

We were not able to obtain explicit expressions for v_{ij}, s_{ij} and t_{ij}. Instead, we present in the following three lemmas upper and lower bounds on these quantities.

Lemma 4

For $i=1,2,\dots,N-1$, $j=i+1,i+2,\dots,N$:

$$\left(\overline{U_i}+\overline{U_j}-1\right)^+\leq v_{ij}\leq \min(\overline{U_i},\overline{U_j})\,. \quad (24)$$

Proof

Recall from definition 2 that $v_{ij}=\lim_{n\to\infty}E(U_i(n)U_j(n))$. Because $U_i(n), U_j(n)\in\{0,1\}$, then:

$$E(U_i(n)\,U_j(n)) = Pr(U_i(n)=1,\ U_j(n)=1). \quad (25)$$

But,

$$(Pr(U_i(n)=1)+Pr(U_j(n)=1)-1)^+\leq Pr(U_i(n)=1,\ U_j(n)=1)\leq \min(Pr(U_i(n)=1),\ Pr(U_j(n)=1)). \quad (26)$$

Since $U_i(n), U_j(n)\in\{0,1\}$, then:

$$\overline{U_i(n)}=Pr(U_i(n)=1),\ \overline{U_j(n)}=Pr(U_j(n)=1). \quad (27)$$

Substituting from eq. (27) into eq. (26), then:

$$\left(\overline{U_i(n)}+\overline{U_j(n)}-1\right)^+\leq Pr\left(U_i(n)=1,U_j(n)=1\right) \leq \min(\overline{U_i(n)},\overline{U_j(n)})\,. \quad (28)$$

Combining eqs. (28) and (25) and taking the limit as n tends to infinity give eq. (24).

Lemma 5

For $i=1,2,\dots,N-1$, $j=i+1,i+2,\dots,N$:

$$\left(\overline{U_i}\sigma_i+\overline{U_j}\sigma_j-1\right)^+\leq t_{ij}\leq \min(\overline{U_i}\sigma_i,\overline{U_j}\sigma_j)\,. \quad (29)$$

Proof

The proof is very similar to that of Lemma 4 and hence it will be omitted.

Lemma 6

$$\sum_{j=1, j\neq i}^{N} s_{ij} \leq N - \sigma_i \sum_{j=1, j\neq i}^{N} \overline{U_j} \sigma_j P_{ji} \text{ , } i = 1, 2,\ldots, N. \quad (30)$$

Proof

Since single service is assumed, then: $\sum_{j=1}^{N} K_j(n) \leq N$. Multiplying both sides by $R_i(n)$ and noting that $R_i(n) \leq 1$, then $\sum_{j=1}^{N} R_i(n) K_j(n) \leq N$. This inequality can be written as follows:

$$\sum_{j=1, j\neq i}^{N} R_i(n) K_j(n) + R_i(n) K_i(n) \leq N.$$

Taking the expectation of both sides and using eq. (10), then:

$$\sum_{j=1, j\neq i}^{N} \overline{R_i(n) K_j(n)} + \sigma_i \sum_{j=1, j\neq i}^{N} U_j(n-1) \sigma_j P_{ji} \leq N$$

Taking the limit as n tends to infinity and applying definition 2 give eq. (30).

Now, we present the main result of this work. The following theorem gives a lower bound on the total number of customers in the system in terms of a linear program.

Theorem 1

In the steady state, the expected number of customers in the described Jackson network is bounded below by the value of the following linear program:

min $\sum_{i=1}^{N} z_{ii}$ subject to the constraints given in eqs. (2), (12), (23), (24), (29), (30) of Lemmas 1– 6 in addition to:

$$\upsilon_{ij} = \upsilon_{ji},\ i = 1, 2,\ldots, N-1,\ j = i+1, i+2, \ldots, N, \quad (31)$$

$$z_{ij} \geq 0, \qquad i,j = 1, 2, \ldots, N, \quad (32)$$

$$s_{ij} \geq 0, \qquad i,j = 1, 2, \ldots, N,\ i \neq j. \quad (33)$$

Proof

Recall from definition 2 that: $z_{ij} = \lim_{n\to\infty} E(U_i(n) X_j(n))$. Since we assume a non-idling policy then $E(U_i(n)\ X_i(n)) = E(X_i(n))$. Hence, $\sum_{i=1}^{N} z_{ii} = \lim_{n\to\infty} \sum_{i=1}^{N} E(X_i(n))$ which proves that the objective function represents the required performance measure. The constraints in eq. (31) follow obviously from the definition of the variables υ_{ij}. Finally, eqs. (32) and (33) represent the non-negativity constraints.

Remark 1

The linear program given in Theorem 1 is used to compute a lower bound on the expected number of customers in the network. To compute an upper bound, only the minimization problem is replaced by a maximization one. Moreover, Little's theorem can be used to compute similar bounds on the total delay time.

4. Numerical examples

In order to make the present approach easy to implement, the solution process is fully automated. A program is developed using the Mathematica [7] package to handle this task. Only the network specifications are given as inputs and then the program generates the linear program given in Theorem 1 and solves it for both minimization and maximization. Several examples are treated to examine the proposed approach. The bounds are compared with results obtained via simulation. During simulation, it is assumed that customers at each node are served according to the FCFS (First Come First Served) policy. Each simulation run lasts for 30,000 time slots.

Example 1 (Comparison with late arrival case):

In [8,12], we presented a parallel study of the present model but working under the late arrival scheme. From a large class of examples we have examined, we identified a class of queueing networks for which the bounds are very tight. Moreover, the difference between bounds approaches a certain upper limit as a function of the load factor. More specifically, we observed that the bounds are very tight in the case of equal arrival rates and uniform or symmetric routing. By a uniform routing, we mean that the served customer joins any other node or leaves the network with the same probability. In the symmetric routing case, it is assumed that the served customer at node i joins node j with the same probability with which served customer at node j joins node i.

In the present example, we generate bounds for this class of networks but working now under the early arrival scheme. The results are shown in fig. 3. Poisson arrivals with equal arrival rates are assumed. We observe that the bounds are tight in the case of light loads. However, when the load factor increases, this tightness decreases especially when the number of nodes is also increased. The difference between bounds here does not converge to a certain limit as appeared in the late arrival scheme.

Example 2 (Exit routing case):

As an attempt to determine a class of networks with more tight bounds, we present here a comparison between uniform and symmetric routing and another type of routing we name it exit routing. In this type of routing, the served customer at any node leaves the network with a constant probability q and joins any other node with equal probabilities. In fig. 4, we plot bounds difference against load factor for the three types of routing. The exit probability q is taken to be 0.7. We observe that the tightness of bounds increases in the case of exit routing. Moreover, increasing number of nodes decreases the tightness with a small rate in the case of exit routing.

In fact, we studied this type of routing in some more details. In fig. 5, we consider two values for the exit probability: 0.7 and 0.9 and in each case we plot the bounds difference against load factor for different values of N. As expected in the previous figure, the tightness of bounds decreases as the number of nodes is increased but with a small rate. Moreover, this tightness increases when the exit probability is also increased.

Remark 2

An important feature of the present technique is the time parameter. The average time required to solve the linear program of Theorem 1 and generate the bounds is very small compared to the time required to generate a simulation estimate. In fig. 6, we compare these two times for different number of nodes. From this comparison, we can say that whenever the tightness of the bounds is acceptable, one can rely on performance bounds technique rather than simulation estimates to produce a quick result.

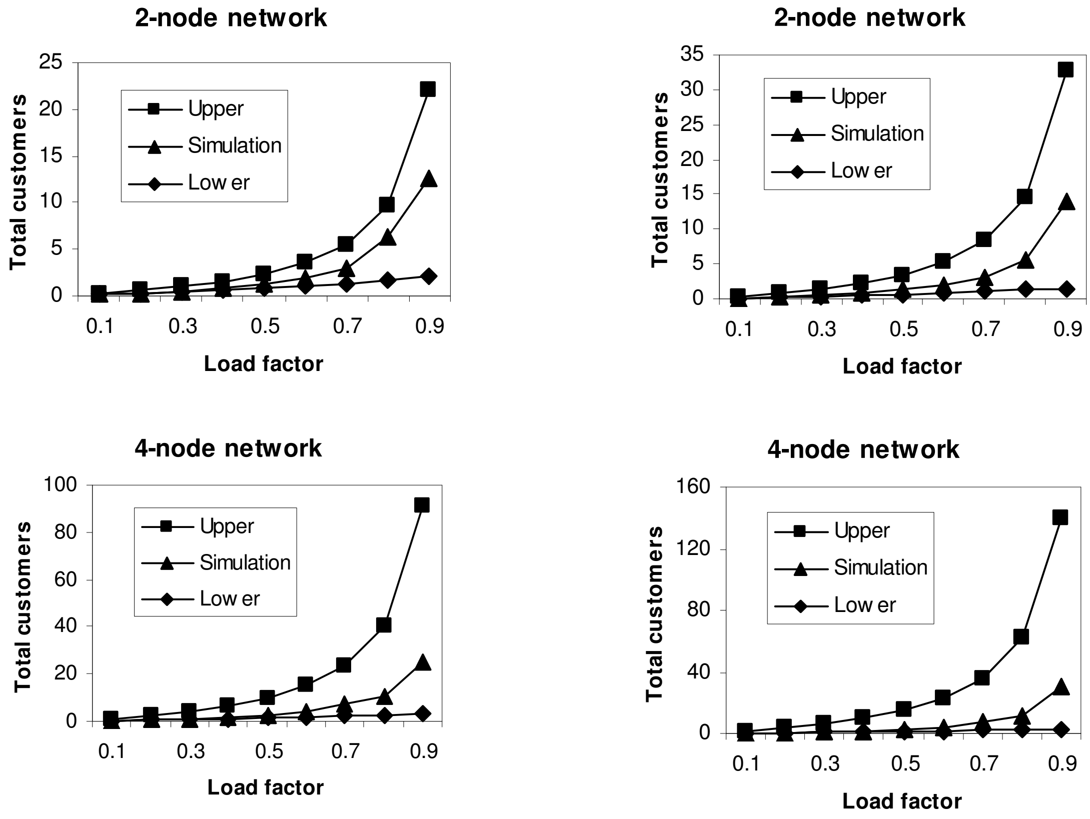
5. Conclusions

In the present work, we extended the performance bounds technique to analyze the discrete time Jackson network working under the early arrival scheme. We were able to overcome the non-Markovian nature of the process by defining a new set of variables and adding a corresponding set of constraints. Obtaining the bounds was made easy by designing a Mathematica [7] package to handle the solution process.

From the obtained results, we observed that the bounds are tight in the case of light loads. When the load factor increases, this tightness decreases especially when the number of nodes is also increased. To bring the bounds more tightly, it seems that another set of constraints has to be added. This is still an open problem.

When comparing the present results with those regarding the late arrival scheme [8], it was observed that the interesting tightness property exists in a certain class of networks working under the late arrival scheme does not hold when the early arrival scheme is applied. It is still an open problem to identify classes of networks that possess tight bounds in the early arrival scheme case.

Another important feature of the present technique is the time parameter. As explained in the numerical results, if the time parameter is important and when the tightness of bounds is acceptable, then quick performance



(a) Uniform routing (b) Symmetric routing

Fig. 3. Results for uniform and symmetric routing.

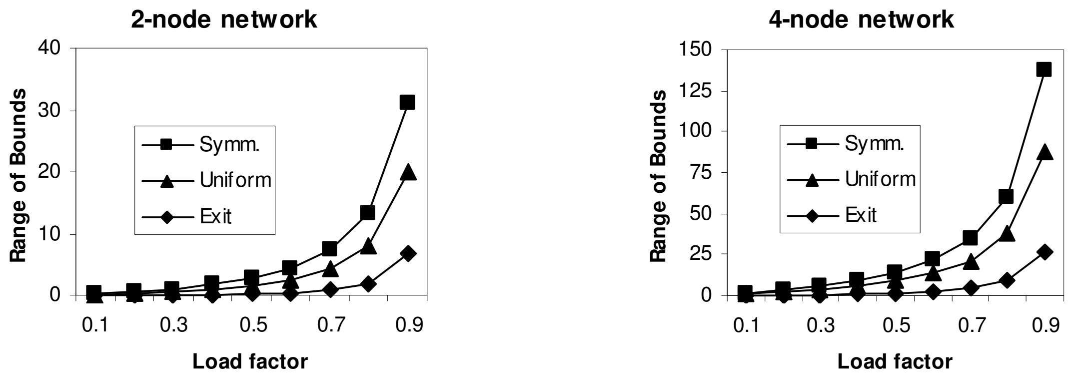


Fig. 4. Comparison between three types of routing.

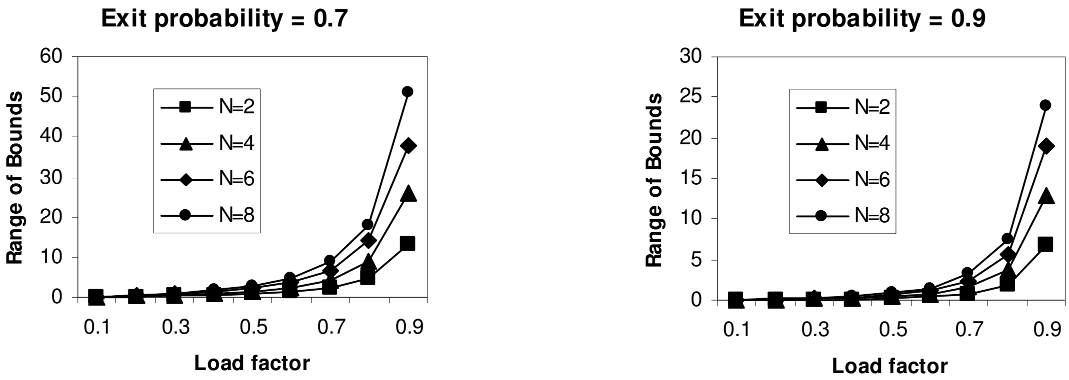


Fig. 5. Bounds tightness for exit routing.

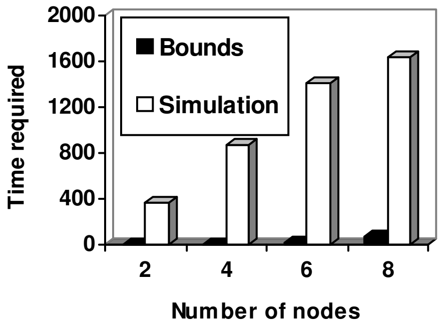


Fig. 6. Run time comparison.

measures estimates can be obtained using the present approach.

In order to obtain a linear program setting, only the second order moments were examined. In fact, examining higher order moments can be incorporated into the present technique in a straightforward manner but it leads to a non-linear program setting. This point needs more investigation and it is another open problem.

The non-Markovian nature of the process emerges mainly from the assumption that departures (at the end of a time slot) are not considered as arrivals to their destination nodes until the beginning of the next time slot. If we change this assumption and allow departures to constitute arrivals before the beginning of the next time slot, then the system state evolves as a Markov chain. Obtaining performance bounds for networks working under this assumption is the subject of our incoming paper.

References

[1] E. Gelenbe and G. Pujolle, Introduction to Queueing Networks, second edition. John Wiley, New York (1997).

[2] Li S.-Q., "A General Solution Technique for Discrete Queueing Analysis of Multimedia Traffic on ATM," IEEE Trans, Comm, Vol. 39, pp. 1115-1132 (1991).

[3] M. W., McKinnon G. N. Rouskas, and H. G. Perros, "Performance Analysis of a Photonic Single-Hop ATM Switch Architecture, with Tunable Transmitters and Fixed Frequency Receivers," Performance Evaluation, Vol. 33, pp. 113-136 (1998).

[4] M. E. Woodward, "Towards the Accurate Modeling of High-Speed Communication Networks with Product-Form Discrete-Time Networks of Queues," Computer Communications. Vol. 21, pp. 1530-1543 (1998).

[5] K. Bharath-Kumar, "Discrete Time Queueing Systems and Their Networks," IEEE Trans, Comm, COM-28, pp. 260-263 (1980).

[6] J. R. Jackson, "Networks of Waiting Lines," Operations Research, Vol. 5, pp. 518-521 (1957).

[7] S. Wolfram, The Mathematica Book, 3rd ed., Wolfram Media/Cambridge University press (1996).

[8] A. Aboul-Hassan S. I. and Rabia "Performance bounds for discrete time

Jackson network with batch arrivals" Unpublished manuscript.

[9] B. L. Nelson, Stochastic Modeling, Analysis and Simulation, McGraw-Hill, New York (1995).

[10] D. Bertsimas, I. Ch., and J. N. Paschalidis and Tsitsiklis, "Optimization of Multiclass Queueing Networks: Polyhedral," and Nonlinear Characterization of Achievable Performance. Ann. App. Probab, Vol. 4, pp. 43-75 (1994).

[11] S. Kumar and P. R. Kumar, "Performance Bounds for Queueing Networks and Scheduling Policies," IEEE Trans. Automat. Control, Vol. 39, pp. 1600-1610.

[12] S. I. Rabia and A. Aboul-Hassan, "A Unified Approach for the Approximate Analysis of Discrete Time Queueing Systems. Presented in the Conference on Stochastic Networks, Madison, (2000).

Received September 16, 2003
Accepted November 19, 2003