Deadlock free routing on a ring : a performance evaluation

Zhen Liu, Philippe Mussi

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

Deadlock Free Routing on a Ring:
A Performance Evaluation

Zhen Liu and Philippe Mussi

N° 2233

Avril 1994

PROGRAMME 1

Architectures parallèles,

bases de données,

réseaux et systèmes distribués

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Deadlock Free Routing on a Ring:
A Performance Evaluation

Zhen Liu and Philippe Mussi *

Programme 1 — Architectures parallèles, bases de données, réseaux
et systèmes distribués
Projet Mistral

Rapport de recherche n°2233 — Avril 1994 — 16 pages

Abstract: This paper is concerned with the performance analysis of a deadlock free protocol for limited capacity ring networks. There are N nodes on the ring. Each node is connected with an emitter and a receiver of messages and is equiped with a buffer of capacity two for receiving and transmitting messages. The ring is of uni–direction so that every node has a communication predecessor and a successor.

The system is modeled by a queueing system with finite capacities and blockings. An approximate analysis is provided for computing various performance measures, such as throughput, mean sojourn time of messages, and mean number of messages on the ring. Techniques in use are based on aggregation of Markov chains. Marginal queue length probability distributions are directly derived from aggregated state transition probabilities.

Key-words: Performance evaluation, Distributed systems, Routing, Blocking queueing networks

(Résumé : tsvp)

This work was partially supported by CNRS under C^3 (Concurrency, Communication, Cooperation) program

* {Zhen.Liu}{Philippe.Mussi}@sophia.inria.fr

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Routage sans blocage sur un anneau:
Une évaluation de performance

Résumé : Ce rapport est consacré à l'évaluation des performances d'un protocole de routage sans blocage pour les réseaux en anneau à capacité limitée. Chacun des N nœuds de l'anneau est connecté à un émetteur et un récepteur de messages et est muni d'un tampon de messages de capacité deux. L'anneau est uni-directionnel et travaille en temps discret. Au début de chaque période de temps, un nœud peut recevoir un message de son émetteur ou de son prédécesseur, et peut transmettre un message à son successeur ou à son récepteur. Un message ne peut être transmis au successeur que si ce dernier n'est pas saturé. Nous supposons également qu'un nœud ne peut recevoir de messages de son émetteur que si son tampon est vide.

Ce système est modélisé par un réseau de files d'attente avec blocages et capacités limitées. Grâce à une analyse approchée, nous calculons diverses mesures de performance, telles que le débit, le temps moyen de séjour des messages, le nombre moyen de messages sur l'anneau. Les techniques utilisées sont basées sur l'aggrégation des chaînes de Markov. Les distributions de probabilité des longueurs de files sont directement déduites des probabilités de transition des états aggrégés.

Mots-clé : Évaluation de performance, Systèmes répartis, Routage, Réseaux de files d'attente avec blocages

1 Introduction

Most protocols proposed for distributed systems are based on some academic assumptions, for example, processes are often supposed to be able to exchange messages in finite time, or some processes must never refuse or delay incoming messages, there are message buffers of infinite capacities, etc... Such properties may be very difficult to ensure in real systems, especially in the case of *Transputers* networks when the basic communication principles are those of OCCAM [3], or analogous.

A *Transputer* processor can only use four communication links. Every two processes can exchange information only by *rendez–vous*. In addition, memory space is restricted on each processor by hardware, and dynamic memory is restricted by software, so that buffering space is greatly limited.

Consider a Transputers network of ring structure. There are N nodes (or processors) on the ring. Each node is connected with an emitter and a receiver of messages and is equiped with a buffer of finite capacity for receiving and transmitting messages. The ring is of uni–direction so that every node has a communication predecessor and a successor. Every node has an emitting process and a receiving process. The former transmits messages in the buffer to the receiver or the successor. The latter receives messages from the emitter or the predecessor and put them into the buffer. We assume that the receiving process never refuses or infinitely delays the arriving messages. Figure 1 illustrates the system architecture.

In order to avoid deadlocks, Roscoe proposed and proved in [1] a protocol for the communications between the nodes. The algorithm relies on a very simple flow control on incoming messages. The receiving process of a node accepts a message from its predecessor as long as the buffer is not full. Whereas it accepts a message from its emitter if and only if the buffering space of the node will not be filled up by the current message. Readily, this mechanism ensures that buffering space for the whole ring will not be filled up. Consequently, no deadlock may occur, as no receiving process is allowed to infinitely refuse messages.

This paper is concerned with the performance aspects of such a protocol. The system will be modeled by a Markovian queueing network with finite capacities and blockings. We provide an approximate analysis of the model in using techniques of aggregations of Markov chains. We derive various performance criteria such as mean message number, mean throughput, mean sojourn time of messages, and probability distributions of buffer occupation.

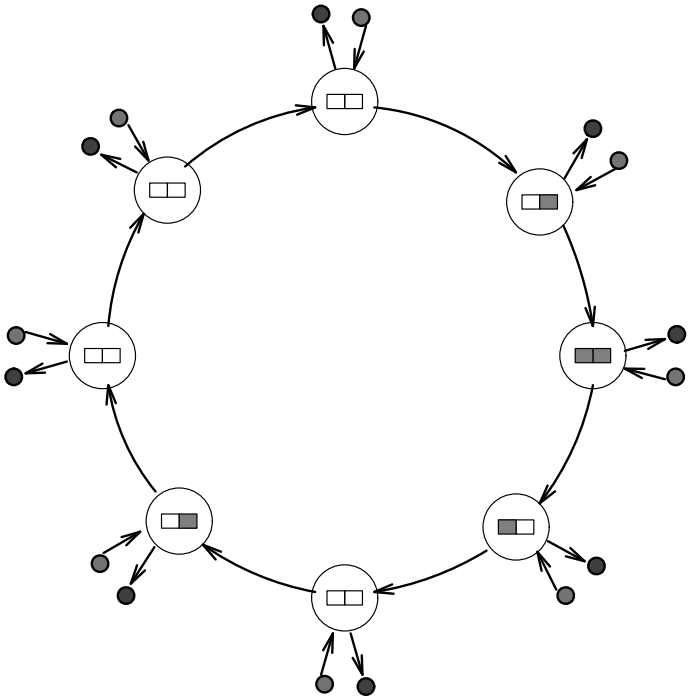

Figure 1: The Transputers ring

The paper is organized as follows. In the next section we describe a probabilistic model and its statistical assumptions. In Section 3 we analyze the model for the steady state regime and compute the performance measures. In Section 4 we provide numerical results in illustrating relations between these measures.

2 Modeling

In the sequel, we study a probabilistic model of a uni-directional ring, with limited capacity buffers on the nodes, which follows the deadlock-free protocol described above.

Let N be the number of nodes on the ring, denoted by $node_1, \cdots, node_N$. They are labeled in such a way that $node_i$, $1 \leq i \leq N$, is the predecessor of $node_{(i \bmod N)+1}$.

Each node is connected with an emitter and a receiver of messages and is equiped with a buffer of capacity two for receiving and transmitting messages. Messages arrive from its predecessor or its emitter, and are transmitted to its successor or its receiver. If the buffer is full, the incoming messages of a node are blocked on its emitter and/or its predecessor. Otherwise, its predecessor can always transmit messages to the node. A message is accepted from the emitter if and only if the buffer is empty. A message can be transmitted to the successor only if the buffer of the latter is not full.

We assume that all messages have the same length, and the transfer times are the same for all messages. Hence our model will be time–slotted, with a time–unit equal to the duration of a transfer. At the beginning of a time slot, a node can receive a message from either the emitter or the predecessor, and can transmit a message either to the successor or to the receiver, all of which terminate at the end of the time slot.

We assume that at the beginning of each time slot, the emitter of $node_i$, $1 \leq i \leq N$, tries to send a message onto the ring with a probability p_i. We assume also that a message in the buffer of $node_i$ decides to quit the ring (i.e., go into the receiver of $node_i$) with probability q_i. These events are called external arrivals and external departures, respectively, and are assumed to be Bernoulli processes. Without loss of generality, we assume that $0 < p_i, q_i < 1$ for all $1 \leq i \leq N$.

3 Performance Analysis

This section is concerned with the performance analysis of the above model. We are particularly interested in the *throughput* of the network, the *number of messages* on the ring, the *sojourn time* of a message, etc. The analysis will focus on the stationary regime of the model. Readers can readily check that the steady state does exist. The discussions are organized as follows. First, we define the performance measures and some notation. Second, we study the stationary regime of the model, and compute the state transition probabilities. The marginal queue length probability distributions are then obtained. Finally the performance measures are shown to be the immediate consequences of these queue length distributions.

3.1 Definitions and notation

Before proceeding the analysis, we need some definitions. The *throughput* T of the network is defined as the mean number of messages leaving the ring per slot of time. We will also use t to denote the *throughput* of a node, which is defined as the mean number of messages leaving the node per slot of time. The *mean sojourn time* S of a message is referred to as the mean number of time slots needed for a message to reach its destination after its acceptance. We denote by C and c the *mean number of messages* on the ring and on a node respectively.

Let $a,\ \alpha,\ d,\ \delta$ denote respectively the events taking place on a node: receive a message from the sender, receive a message from the predecessor, transmit a message to the receiver, transmit a message to the successor. The notations $\neg a,\ \neg\alpha,\ \neg d,\ \neg\delta$ are referred to as the case where these events do not take place.

Let the random variables $\chi,\ \hat{\chi} \in \{0, 1, 2\}$ denote respectively the number of messages to be transmitted on the node at the beginning and at the end of a time slot. Let P_0, P_1 and P_2 be respectively the probability that the node has 0, 1, and 2 messages to be transmitted.

3.2 Marginal Distributions

The state of the system can be described by the vector $\vec{\chi} = (\chi_1, \cdots, \chi_N)$, where the index i, $i = 1, \cdots, N$, is used to refer to node i. There are in total $3^N - 1$ possible states (due to the deadlock free property, the state $\vec{\chi} = (2, 2, \cdots, 2)$ will

never occur). One can easily check that under the assumption that $0 < p_i, q_i < 1$, the system is irreducible and aperiodic. The stationary regime hence exists [2].

It is possible to write out all the possible state transitions of the Markov chain. Nevertheless the $(3^N - 1) \times (3^N - 1)$ transition probability matrix turns out to be analytically (and numerically) intractable.

In this paper we consider the computation of marginal distributions. We approximate joint probabilities by products of marginal probabilities. We analyze all the possible state transitions from χ to $\hat{\chi}$ for an arbitrarily fixed node, and compute their probabilities.

Observe that at the beginning of a time slot, all what occurs on a node are the events $a,\ \alpha,\ d,\ \delta, \neg a,\ \neg\alpha,\ \neg d,\ \neg\delta$ and their feasible combinations. The conditions under which they take place are determined by the values of the random variables χ^-, χ, χ^+ at that instant. The following table summarize the aggregated states transitions and their probabilities, where, for sake of clearness, we add exponents $+$ and $-$ to refer to the successor and the predecessor of the node, respectively.

State Transitions				
No.	event	(χ^-,χ,χ^+)	$\hat{\chi}$	approximate transition probability
0	a , α ,$\neg d$,$\neg\delta$	(1-2,0, $\cdot$)	2	$(P_1^- + P_2^-)P_0 p(1-q^-)$
1	a ,$\neg\alpha$,$\neg d$,$\neg\delta$	($\cdot$,0, $\cdot$)	1	$P_0 p[P_0^- + (P_1^- + P_2^-)q^-]$
2	$\neg a$, α ,$\neg d$,$\neg\delta$	(1-2,0, $\cdot$)	1	$P_0(P_1^- + P_2^-)(1-p)(1-q^-)$
3	$\neg a$, α ,$\neg d$,$\neg\delta$	(1-2,1, $\cdot$)	2	$P_1(P_1^- + P_2^-)(1-q^-)[(1-q) + P_2^+]$
4	$\neg a$, α ,$\neg d$, δ	(1-2,1,0-1)	1	$P_1(P_1^- + P_2^-)(1-q^-)(P_0^+ + P_1^+)(1-q)$
5	$\neg a$, α , d ,$\neg\delta$	(1-2,1, $\cdot$)	1	$P_1(P_1^- + P_2^-)(1-q^-)q$
6	$\neg a$,$\neg\alpha$, d ,$\neg\delta$	($\cdot$,1, $\cdot$)	0	$P_1 q[P_0^- + (P_1^- + P_2^-)q^-]$
7	$\neg a$,$\neg\alpha$, d ,$\neg\delta$	($\cdot$,2, $\cdot$)	1	$P_2 q$
8	$\neg a$,$\neg\alpha$,$\neg d$, δ	($\cdot$,1,0-1)	0	$P_1(P_0^+ + P_1^+)(1-q)[P_0^- + (P_1^- + P_2^-)q^-]$
9	$\neg a$,$\neg\alpha$,$\neg d$, δ	($\cdot$,2,0-1)	1	$P_2(P_0^+ + P_1^+)(1-q)$
10	$\neg a$,$\neg\alpha$,$\neg d$,$\neg\delta$	($\cdot$,0, $\cdot$)	0	$P_0(1-p)[P_0^- + (P_1^- + P_2^-)q^-]$
11	$\neg a$,$\neg\alpha$,$\neg d$,$\neg\delta$	($\cdot$,1, 2)	1	$P_1 P_2^+(1-q)[P_0^- + (P_1^- + P_2^-)q^-]$
12	$\neg a$,$\neg\alpha$,$\neg d$,$\neg\delta$	($\cdot$,2, 2)	2	$P_2 P_2^+(1-q)$

Now using the relation

$$P_i = P[\chi = i] = P[\hat{\chi} = i]\,, \qquad i = 0, 1, 2,$$

yields three equations for the marginal queue length distributions:

$$\begin{aligned} P_2 &= (P_1^- + P_2^-)P_0 p(1-q^-) \\ &\quad + P_1(P_1^- + P_2^-)(1-p)(1-q^-)[(1-q) + P_2^+] \\ &\quad + P_2 P_2^+(1-q) \end{aligned} \tag{1}$$

$$\begin{aligned} P_1 &= P_0 p[P_0^- + (P_1^- + P_2^-)q^-] \\ &\quad + P_0(P_1^- + P_2^-)(1-p)(1-q^-) \\ &\quad + P_1(P_1^- + P_2^-)(1-q^-)(P_0^+ + P_1^+)(1-q) \\ &\quad + P_1(P_1^- + P_2^-)(1-q^-)q \\ &\quad + P_2 q \\ &\quad + P_2(P_0^+ + P_1^+)(1-q) \\ &\quad + P_1 P_2^+(1-q)[P_0^- + (P_1^- + P_2^-)q^-] \end{aligned} \tag{2}$$

$$\begin{aligned} P_0 &= P_1 q[P_0^- + (P_1^- + P_2^-)q^-] \\ &\quad + P_1(P_0^+ + P_1^+)(1-q)[P_0^- + (P_1^- + P_2^-)q^-] \\ &\quad + P_0(1-p)[P_0^- + (P_1^- + P_2^-)q^-] \end{aligned} \tag{3}$$

In the sequel, we will restrict ourselves to the symmetric case, i.e. :

$$p_i = p, \quad i = 1, \cdots, N \quad ; \quad q_i = q, \quad i = 1, \cdots, N \tag{4}$$

It follows from the nature of the ring network under consideration, that the stationary marginal queue length distributions are identical, i.e. :

$$P_j^i = P_j, \quad i = 1, \cdots, N, \quad j = 0, \cdots, 2. \tag{5}$$

Hence the above equations can be simplified as

$$P_2 = (1-q)\left\{(P_1 + P_2)P_0 p + P_1(P_1 + P_2)(1-p)[(1-q) + P_2] + P_2^2\right\} \tag{6}$$

$$\begin{aligned} P_1 &= P_0 p[P_0 + (P_1 + P_2)q] + P_0(P_1 + P_2)(1-p)(1-q) \\ &\quad + P_1(P_1 + P_2)(1-q)^2(P_0 + P_1) + P_1(P_1 + P_2)(1-q)q \\ &\quad + P_2 q + P_2(P_0 + P_1)(1-q) + P_1 P_2(1-q)[P_0 + (P_1 + P_2)q] \end{aligned} \tag{7}$$

$$P_0 = \{P_1 q + P_1(P_0 + P_1)(1-q) + P_0(1-p)\}\,[P_0 + (P_1 + P_2)q] \tag{8}$$

In addition, we know that, in steady state, arrival rate and departure rate of the network are equal. Therefore,

$$pP_0 = q(P_1 + P_2) \tag{9}$$

Moreover, we have the following trivial relation:

$$P_0 + P_1 + P_2 = 1 \tag{10}$$

Owing to the theory of Markov chains, stationary distribution is unique. Hence marginal queue length probabilities (P_0, P_1, P_2) can be obtained with either three of the equations (6)|(10). It follows from (9) and (10) that

$$P_0 = \frac{q}{p+q} \tag{11}$$

and

$$P_1 + P_2 = \frac{p}{p+q} \tag{12}$$

Replacing the corresponding terms in (6) by (11) and (12) immediately yields a quadratic polynomial in terms of P_2:

$$F(P_2) = b_2 P_2^2 + b_1 P_2 + b_0 = 0 \tag{13}$$

where

$$b_2 = (1-q)\frac{p^2+q}{p+q} \tag{14}$$

$$b_1 = (1-q)(1-p)\frac{pq(p+q-1}{(p+q)^2} - 1 \tag{15}$$

$$b_0 = (1-q)\frac{p^2}{(p+q)^2}(1-p+pq) \tag{16}$$

In view of (12), $0 < P_2 < p/(p+q)$. We are going to show that $F(X) = 0$ has one and only one real root between 0 and $p/(p+q)$. Indeed, it is easy to see that

$$F(0) = b_0 > 0 \tag{17}$$

With some simple manipulations one readily get

$$(p+q)^3 F(\frac{p}{p+q}) \quad = \quad p^3q + p^2q^2 - 2p^4q^2 - p^4q - 2p^3q^3 - 2p^3q^2 - p^2q^3 - p^2q - pq^2$$

Using the facts that $p^3q < p^2q$ and $p^2q^2 < pq^2$ yields immediately

$$F(\frac{p}{p+q}) < 0 \tag{18}$$

In addition, observe that $b_2 = \lim_{X\to\infty} F(X)/X^2$ is non–negative, so that the second real root of $F(X) = 0$ is greater than $p/(p+q)$.

Therefore, equations (11)|(13) yield an unique solution of the steady state probability distribution of (P_0, P_1, P_2) for any pair of (p, q). Such a solution will be denoted as (Π_0, Π_1, Π_2).

3.3 Performance measures computation

The computation of the performance measures then becomes immediate. Node throughput t, ring throughput T, mean number of messages on a node c and on the ring C, and mean number of visited nodes of a message L are given by:

$$t = q(\Pi_1 + \Pi_2) = \frac{pq}{p+q} \tag{19}$$

$$T = Nt = \frac{Npq}{p+q} \tag{20}$$

$$c = \Pi_1 + 2\Pi_2 = \frac{p}{p+q} + \Pi_2 \tag{21}$$

$$C = Nc \tag{22}$$

$$L = \frac{1}{q} \tag{23}$$

Using Little's formula [4] yields the mean sojourn time of a message S

$$S = \frac{C}{T} = \frac{p + \Pi_2(p+q)}{pq} \tag{24}$$

4 Numerical Results

In this section, we provide numerical results of the performance measures as functions of external arrival and departure probabilities. We also illustrate the relations between the performance measures.

For any given pair of parameters (p, q), we solve numerically (13), and choose the smallest root Π_2 to get the steady state probability distribution (Π_0, Π_1, Π_2) for the number of messages on each node.

Recall that the mean number of hops of a message $L = 1/q$. We hence compute mean number of messages on a node c, throughput of a node t, and mean sojourn time of a message on the ring S, as functions of p and L. The results are illustrated by Figure 2, where each curve corresponds to a particular value of p: $p = i/20, \quad i = 1, \cdots, 19$.

In figure 5, we illustrate the relation between S, the mean sojourn time of messages, and t, the mean node throughput.

5 Conclusions

In this paper, we have studied the performance aspects of a deadlock free protocol for a limited capacity ring network. A probabilistic model together with an approximate analysis have been proposed. Numerical results have also been provided.

In the analysis, we assumed that the ring was symmetric, i.e., $p_1 = p_2 = \cdots = p_N$, and $q_1 = q_2 = \cdots = q_N$. However, numerical computation is possible for general cases. Indeed, using (1), (2), (3), (11) and (12), one can readily obtain polynomial equations in terms of P_1 or P_2 for any arbitrary node.

For degenerated cases where some p or q equals 0 or 1, the analysis will be simpler as there will be less state transitions.

Further research topics include analyzing models with Bernoulli external arrival processes replaced by other kind of processes (such as geometry or exponential processes), and with Bernoulli external departure processes replaced by more general distributions on the number of hops of a message.

Acknowledgment

The authors are grateful to Dr. Philippe NAIN for fruitful discussions.

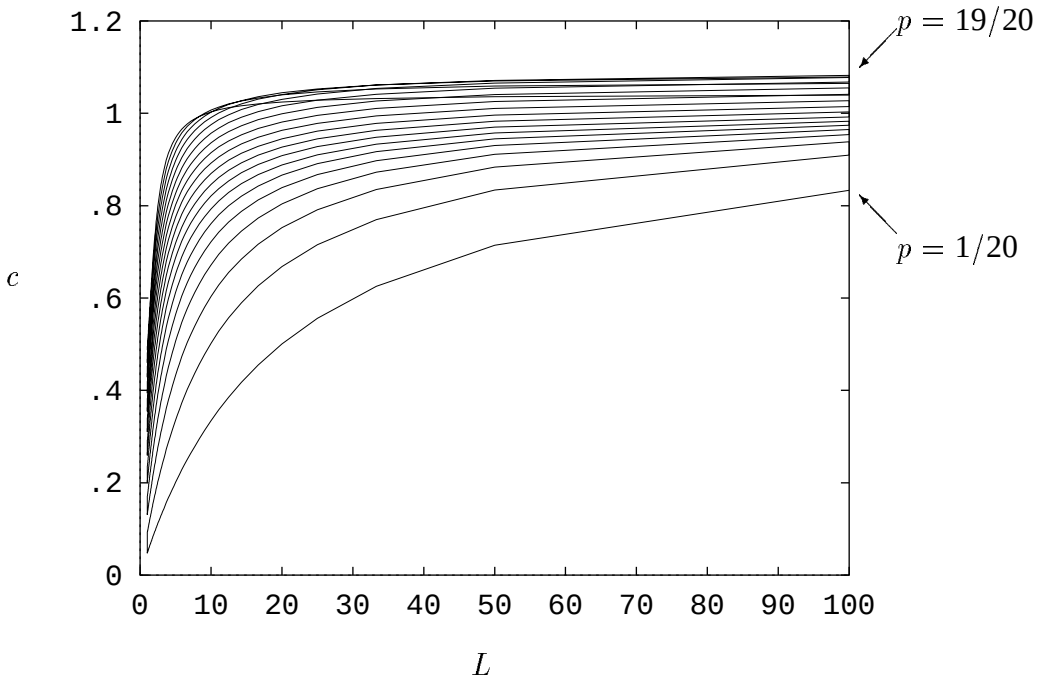

Figure 2: Mean number of messages on a node versus mean number of hops

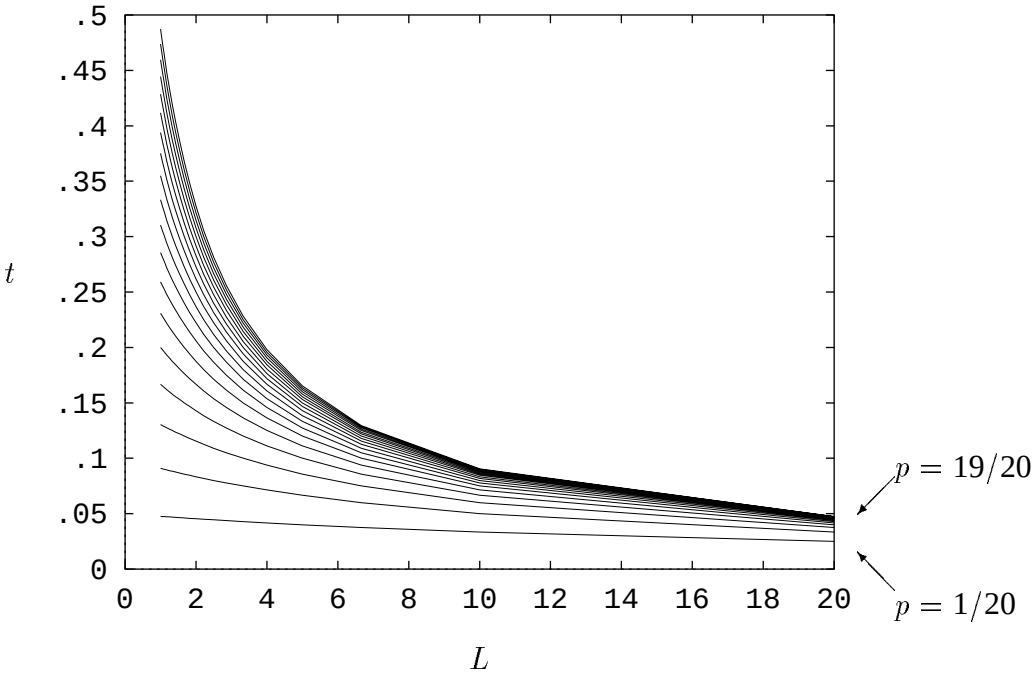

Figure 3: Mean node throughput versus mean number of hops

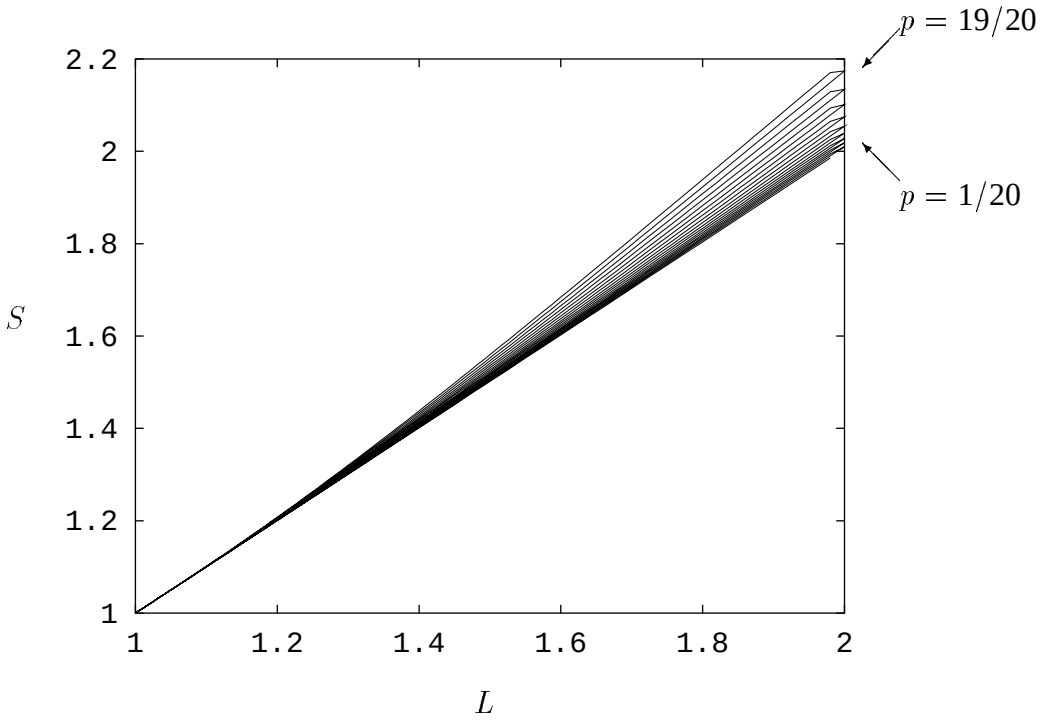

Figure 4: Mean sojourn time versus mean number of hops

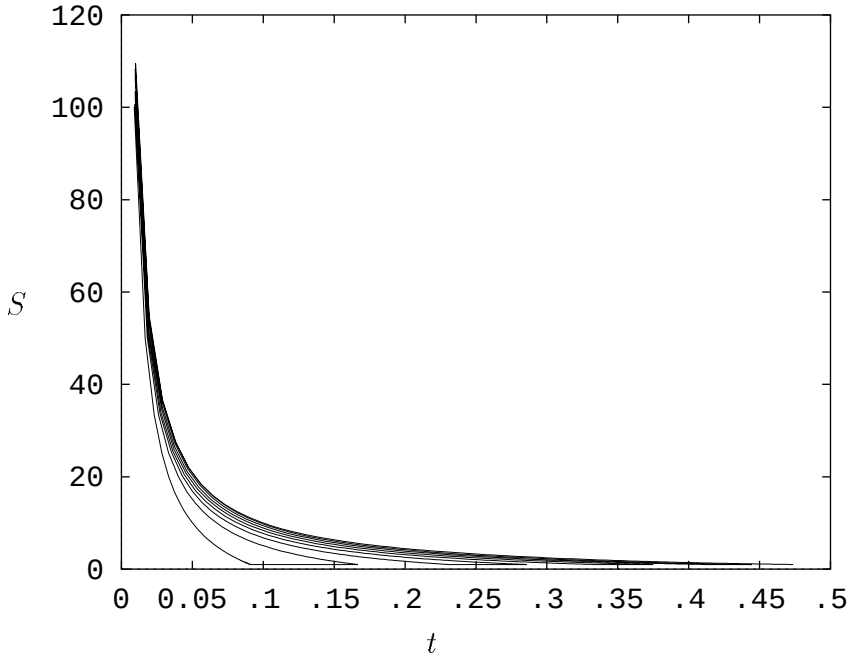

Figure 5: Mean sojourn time versus mean node throughput

References

[1] A. W. Roscoe, *Routing messages through networks : an exercise in deadlock avoidance*. Proceedings of OPPT, Grenoble, September 14–16th, 1987, T. Muntean ed.

[2] D. Freedman, *Markov Chains*, Springer Verlag, 1983.

[3] INMOS Ltd., `OCCAM2` *reference manual*, Prentice–Hall, 1988.

[4] L. Kleinrock *Queueing Systems*, John Wiley & Sons, 1975.

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399