Performance Analysis of Packet Switching Interconnection Networks with Finite Buffers

Aristotel Tentov and Aksenti Grnarov

Computer Science Department, Faculty of Electrical Engineering
University "St. Kiril i Metodij"
Karpos b.b., 91000, Skopje, Macedonia
E-mail: toto@cerera.etf.ukim.edu.mk

Abstract

In this paper, a mathematical method for analysis of synchronous packet-switching interconnection networks with finite buffering capacity at the output of switching elements is presented. The proposed mathematical method is general in that it analyzed interconnection networks under uniform and nonuniform traffic with blocking. The existing methods for analysis of buffered interconnection networks have assumed either single or infinite buffers at each input (or output) port of a switch, as well as uniform traffic pattern of the networks. Firstly, in the paper, a general model of synchronous buffered switching element, using output buffering, under assumption of finite buffer size for a very general class of traffic, is presented. Traffic can be uniform or nonuniform. It is assumed that the subsequent stages of the network are nearly independent and a model is extended for entire network under this assumption. Analytical results obtained with proposed model are then compared with each other, and it is shown that the proposed mathematical method is more general then the known models of interconnection networks.

Index Terms - multistage interconnection networks, performance analysis, finite buffer size, nonuniform traffic distribution, blocking probability.

1. Introduction

There is a great demand for application-specific multistage interconnection networks (MIN) in areas such as multiprocessor systems, high-speed communications based on the asynchronous transfer mode (ATM) and parallel computer systems.

MIN are used for interconnection of a large number of processors and memory modules in multiprocessor systems and for the resource arbitration and token distribution in the data flow computers [5], as main building blocks of switch fabrics in ATM switches [11], and for the high speed packet-switching computer communications in parallel computers [1]. In all these applications, MIN are a critical part of the overall system performance. Consequently, extensive studies have been done on characterizing their performance behavior.

Dias and Jump [7] analyzed the performance of delta networks with single buffers based on the timed Petri-net model. Robertazzi [13] present analysis of the performance of a packet switch based on single-buffered banyan networks, which was presented in the earlier work done by Jenq [14]. Kruskal and Snir [3], [4] studied the performance of banyan networks with infinite buffers. The performance of MIN's (multiple but finite buffers at each input port of a switch) has only partially been known through simulation [5], [7]. Common characteristic of all of this performance analysis was that they have considered only banyan (delta) networks based on (2x2) crossbar switches. These networks are unique path networks, which provide a unique path for each input-output connection.

The performance of single and multiple buffered delta networks constructed from switching elements (SE's) of arbitrary sizes and types (crossbar or bus), as well as the performances of single and multiple buffered, multiple-path networks (plus-minus-2^i networks) have been analytically studied by H. Yoon *et al.* [8]. Theimer *et al.* [10] expanded Jenq's model by introducing a blocked state for the single-buffered banyan network composed of $2x2$ switching elements (SE). Hsiao and Chen [12] also considered the blocked state for the single-buffered banyan networks. Ding and Bhuyan [6] showed that the performance of MIN can be improved by adopting the concept of small clock cycle using a model similar to Yoon's [8]. This models are based on some unrealistic assumptions or too complex to be generalized for multibuffered MIN's or large size SE.

Mun and Youn [15] modeled multibuffered MIN's with

1089-6503/96 $5.00 © 1996 IEEE
Proceedings of EUROMICRO-22

$2x2$ SE, with finite buffers on each *input* port. Their model considered the correlation of packet movements between two adjacent stages as well as subsequent network cycles, and it takes the blocked state into account. This model however is not generalized for nonuniform network traffic and for switches with different sizes. Labarta, Domingo and Casals [9] analyzed omega networks based on $2x2$ SE, with finite buffers on each output port, under assumption of uniform traffic.

In our previous paper [2], we presented an analytic model, which can be used to analyze MIN's constructed from SE's of an arbitrary size and type (crossbar or bus), with finite size buffers on each output port, under a very general class of uniform and nonuniform traffic. One of the shortcomings of this model is that it didn't take the effect of blocking into account.

In this paper, a general analytic model for the performance evaluation of MIN's is presented. It effectivelly models MIN's from SE's of various sizes and types, with finite buffers on output ports, under assumption of uniform and nonuniform traffic thaking into account effect of blocking. Characteristics of the model, as well as necessary assumption which were made in the process of its development give the assurance that the model is general and that it covers a great deal of different types of MIN's.

The rest of the paper is organized as follows. Delta networks and their main characteristics are briefly described in Section 2. The new analytic method for performance analysis of MIN's with finite output buffers is introduced in Section 3. In section 4, the proposed analytic model is used to analyze the performance of synchronous MIN's in the case of uniform and nonuniform traffic pattern. Conclusions are given in Section 5.

2. Delta Networks

Interconnection networks connect source and destination elements through stages of switches. An *(N x N)* delta network consists of *n* stages of *N/s (s x s)* crossbar switches, where $N=s^n$. A packet movement through the network can be controlled locally at each SE by a single *base-s* digit of the destination address. An example of an *(8 x 8)* delta network is given in Fig. 1.

Basic building block of an interconnection network is *s-input, s-output (sxs)* buffered switch (Fig. 2). Each input port can accept one packet per clock cycle, and route it to the appropriate output port. Each output port has a FIFO buffer. Conflicts between messages simultaneously routed to the same output port are resolved by queuing the messages. In this paper, it is assumed that the output buffer at each output port has a finite length (capacity). Each output port in a clock cycle can accept a number of arriving packets from the input ports, with limitation imposed by the finite capacity.

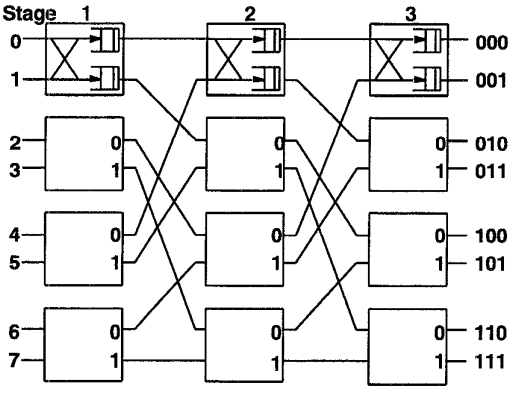

Figure 1. Model of 8x8 Delta network

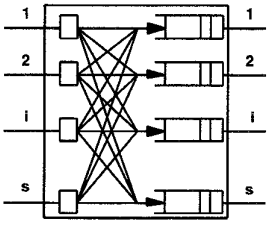

Figure 2. Model of sxs switch (capacity of output queues is m)

A synchronous type of switch operates at a rate of τ (called stage cycle), which consists of two phases:

- In the first phase (τ_1), the availability of the buffer space at the subsequent stage along the destined path of a packet in the current output buffer is determined, and the packet is informed whether it may go to the next stage or should stay in the current buffer.
- In the second phase (τ_2), packets are moving forward one stage if there are space in the subsequent stage, or they remain blocked in the current stage.

3. Performance analysis

Interconnection network is analyzed under the following assumptions:

- Packets are generated at each source node on the first stage with equal probability p, with the assumption of maximum load environment. At each network input port, packets are generated at the rate the network accepts them (there is always a packet waiting at each input), and at the network output ports, packets are served immediately;
- Number of buffers (m) at each output port of SE's is finite and equal for all stages;
- The number of packets arriving at successive cycles to an output port are independent, identically distributed random variables. These random variables may have a different distribution at different ports, and are clearly dependent from port to port;
- The number of cycles required to forward a packet (service time) for successive packets at an output port is random variable, which may vary from port to port.
- Due to the low correlation between successive stages [4], it is assumed that they are independent and characterized by a single switching element.

Accepted packets at each input port, are first submitted to the routing policy and then conceptually join a pool of output buffers (m-buffers). The routing decision is assumed to be fair, hence independent of buffer utilization. A given port can only serve packets routed to that particular port. Arriving packets who find no space in their respective queues are blocked in the previous stage. Accepted (non rejected) packets are served on a First-Come-First-Serve (FCFS) basis. Consequently, the major difference between between input and output buffering is in the different possible number of incoming packets. While in the case of input buffers, only a single packet may be transmitted between particular ports of stages per stage cycle, in the case of output buffering maximum s packets can be destined to an output port per stage cycle. Thus, a nonboundary state j can be reached from s previous states, j state, and state $j+1$. The complete state transition diagram of an m-buffer output port, under such assumptions is given in Fig. 3.

Before proceeding with analisys, the following variables are defined in the same manner as in [8] and [2]:

- n - number of switching stages;
- m - number of buffers in each output buffer (port);
- $p_j(k,t)$ - probability that there are j packets in a buffer module of an SE at stage k at the beginning of the $t-th$ stage cycle, $0 \leq j \leq m$;
- $p_{b_j}(k,t)$ - probability that the buffer module of an SE at stage k at is blocked the beginning of the $t-th$ stage cycle, $0 \leq j \leq m$;

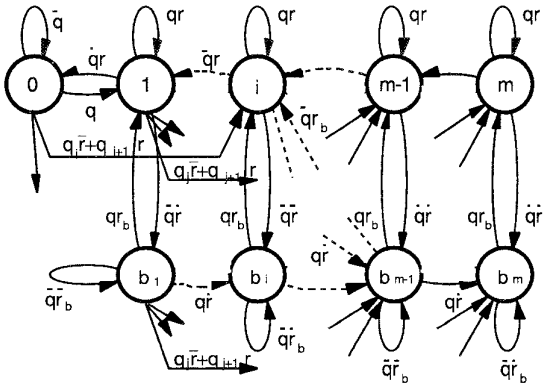

Figure 3. State transition diagram of the output port with m buffers

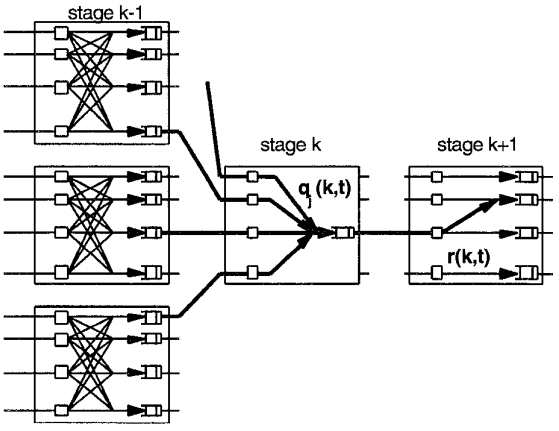

Figure 4. Illustration of $q_j(k,t)$ and $r(k,t)$

- $q_j(k,t)$ - probability that j packets are ready to come to a buffer of an SE at stage k during the $t-th$ stage cycle;
- $r(k,t)$ - probability that a packet in a *nonblocked* buffer of an SE at stage k is able to move forward during the $t-th$ stage cycle, given that there is at least one packet in the buffer module, i.e. the buffer module is not empty;
- $r_b(k,t)$ - probability that a packet in a *blocked* buffer of an SE at stage k is able to move forward during the $t-th$ stage cycle, given that there is at least one packet in the buffer module, i.e. the buffer module is not empty;

Based on the state transition diagram (Fig. 3.), definitions of the state variables and the illustration presented on the

Fig. 4., the following set of state equations holds for delta networks with m output buffers in each output port:

$$q_j(k,t) = \binom{s}{j} \cdot \left(\frac{p}{j}\right)^j \cdot \left(1-\frac{p}{j}\right)^{(s-j)}, \quad k=1 \quad (1)$$

$$q_j(k,t) = \binom{s}{j} \cdot \left(\frac{\overline{p_0}(k-1,t)}{j}\right)^j \cdot \left(1-\frac{\overline{p_0}(k-1,t)}{j}\right)^{(s-j)}, \quad 2 \le k \le n \quad (2)$$

for the uniform traffic.

In the case of nonuniform traffic each input is likely to have distinct favorite output port (e.g., the output port connecting a processor to its private memory). Each input port sends arriving packets to its favorite output port with probability q, and sends them with probability $(1-q)/s$ to each output port, including its favorite output port. The distribution of packets at the output ports is the sum of two terms: the first term accounts for normal packets and is essentially the same as equation (2) with p replaced by $p(1-q)$, and the second term accounts for favored packets. Under this assumptions, $q_j(k,t)(2 \le k \le m)$ is given by the following equation:

$$\begin{aligned} q_j(k,t) = {} & \binom{s-1}{j} \cdot \left(\frac{\overline{p_0}(k-1,t)}{j}\right)^j \\ & \cdot \left(1-\frac{\overline{p_0}(k-1,t)}{j}\right)^{(s-j)} \\ & \cdot \left(1-p\cdot\left(\frac{q+(1-q)}{s}\right)\right) \\ & + \binom{s-1}{j-1} \cdot \left(\frac{\overline{p_0}(k-1,t)}{j}\right)^{(j-1)} \\ & \cdot \left(1-\frac{\overline{p_0}(k-1,t)}{j}\right)^{(s-j+1)} \\ & \cdot \left(q+\frac{(1-q)}{s}\right) \end{aligned} \quad (3)$$

$$\begin{aligned} r(k,t) = {} & \overline{p_0}(k,t) \cdot [p_{m-1}(k,t) + p_{b_{m-1}}(k,t) \\ & + p_{b_m}(k,t) \cdot r(k+1,t) + p_m(k,t) \\ & \cdot r(k+1,t)] \end{aligned} \quad (4)$$

$$\begin{aligned} p_j(k,t) = {} & \sum_{l=1}^{j+1} p_l(k,t-1) \cdot q_{j-l+1}(k,t) \cdot r(k,t) \\ & + \sum_{l=0}^{j} \cdot p_l(k,t-1) \cdot q_{j-l}(k,t) \cdot \overline{r}(k,t) \\ & + \overline{q}(k,t) \cdot r(k,t) \cdot p_{j+1}(k,t-1) \\ & + \overline{q}(k,t) \cdot r_b(k,t) \cdot p_{j+1}(k,t-1) \\ & + q(k,t) \cdot r_b(k,t) \cdot p_{b_j}(k,t-1) \\ & + q(k,t) \cdot r(k,t) \cdot p_j(k,t-1), \\ & 0 \le j \le s, 1 \le k \le n \end{aligned} \quad (5)$$

$$\begin{aligned} p_j(k,t) = {} & \sum_{l=j-s+1}^{j+1} p_l(k,t-1) \cdot q_{j-l+1}(k,t) \cdot r(k,t) \\ & + \sum_{l=j-s}^{j} p_l(k,t-1) \cdot q_{j-l}(k,t) \cdot \overline{r}(k,t) \\ & + \overline{q}(k,t) \cdot r(k,t) \cdot p_{j+1}(k,t-1) \\ & + \overline{q}(k,t) \cdot r_b(k,t) \cdot p_{j+1}(k,t-1) \\ & + q(k,t) \cdot r_b(k,t) \cdot p_{b_j}(k,t-1) \\ & + q(k,t) \cdot r(k,t) \cdot p_j(k,t-1), \\ & s < j < m, 1 \le k \le n \end{aligned} \quad (6)$$

$$\begin{aligned} p_{b_j}(k,t) = {} & \sum_{l=1}^{j+1} p_{b_l}(k,t-1) \cdot q_{j-l+1}(k,t) \cdot r(k,t) \\ & + \sum_{l=0}^{j} p_{b_l}(k,t-1) \cdot q_{j-l}(k,t) \cdot \overline{r}(k,t) \\ & + \overline{q}(k,t) \cdot \overline{r}_b(k,t) \cdot p_{b_j}(k,t-1) \\ & + \overline{q}(k,t) \cdot \overline{r}(k,t) \cdot pj(k,t-1) \\ & + q(k,t) \cdot \overline{r}(k,t) \cdot p_{j-1}(k,t-1), \\ & 0 \le j \le s, 1 \le k \le n \end{aligned} \quad (7)$$

$$\begin{aligned} p_{b_j}(k,t) = {} & \sum_{l=j-s+1}^{j+1} p_{b_l}(k,t-1) \cdot q_{j-l+1}(k,t) \cdot r(k,t) \\ & + \sum_{l=j-s}^{j} p_{b_l}(k,t-1) \cdot q_{j-l}(k,t) \cdot \overline{r}(k,t) \\ & + \overline{q}(k,t) \cdot \overline{r}_b(k,t) \cdot p_{b_j}(k,t-1) \\ & + \overline{q}(k,t) \cdot \overline{r}(k,t) \cdot pj(k,t-1) \\ & + q(k,t) \cdot \overline{r}(k,t) \cdot p_{j-1}(k,t-1), \\ & s < j < m, 1 \le k \le n \end{aligned} \quad (8)$$

$$
\begin{aligned}
p_m(k,t) &= \sum_{l=j-m+1}^{m} p_l(k,t-1)\cdot q_{m-l+1}(k,t)\cdot r(k,t) \\
&\quad + \sum_{l=m-s}^{m} p_l(k,t-1)\cdot q_{m-l}(k,t)\cdot \overline{r}(k,t) \\
&\quad + q(k,t)\cdot r_b(k,t)\cdot p_{b_m}(k,t-1) \\
&\quad + q(k,t)\cdot r(k,t)\cdot p_m(k,t-1), \\
&\quad 1 \le k \le n
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
p_{b_m}(k,t) &= \sum_{l=j-m+1}^{m} p_{b_l}(k,t-1)\cdot q_{m-l+1}(k,t)\cdot r(k,t) \\
&\quad + \sum_{l=m-s}^{m} p_{b_l}(k,t-1)\cdot q_{m-l}(k,t)\cdot \overline{r}(k,t) \\
&\quad + \overline{q}(k,t)\cdot \overline{r}_b(k,t)\cdot p_{b_m}(k,t-1) \\
&\quad + \overline{q}(k,t)\cdot \overline{r}(k,t)\cdot p_m(k,t-1) \\
&\quad + q(k,t)\cdot \overline{r}(k,t)\cdot p_{m-1}(k,t-1), \\
&\quad 1 \le k \le n
\end{aligned} \tag{10}
$$

In the steady state, a packet arrives at an output port of the network with probability $r(n)$, which is the normalized throughput S.

$$S = r(n)\cdot \overline{p_0}(n) + r_b(n)\cdot \overline{p_{b_0}}(n) \tag{11}$$

Delay normalized with respect to the number of switching stages, i.e. average number of stage cycles taken for a packet to pass a single stage is given by the following equation:

$$D = \left(\frac{1}{n}\right)\cdot \sum_{k=1}^{n} \frac{1}{R(k)} \tag{12}$$

where $R(k)$ is:

$$R(k) = r(k)\cdot \sum_{i=1}^{m} \left(\frac{p_i(k) + p_{b_i}(k)}{\overline{p_0}(k)}\right)\cdot \frac{1}{i} \tag{13}$$

4. Analytic Results

The equations presented in previous section can be used to determine the normalized throughput and the normalized delay of asynchronous and synchronous network with the following parameters:

- SE size s,
- Network size $n (n = \log_s N)$,
- Buffer size m,

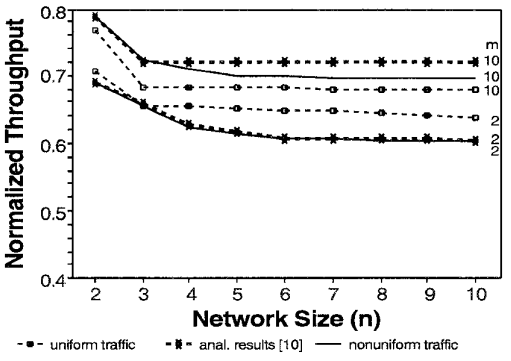

Figure 5. Normalized throughput vs. network size $(s = 2, q(1) = 1)$

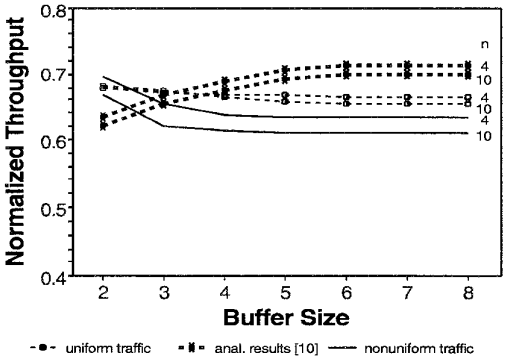

Figure 6. Normalized throughput vs. buffer size $(s = 2, q(1) = 1)$

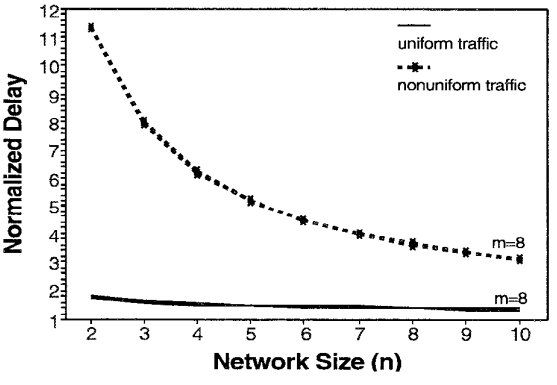

Figure 7. Normalized delay vs. network size $(s = 2, q(1) = 1)$

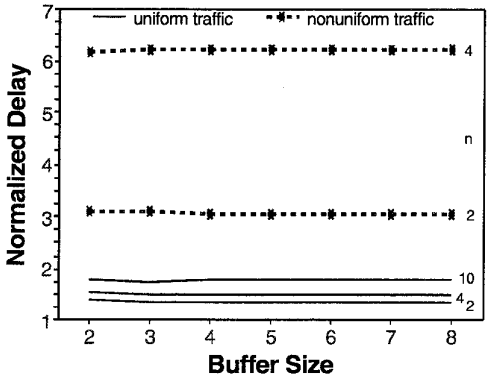

Figure 8. Normalized delay vs. buffer size $(s = 2, q(1) = 1)$

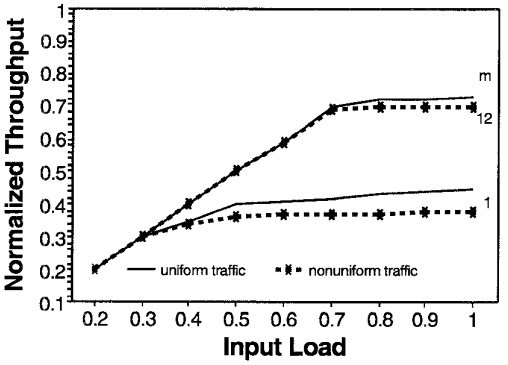

Figure 9. Normalized throughput vs. input load $(s = 2, q(1) = 1, n = 6)$

- Input load applied to the network q(1).

Among the many possible combinations of parameters, the most interesting cases are computed and plotted in Fig.5-9. Fig. 5 and 6 show normalized throughput versus network size, and normalized throughput versus buffer size, respectively, for various buffer sizes and for various network sizes. Fig. 7 and 8 show normalized delay versus network size, and normalized delay versus buffer size, respectively, for various buffer sizes and variuos network sizes. Finally, normalized throughput versus input load, are shown in Fig. 9, in the case of (64 x 64) delta networks.

From the figures it can be seen that the derived values for normalized delay and normalized throughput are more pesimistic in the case of nonuniform traffic, compared with uniform traffic. This means that the presented model with the assumption of nonuniform traffic, is closer to the real situation, compared with models which assumed uniform traffic pattern distribution, despite the fact that it is based on the assumption of nonindependent subsequent stages.

5. Conclusions

In this paper the new analytic model for analysis MIN with finite size of output buffers, under the assumption of blocking and nonuniform traffic, is introduced. A model who analyzed MIN under such conditions of networks does not exist in the known literature. Previously, the performance of this networks had been known for limited cases only through simulation, except for the case of unlimited buffers. The analytic model for the synchronous network is based on the work of Robertazzi [14], Yoon [8], and especially on the our paper [2]. Yoon et al. [8] in their paper, have also analyzed synchronous delta networks, under the assumption of input buffering, blocking and uniform network traffic. Main characteristic of presented model is the assumption that the buffering is done at the output of the switch instead at input. Also, it is assumed that the buffers have limited length, which is true in the real systems. Main achievement obtained with this model is the opportunity to analyze network behaviour in the case of blocking and nonuniform distribution of input traffic which is of course more realistic than assumption of blocking and uniform traffic. Model is presented on the delta networks, but this is done only for the simplicity of explanation. It is general enough for the types of networks for which it is developed, and it can handle networks with arbitrary network and switch sizes.

References

[1] —. Design of a broadcast packet switching network. *Proc. IEEE INFOCOM*, pages 667–675, April 1986.

[2] A. Tentov and A. Grnarov. Performance analysis of synchronous multibuffered packet-switching networks. *Microprocessing and Microprogramming*, 38(1-5), September 1993.

[3] C. P. Kruskal and M. Snir. The performance of multistage interconnection networks for multiprocessors. *IEEE Trans. Comput.*, C-32(12):1091–1098, December 1983.

[4] C. P. Kruskal and M. Snir. The distribution of waiting times in clocked multistage interconnection networks. *IEEE Trans. Comput.*, 37(11):1337–1352, November 1988.

[5] C. Y. Chin and K. Hwang. Packet switching networks for multiprocessors and data flow computers. *IEEE Trans. Comput.*, C-33(11):1014–1021, April 1983.

[6] J. Ding and L. N. Bhuyan. Finite buffer analysis of multistage interconnection networks. *IEEE Trans. Comput.*, 43(2):243–247, February 1994.

[7] D. M. Dias and J. R. Jump. Analysis and simulation of buffered delta networks. *IEEE Trans. Comput.*, C-30(4):273–282, April 1981.

[8] H. Yoon, K. Y. Lee and M. T. Liu. Performance analysis of multibuffered packet-switching networks in multiprocessor systems. *IEEE Trans. Comput.*, 39(3):319–327, March 1990.

[9] J. Labarta, J. Domingo and O. Casals. Performance evaluation of packet switched omega networks with finite buffers. *Queuing Networks with Blocking*, 1989.

[10] T. H. Theimer, E. P. Rathgeb and M. N. Huber. Performance analysis of buffered banyan networks. *IEEE Trans. Commun.*, C-39(2):269–277, February 1991.

[11] P. Newman. Atm technology for corporate networks. *IEEE Communications Magazine*, 30(4):90–101, April 1992.

[12] S. H. Hsiao and C. Y. R. Chen. Performance analysis of single-buffered multistage interconnection networks. *Proc. Third IEEE Symp. Parallel Distributed Processing*, pages 864–867, December 1991.

[13] T.Robertazzi. *Computer Networks and Systems: Queueing Theory and Performance Evaluation.* Springer Verlag, New York, 1990.

[14] Y. C. Jenq. Performance analysis of a packet switch based on single-buffered banyan networks. *J. Select. Areas Commun.*, SAC-3(6):1014–1021, December 1983.

[15] Y. Mun and H. Y. Youn. Performance analysis of finite buffered multistage interconnection networks. *IEEE Trans. Comput.*, 43(2):153–162, February 1994.