

An Exact Model For Analysis of Shared Buffer Delta Networks With Arbitrary Output Distribution

Mahmoud Saleh

Dept. of Comp. Science and Comp. Engg.
LaTrobe University
Melbourne, 3083, Australia
Email: saleh@cs.latrobe.edu.au

Mohammed Atiquzzaman

Dept. of Elec. & Computer Systems Engg.
Monash University, Clayton,
Melbourne 3168, Australia
Email: atiq@eng.monash.edu.au

Abstract

In this paper, we have developed an exact model to evaluate the performance of Multistage Interconnection Networks using internal shared buffering. The model assumes general output distribution which enables one to study the performance of such networks under any desired output distribution. Among many possible output distributions, uniform, hot spot and favorite distributions are taken as examples in this paper. The model is validated by the comparison of numerical results with event simulation results.

1 Introduction

Multistage Interconnection Networks (MINs) have received increasing attention as switching architectures for broadband integrated services digital network (B-ISDN) and transport systems based on Asynchronous Transfer Mode (ATM). An ATM switch transfers all information in fixed length packets called *cells*, and is characterized by simplified protocols, high speed links, and high capacity switching nodes. MINs are particularly useful as the switching fabric of ATM switches because of the promising features they offer, such as modularity and decentralized routability. MINs have also been studied and implemented for interconnecting a large number of processors and memories in a multiprocessor system.

A MIN consists of a number of stages of small switching elements (SE) which are interconnected by a permutation function. In a Delta network [12], the destination address is decoded and used for routing in a particular stage's switching element (SE). Therefore, no central controller is needed for global routing. Delta network and its equivalent topologies such as Omega and inverse cube are among the blocking type networks. This means that packets may contend for the same outlet in an SE which results in a loss in the performance of the network. The performance of such networks can be increased by using a sorting network at the input of the network, or by having multiple paths between input/output pairs, or by using buffers to store the conflicting packets. Multiple path networks need additional control mechanisms to manage multiple submission of packets to different paths. Internally buffered networks employ buffers at the SEs inside the network. The packets losing contention at an SE are stored in the buffers

in the SE. The location of buffers in an SE is crucial in the implementation and performance of the network. Networks with buffers located at the inlets of the SEs suffer from head of line (HOL) blocking resulting in reduced throughput. Input queues with bypass mechanisms have been proposed to reduce the effect of HOL contention. Buffers may be placed at the outlets of the SEs, and the packets destined to a particular outlet of an SE are queued at the corresponding buffer. An output buffered $d \times d$ SE requires reduced buffer access time and internal speedup which is d times the switching speed of an input-buffered SE.

The use of dedicated buffers at the inputs or outputs results in the networks using such buffering schemes to have low buffer utilization for most unbalanced traffics. Shared buffers may be used in the SEs to increase the buffer utilization and the performance of the network. Buffers in a shared buffer SE may be used to accommodate traffic for all inlets and outlets of the SE in such a way that a packet coming to an inlet may be placed into any available shared buffer in the SE, and a packet in a buffer can be forwarded to any of the outlets. An SE employing shared buffers does not suffer from HOL blocking. In addition, unlike output buffering, buffer resources in shared buffering are allocated to the outputs which most need them, and not dedicated to a particular output regardless of its needs. Consequently, MINs constructed from shared buffer SEs have higher throughput, lower delay and better buffer utilization than networks constructed from input or output buffered SEs. Moreover, given the same amount of buffer, the shared buffer is the best choice in terms of packet loss rate [7, 10]. Since one of the important performance criteria for ATM networks is packet loss rate, a shared buffer architecture is very suitable for implementing ATM networks in B-ISDN.

Performance evaluation studies may be accomplished by simulation or analytical modeling. Although simulation enables one to closely study the behavior of a network, using simulation to estimate the probability of rare events and their effect on performance is problematic, because vast computational resources may be required to generate a sufficient number of events from which statistical estimates may be formed with adequate statistical

confidence [8]. In analytical modeling, on the other hand, the results are obtained much faster with no special attention to calculation of very small probabilities.

Turner [3] developed a model for Delta network with shared buffer SEs under uniform traffic distribution. His model assumes independence between buffer slots, and uses a flow control mechanism to avoid packet loss inside the network. In that model, the state space of the buffer in a shared buffer SE is represented by a vector whose elements represents the number of packets available in the buffer at a particular cycle. Turner's model was improved by Monterosso and Pattavina [1] and Bianchi [14]. The model presented in [1] considers a bidimensional representation of the states in which it is known how many packets in the shared buffer SE are destined to any outlet of the SE. Moreover, the model allowed packets to be lost inside the network, too. Bianchi [14] proposed two alternative models to [3] which offer accuracy at the expense of complexity. A model for a network using shared buffer SEs, operating under a uniform traffic pattern and global flow control policy, has been reported in [9].

Gianatti and Pattavina [15] studied shared buffer networks with nonuniform traffic patterns. However, in their model, the outputs of the MIN are divided such that a group of outputs are hot and the rest are cold. The number of SEs in the hot group is determined by $\log_d N$, where N is the network size, and d is the size of an SE. For example, for $N = 64$, and $d = 2$, they consider 32 hot, and 32 cold outputs. Hence, the model is not suitable for studying networks with a single hot output, i.e. networks where an output becomes more popular than the others.

Most of the above models use local flow control [3] to control packet movement between stages. In local flow control, a packet can be forwarded to the next stage depending on its state at the beginning of a cycle, whereas in global flow control simultaneous operation of forwarding and receiving packets during a cycle is allowed. Therefore, global flow control results in a higher throughput and better buffer utilization than local flow control.

The objective of this paper is to study the performance of Delta network with global flow control and operating under any output distribution using analytical and simulation techniques.

This paper is organized as follows. In Section 2, we describe the modeling assumptions and general load distribution. Construction of the corresponding simulator, and additional considerations are explained in Section 3. In Section 4, we examine our model with some numerical results, and compare the results with the simulation. Concluding remarks and further possible work are given in Section 5.

2 The Vectorial Model

We describe the state of a shared buffer SE of size d with a pair (s, \mathbf{V}) in which s is the total number of currently full buffers, and \mathbf{V} is a vector of size d whose elements indicate the number of packets which are to pass through

a particular outlet. In other words

$$\mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}, \sum_{j=1}^d v_j = s \leq B,$$

where v_j indicates the number of packets that pass through the j^{th} outlet of an SE, and B is the total capacity of the shared buffer in the SE. A variant of this approach has been used in [1] for a vectorial model developed for uniform traffic. In the pair (s, \mathbf{V}) , s is redundant, since the total number of packets which are in an SE's buffers is already known from \mathbf{V} . However, for convenience, we still use s as a separate argument in our notation.

In order to realize a general traffic pattern in a Delta network, all the SEs in a stage of the network should be distinct. We number stages and SEs in a stage in a Delta network as exemplified in Fig. 1.

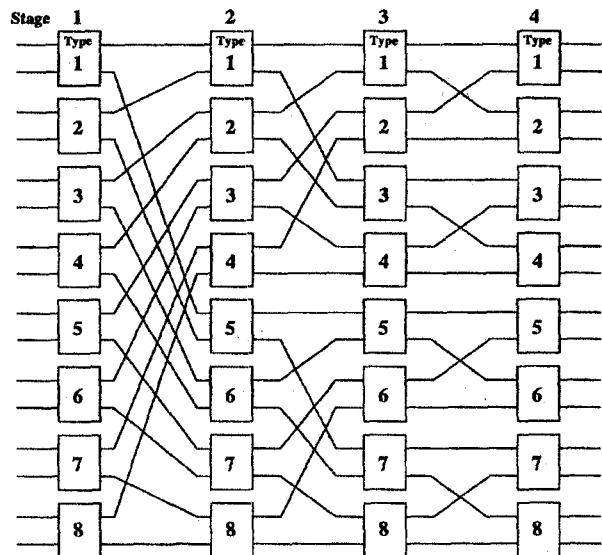


Figure 1: A 16×16 Delta-2 MIN.

For the purpose of analysis, we assume that the process of forwarding and accepting packets in each SE is accomplished in two phases [9]. In the *forward* phase, depending on the state of the SE and its downstream SEs, a number of packets may leave the SE, and the switch goes to an *intermediate* state. During the *receive* phase, the packets offered from upstream SEs are placed in the buffers, the corresponding acknowledgments are sent to the upstream SEs, and the SE goes to the final state. If the number of arriving packets is greater than the number of available buffers in the SE, a number of packets equal to the number of available spaces are selected randomly. The possible transitions of states in an SE for $d = 2$ and $B = 3$ are illustrated in Fig. 2.

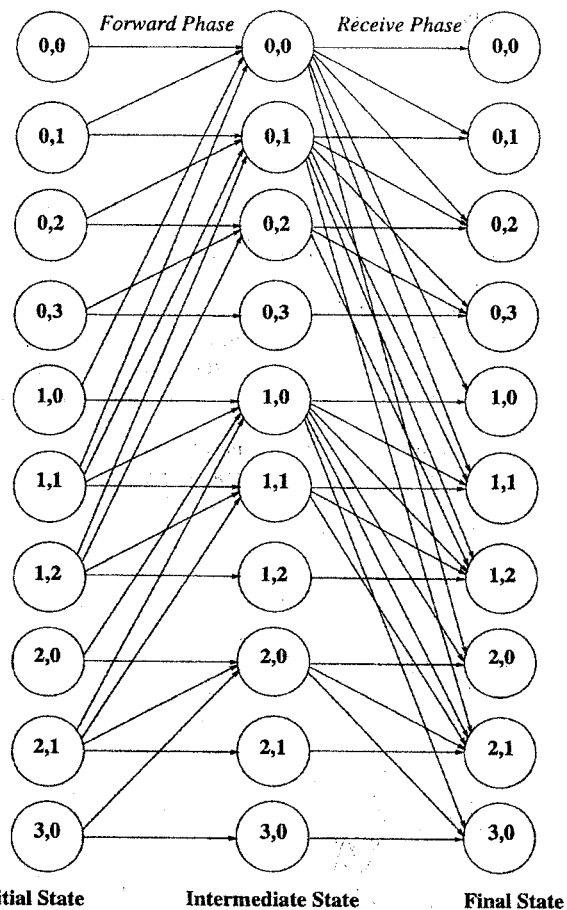


Figure 2: State diagram of a two phase network operation in an SE with $d = 2$, and $B = 3$.

2.1 Notation

Under a general output distribution, the mixture of the traffic in every SE in a MIN be different, therefore, all SEs in a stage are labeled with different numbers. A *type* r SE at stage i is the SE which is located at stage i and whose label number is r . The following notation will be used in the vectorial model.

$SE_{i,r}$: An SE of type r at stage i .

$\pi_{i,r,t}(s, \mathbf{V})$: Probability that $SE_{i,r}$ is in state (s, \mathbf{V}) at the beginning of cycle t .

$\tau_{i,r,t}(s1, \mathbf{V1}, s3, \mathbf{V3})$: Probability that $SE_{i,r}$ is in state $(s3, \mathbf{V3})$ at the beginning of the receive phase, given that it was in state $(s1, \mathbf{V1})$, at the beginning of the forward phase of cycle t , where $s3 \leq s1$.

$\sigma_{i,r,t}(s3, \mathbf{V3}, s2, \mathbf{V2})$: Probability that $SE_{i,r}$ is in state $(s2, \mathbf{V2})$ at the end of the receive phase of cycle t , given that it was in state $(s3, \mathbf{V3})$ at the beginning of the receive phase of the same cycle, where $s3 \leq s2$.

$\bar{\pi}_{i,r,t}(s3, \mathbf{V3})$: Probability that $SE_{i,r}$ is in state $(s3, \mathbf{V3})$ at the beginning of the receive phase of cycle t .

$a_{i,r,j,t}$: Probability that a packet is offered to inlet j of $SE_{i,r}$ during cycle t .

$b_{i,r,j,t}$: Probability that, during cycle t , a successor of $SE_{i,r}$ provides an acknowledgment to the j^{th} outlet of the SE, given that a packet was submitted to the successor through outlet j during the same cycle.

$u_{i,r,j}$: Probability that a packet in $SE_{i,r}$ is destined to its j^{th} outlet, where $1 \leq j \leq d$.

2.2 Load Distribution

In general, l_{mj} , the probability that output j of a network is referenced input m during a cycle, is equal to the probability of a packet being offered at that input times the probability that the packet is destined to the output under consideration.

$$l_{mj} = \rho_m q_{mj}, \quad (1)$$

where, ρ_m is the probability that a packet is offered at input m , and q_{mj} is the probability that a packet at input m is destined to output j . Therefore, load distribution \mathbf{L} of a network of size N may be expressed as the product of input load column vector \mathbf{P} and output distribution matrix \mathbf{Q} , where ρ and q have the same meaning as in Eq. (1).

$$\mathbf{L} = \mathbf{P}\mathbf{Q} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_N \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & \cdots & \cdots & q_{NN} \end{bmatrix}. \quad (2)$$

In this paper we assume that the Delta network has input rate ρ at every input j ; $1 \leq j \leq N$. Hence the load distribution \mathbf{L} reduces to

$$\begin{aligned} \mathbf{L} &= \rho \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & \cdots & \cdots & q_{NN} \end{bmatrix} \\ &= \begin{bmatrix} \rho q_{11} & \rho q_{12} & \cdots & \rho q_{1N} \\ \rho q_{21} & \rho q_{22} & \cdots & \rho q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho q_{N1} & \cdots & \cdots & \rho q_{NN} \end{bmatrix}. \quad (3) \end{aligned}$$

Output distribution matrix \mathbf{Q} depends on the specific output distribution chosen for the network. In uniform output distribution each output receives an equal ratio of the traffic coming from any input. Therefore, matrix \mathbf{Q} for uniform traffic is expressed as

$$\mathbf{Q}_u = \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \cdots & \frac{1}{N} \end{bmatrix}. \quad (4)$$

In hot spot distribution [16], a fraction h of the traffic from every input is directed to the *hot* output, and the rest is equally distributed to all of the outputs. For example if output 0 is hot, distribution matrix \mathbf{Q} will be

$$\mathbf{Q}_h = \begin{bmatrix} h + \frac{1-h}{N} & \frac{1-h}{N} & \dots & \frac{1-h}{N} \\ h + \frac{1-h}{N} & \frac{1-h}{N} & \dots & \frac{1-h}{N} \\ \vdots & \vdots & \ddots & \vdots \\ h + \frac{1-h}{N} & \dots & \dots & \frac{1-h}{N} \end{bmatrix}. \quad (5)$$

All to one distribution is a special case of hot spot distribution where $h = 1$. For *favorite* output distribution where input j sends a fraction f of its traffic to output j and equally distributes the rest to every output, matrix \mathbf{Q} will have the form

$$\mathbf{Q}_f = \begin{bmatrix} f + \frac{1-f}{N} & \frac{1-f}{N} & \dots & \frac{1-f}{N} \\ \frac{1-f}{N} & f + \frac{1-f}{N} & \dots & \frac{1-f}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-f}{N} & \dots & \dots & f + \frac{1-f}{N} \end{bmatrix}. \quad (6)$$

Single source to single destination (SSSD), also known as identity distribution, is a special case of favorite distribution where $f = 1$.

$u_{i,r,j}$, the probability that a packet in $\text{SE}_{i,r}$ is destined to the j^{th} outlet of $\text{SE}_{i,r}$, is determined by:

$$u_{i,r,j} = \frac{e_{i,r,j}}{\sum_{h=1}^d e_{i,r,h}}, \quad (7)$$

where $e_{i,r,j}$ is the sum of the distribution of all outputs which are accessible from outlet j in $\text{SE}_{i,r}$.

$$\begin{aligned} e_{i,r,j} &= \sum_{h=\underline{L}_{i,r,j}}^{\bar{I}_{i,r,j}} \sum_{c=\underline{Q}_{i,r,j}}^{\bar{O}_{i,r,j}} l_{hc}, \\ \underline{L}_{i,r,j} &= \begin{cases} rd + j & , i = 1 \\ \underline{L}_{i-1,\psi,0} & , 1 < i \leq k \end{cases} \\ \bar{I}_{i,r,j} &= \begin{cases} rd + j & , i = 1 \\ \bar{I}_{i-1,\psi,d-1} & , 1 < i \leq k \end{cases} \\ \underline{Q}_{i,r,j} &= \begin{cases} rd + j & , i = k \\ \underline{Q}_{i+1,\zeta,0} & , 1 \leq i < k \end{cases} \\ \bar{O}_{i,r,j} &= \begin{cases} rd + j & , i = k \\ \bar{O}_{i+1,\zeta,d-1} & , 1 \leq i < k \end{cases} \end{aligned} \quad (8)$$

$\underline{L}_{i,r,j}$, $\bar{I}_{i,r,j}$, $\underline{Q}_{i,r,j}$, and $\bar{O}_{i,r,j}$ are the lower bound of inputs, upper bound of inputs, lower bound of outputs, and upper bound of outputs which are accessible from outlet j of $\text{SE}_{i,r}$, respectively. These limits can be derived from the permutation function of a Delta network [12]. ψ and ζ in Eq. (8) are the types of SEs which are accessible from $\text{SE}_{i,r}$ at the previous and next stage respectively.

2.3 Description of the Model

In a shared buffer SE, the buffer will be in the intermediate state $\tilde{\pi}_{i,r,t}(s3, \mathbf{V3})$, if the SE is initially in state $\pi_{i,r,t}(s1, \mathbf{V1})$, and transition $\tau_{i,r,t}(s1, \mathbf{V1}, s3, \mathbf{V3})$ takes

place, enumerating for all possible initial states. In other words

$$\tilde{\pi}_{i,r,t}(s3, \mathbf{V3}) = \sum_{\mathbf{V1}} \pi_{i,r,t}(s1, \mathbf{V1}) \tau_{i,r,t}(s1, \mathbf{V1}, s3, \mathbf{V3}). \quad (9)$$

Similarly, the final state $\pi_{i,r,t+1}(s2, \mathbf{V2})$ is obtained, if the SE is in the intermediate state $\tilde{\pi}_{i,r,t}(s3, \mathbf{V3})$, and transition $\sigma_{i,r,t}(s3, \mathbf{V3}, s2, \mathbf{V2})$ takes place, summing for all possible intermediate states. The final state of an SE at cycle t is equal to the initial state of the SE at cycle $t+1$. Hence

$$\pi_{i,r,t+1}(s2, \mathbf{V2}) = \sum_{\mathbf{V3}} \tilde{\pi}_{i,r,t}(s3, \mathbf{V3}) \sigma_{i,r,t}(s3, \mathbf{V3}, s2, \mathbf{V2}). \quad (10)$$

In the rest of this paper, we consider the network in its steady state condition, and drop subscript t .

We define $Y_{i,r}^{d,h}$ as the list of all combinations of input traffic $a_{i,r}$'s (a total of $\binom{d}{h}$ elements) in stage $\text{SE}_{i,r}$ such that

$$\begin{aligned} Y_{i,r}^{d,h} &= \left\{ \sum a_{i,r,l_1} \times \dots \times a_{i,r,l_h} \times (1 - a_{i,r,k_1}) \times \dots \times \right. \\ &\quad \left. (1 - a_{i,r,k_\omega}) \mid l_1, \dots, l_h \in \{1, 2, \dots, d\}, \right. \\ &\quad \left. k_1, \dots, k_\omega \in \{1, 2, \dots, d\}, l_1 < \dots < l_h, \right. \\ &\quad \left. k_1 < \dots < k_\omega, l_1, \dots, l_h \neq k_1, \dots, k_\omega \right\}, \end{aligned} \quad (11)$$

where $\omega = d - h$. For example, if $d = 4$ and $h = 2$, then $Y_{i,r}^{4,2}$ will be:

$$\begin{aligned} Y_{i,r}^{4,2} &= \{ a_{i,r,1} a_{i,r,2} (1 - a_{i,r,3}) (1 - a_{i,r,4}) + \\ &\quad a_{i,r,1} (1 - a_{i,r,2}) a_{i,r,3} (1 - a_{i,r,4}) + \\ &\quad a_{i,r,1} (1 - a_{i,r,2}) (1 - a_{i,r,3}) a_{i,r,4} + \\ &\quad (1 - a_{i,r,1}) a_{i,r,2} a_{i,r,3} (1 - a_{i,r,4}) + \\ &\quad (1 - a_{i,r,1}) a_{i,r,2} (1 - a_{i,r,3}) a_{i,r,4} + \\ &\quad (1 - a_{i,r,1}) (1 - a_{i,r,2}) a_{i,r,3} a_{i,r,4} \}. \end{aligned} \quad (12)$$

To calculate $\sigma_{i,r}(s3, \mathbf{V3}, s2, \mathbf{V2})$, we consider two different cases depending on whether, after the intake of packets in the current cycle, the SE's buffers are all full or not.

1. $s2 < B$

In this case, every packet which would have wanted to enter $\text{SE}_{i,r}$ has actually entered the SE, and none is blocked due to lack of enough buffer space. Hence, it is only required to consider how many packets were offered to (and entered from) each inlet of the SE. This is equal to calculating the multinomial distribution of all offered packets.

$$\sigma_{i,r}(s3, \mathbf{V3}, s2, \mathbf{V2}) = (s2 - s3)! Y_{i,r}^{d,s2-s3} \prod_{j=1}^d \frac{u_{i,r,j}^{(v2_{i,r,j} - v3_{i,r,j})}}{(v2_{i,r,j} - v3_{i,r,j})!}. \quad (13)$$

2. $s2 = B$

In this case, it is possible that a packet which was offered to an inlet of $SE_{i,r}$ was not accepted due to fewer number of available buffers than the total number of offered packets. We assume that in case of contention, a number of packets equal to the number of available buffer spaces are accepted, regardless of their destinations.

$$\sigma_{i,r}(s3, \mathbf{V3}, s2, \mathbf{V2}) = (s2 - s3)! \prod_{j=1}^d \frac{u_{i,r,j}^{(v2_{i,r,j} - v3_{i,r,j})}}{(v2_{i,r,j} - v3_{i,r,j})!} \sum_{h=s2-s3}^d Y_{i,r}^{d,h} \quad (14)$$

$\tau_{i,r}$, the probability of forwarding the packets in $SE_{i,r}$ such that transition from $\mathbf{V1}$ to $\mathbf{V3}$ takes place is equal to the product of binomial distribution of the packets forwarded from each outlet of the SE.

$$\tau_{i,r}(s1, \mathbf{V1}, s3, \mathbf{V3}) = \prod_{j=1}^d \beta[\min(1, v1_j), v1_j - v3_j, b_{i,r,j}], \quad (15)$$

where

$$\beta(n, k, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (16)$$

$b_{i,r,j}$ consists of two parts. If the succeeding SE at the next stage is in the final states where total packets are less than B , after a packet from current outlet j was offered as well as other outlets, outlet j will definitely receive an acknowledgment. Otherwise, the probability that an acknowledgment will be received by outlet j depends on whether it wins the contention with other offered packets to the corresponding SE in the next stage.

$$b_{i,r,j} = \sum_{s2 < B} \pi_{i+1,\zeta}(s2, \mathbf{V2}) + \sum_{s2=B}^{B-1} \sum_{s3=B-d}^{B-1} \bar{\pi}_{i+1,\zeta}(s3, \mathbf{V3}) (s2 - s3)! \times \prod_{c=1}^d \frac{u_{i+1,\zeta,c}^{(v2_{i+1,\zeta,c} - v3_{i+1,\zeta,c})}}{(v2_{i+1,\zeta,c} - v3_{i+1,\zeta,c})!} \times \sum_{h=s2-s3}^d \frac{(s2 - s3)}{h} Y_{i+1,\zeta}^{d,h} \quad (17)$$

Subscript ζ in Eq. (17) denotes the type of SE which should be considered at the next stage.

If there is at least one packet destined to outlet j of $SE_{i,r}$, then a packet will definitely be offered to the j^{th} inlet of $SE_{i+1,r}$. Therefore, $a_{i,r,j}$, the probability that a packet is offered at the j^{th} inlet of $SE_{i,r}$, is determined as:

$$a_{i,r,j} = \begin{cases} \rho & , i = 1 \\ 1 - \sum_{\mathbf{V}; v_j=0} \pi_{i-1,\psi}(s, \mathbf{V}) & , i > 1 \end{cases} \quad (18)$$

where ψ is the type of the SE to which $a_{i,r,j}$ is connected. As in Eq. (18), the probability that a packet is offered to

an inlet of any SE at the first stage is equal to the input load of the network.

2.4 Performance Evaluation

In steady state condition of the network, the throughput, packet loss, and delay of various SE types can be computed.

Throughput of outlet j of $SE_{i,r}$ is equal to the sum of all possible transitions from initial state $(s1, \mathbf{V1})$ to intermediate state $(s3, \mathbf{V3})$, so that a packet leaves the SE from outlet j

$$\theta_{i,r,j} = \sum_{\mathbf{V1}} \pi_{i,r}(s1, \mathbf{V1}) \sum_{\mathbf{V3}} \tau_{i,r}(s1, \mathbf{V1}, s3, \mathbf{V3}) \quad (19)$$

Summing the throughputs of all outlets of $SE_{i,r}$, we get the overall throughput of that SE:

$$\theta_{i,r} = \sum_{j=1}^d \theta_{i,r,j} \quad (20)$$

Finally, the throughput of stage i is given by:

$$\Theta_i = \sum_{r=1}^{N/d} \theta_{i,r} \quad (21)$$

Since there is no packet loss inside the network, the overall throughputs of all stages are the same.

$$\eta = \frac{pN - \Theta_i}{pN} = \frac{p - \Theta_i/N}{p} \quad (22)$$

where Θ_i/N is the throughput per link at any stage i .

Delay of a packet leaving an outlet of an SE may be calculated using Little's formula for delay in which waiting time in a queue is equal to the average queue length divided by the arrival rate of the queue. In our vectorial model, the length of the logical queue of the outlet of $SE_{i,r}$ is known from the state vector of the SE. Thus, the delay of that outlet is determined by:

$$w_{i,r,j} = \frac{1}{\theta_{i,r,j}} \sum_{j=1}^B j \pi_{i,r}(s, \mathbf{V}) \quad (23)$$

The average delay in $SE_{i,r}$ is equal to the sum of the delays in logical queues of all outlets divided by d :

$$w_{i,r,av} = \frac{1}{d} \sum_{j=1}^d w_{i,r,j} \quad (24)$$

The average delay at stage i is equal to the sum of all $w_{i,r,av}$ for $1 \leq r \leq N/d$, divided by the number of SEs in the stage (N/d):

$$w_i = \frac{1}{N/d} \sum_{r=1}^{N/d} w_{i,r,av} \quad (25)$$

Finally, the average overall delay is obtained by summing the delays in different stages of the network:

$$W = \sum_{i=1}^k w_i, \quad (26)$$

where k is the number of stages in the network.

3 Simulation Study

The model presented in Section 2 is validated with a simulation study. The same assumptions as made for the analysis apply to the simulation of the network, and the following operations are carried out:

- At each cycle, a packet is generated with probability ρ (offered load to the network input). The generated packet is independent of the packets generated in previous cycles and at other input ports. Each packet consists of the following information:
 1. a source tag which denotes the input link at which the packet arrived,
 2. a destination tag denoting the output link to which the packet is destined, and
 3. the current cycle number, used for measurement of the packet delay in the network.
- Simulation results from the first several hundred cycles of the network operation are ignored to allow the network to reach steady state condition. The simulation program is then allowed to run until the change in the average throughput between consecutive cycles becomes less than 10^{-6} .
- Conflict in the buffers for accessing a particular outlet as well as contention to seize a buffer space in the next stage are resolved using a random number generator with a different seed value from that of the packet generator.

The network operates as follows:

1. The packets at the last stage buffers are sent to the output links of the network, and the instantaneous throughput and delay are measured for every link.
2. For each SE at stages $k - 1$ to 1:
 - The SE buffers are examined for packets passing the different outlets of the SE, copies of all packets passing different outlets are placed in the corresponding outlet lists, forming logical output queues, and the lists are sent to the corresponding inlets of the next stage.
 - If the number of available buffer spaces in the SE is less than the number of packets in the different lists at the inlets to the SE, a number of packets equal to the number of available spaces are chosen at random from the available

lists. Packets which are not accepted stay in the buffers at the previous stage until they can be forwarded in the subsequent cycles.

3. A new packet is generated at every input of stage 1 with probability ρ , taking to account the type of output distribution. The generated packet is then placed in the first stage's relevant buffer if there is any room. Otherwise, it is discarded and the packet loss counter is incremented by one.

4 Numerical Results

The model developed in Section 2 can be used for any output traffic distribution. For any distribution, the only thing that needs to be changed is the load distribution matrix L discussed in Section 2.2. In this section we examine the model under uniform, hot spot and favorite distributions.

The normalized throughput of a Delta network for $N = 64$ and $B/d = 2$ is illustrated in Figs. 3 and 4, for hot spot and favorite distribution, respectively. The proposed model is quite accurate when the input load is less than 0.7. The model gives accurate results for higher input rates under both hot spot and favorite distributions as hot or favorite ratio increases.

Favorite output distribution has a lesser impact on the overall throughput of Delta network than hot spot. This is due to the fact that in hot spot, traffic inside the network is concentrated in the hot switch as it flows towards the last stage. Therefore, all of the traffic including the portion which is destined to non-hot outputs is jammed in the concentrated switches, which adversely reduces the throughput of non-hot outputs. In favorite distribution, on the other hand, there are as many favorite outputs as inputs. So, the traffic inside the network is more balanced. Figs. 5 and 6 compare the average delay per stage for the same network and buffer size configurations. In these figures, too, the difference in the results with hot spot and favorite distribution confirms the discussion with respect to the difference in the nature of the two distributions.

Buffer Occupancy for hot spot and favorite distributions is illustrated in Fig. 7 and Fig. 8 respectively. For hot spot distribution, the logical output queue for the hot outlet tends to reach total buffer capacity for input loads of as low as 0.4, and hot spot value of 0.04. For a network with bigger B/d , the hot logical queue tends to allocate total buffer capacity more rapidly under the same input rate and hot spot value. The rate of hot occupancy is nonlinear as shown in Fig. 7. In a network with favorite output distribution such as the one shown in Fig. 8, on the other hand, although the total buffer occupancy is high, the outlet under study (outlet 0, to have a reasonable comparison with the hot spot traffic), has a much lesser occupancy rate than the hot spot distribution. Moreover, the rate of change in buffer occupancy, when the favorite value increases, is linear.

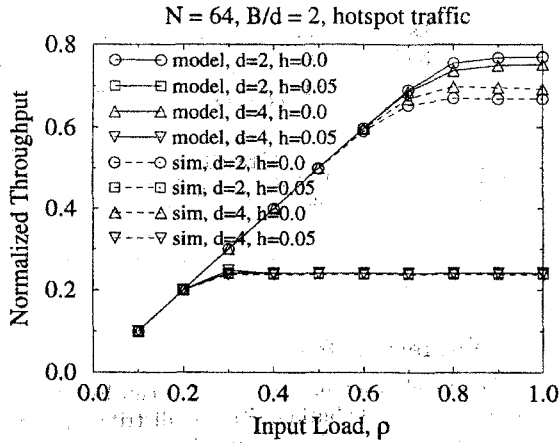


Figure 3: Throughput versus input load for $N = 64$ and $B/d = 2$ (hot spot).

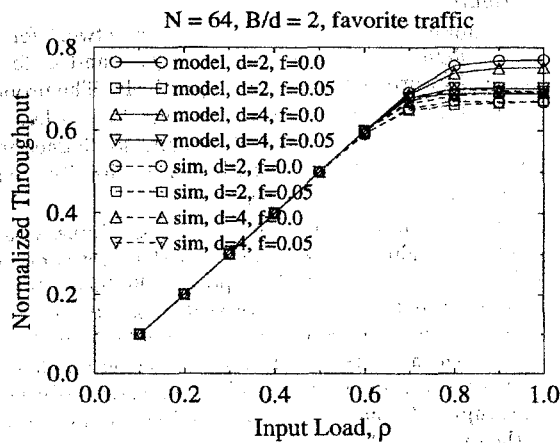


Figure 4: Throughput versus input load for $N = 64$ and $B/d = 2$ (favorite).

5 Conclusion

We have developed an analytical model to study the performance of multistage networks constructed from shared buffer switching elements with any SE size and buffer size. The proposed model may be used for analysis of multistage networks under an arbitrary traffic pattern. We have shown the results from the model for uniform, favorite, and hotspot output distributions, and compared their accuracy with simulation results. Numerical results show a close agreement with the results obtained from the simulation.

We have studied the performance merits of Delta network with global control policy. However, the model may be easily modified to be used for local flow control as well.

The proposed model is of modest computational cost when used for networks built from SEs with small number of inlets and outlets. However, since the number of states in an SE grows exponentially as the SE size in-

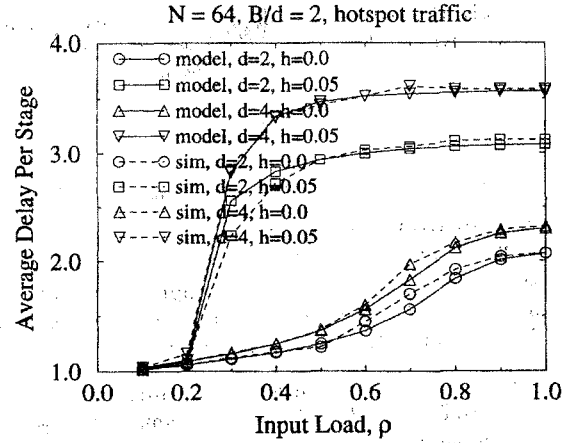


Figure 5: Average delay per stage versus input load for $N = 64$ and $B/d = 2$ (hot spot).

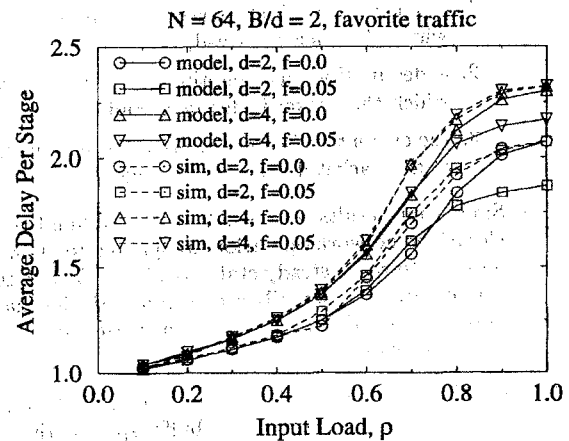


Figure 6: Average delay per stage versus input load for $N = 64$ and $B/d = 2$ (favorite).

creases, the model becomes computationally very expensive in that case. Nevertheless, it is still advantageous to use the model over simulation methods for measuring parameters such as packet loss and buffer occupancy.

References

- [1] A. Monterosso and A. Pattavina, "Performance analysis of multistage interconnection networks with shared-buffered switching elements for ATM switching," in *Proc. IEEE INFOCOM '92: Conference on Computer Communications*, Florence, Italy, pp. 124-131, May 1992.
- [2] H. Yoon, K.Y. Lee, and M.T. Liu, "Performance analysis of multibuffered packet-switching networks in multiprocessor systems," *IEEE Transactions on Computers*, vol. 39, no. 3, pp. 319-327, March 1990.

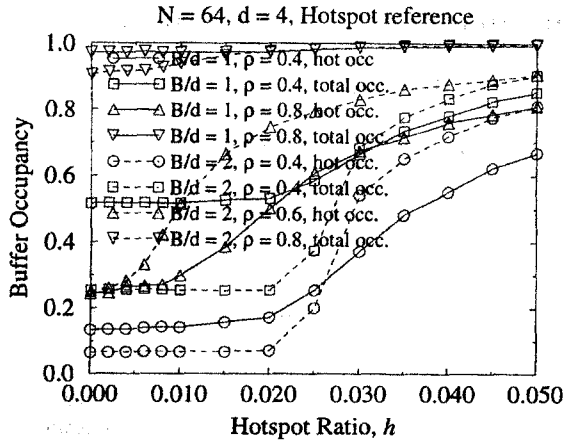


Figure 7: Buffer Occupancy of the first type SE at the first stage for $N = 64$, and hot spot output distribution.

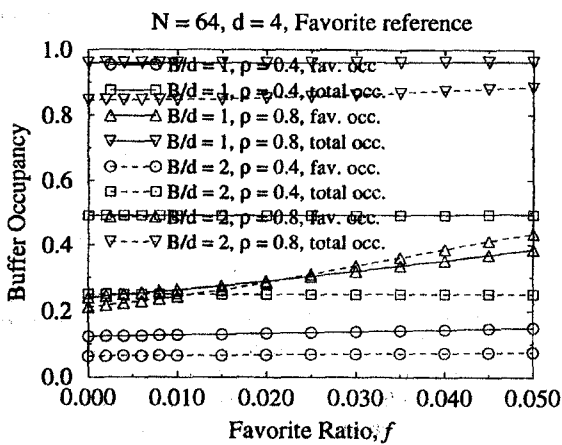


Figure 8: Buffer Occupancy of the first type SE at the first stage for $N = 64$, and favorite output distribution.

[3] J.S. Turner, "Queueing analysis of buffered switching networks," *IEEE Transactions on Communications*, vol. 41, no. 2, pp. 412-420, February 1993.

[4] M. Atiquzzaman and M.S. Akhtar, "Performance of buffered multistage interconnection networks in non uniform traffic environment," *Journal of Parallel and Distributed Computing*, Vol 30, No. 1, pp. 52-63, October 1995.

[5] Y-C. Jenq, "Performance analysis of a packet switch based on single-buffered Banyan network," *IEEE Journal on Selected Areas in Communications*, vol. SAC-1, no. 6, pp. 1014-1021, December 1983.

[6] T. Szymanski and S. Shaikh, "Markov chain analysis of packet-switched Banyans with arbitrary switch sizes, queue sizes, link multiplicities and speedups," in *Proc. IEEE INFOCOM '89: 8th Annual Joint*

Conference of the IEEE Computer and Communication Societies, Ontario, Canada, pp. 960-971, April 1989.

[7] Y. Sakurai, N. Ido, S. Gohara, and N. Endo, "Large-scale ATM multistage switching network with shared buffer memory switch," *IEEE Communications Magazine*, pp. 90-96, January 1991.

[8] V.S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, pp. 70-81, March 1994.

[9] M. Saleh and M. Atiquzzaman, "Queueing analysis of shared buffer switches for ATM networks," in *Proc. GLOBECOM 94: IEEE Global Telecommunications Conference*, San Francisco, California, pp. 1070-1074, November 1994.

[10] H. Kuwahara, N. Endo, M. Ogino, T. Kozaki, Y. Sakurai, and S. Gohara, "A shared buffer memory switch for an ATM exchange," in *Proc. IEEE International Conference on Communications*, pp. 118-122, 1989.

[11] M. Saleh and M. Atiquzzaman, "Buffer occupancy in ATM switches with single hot spot," *Electronics Letters*, vol. 31, no. 1, pp. 13-15, 5 January 1995.

[12] J.H. Patel, "Processor-memory interconnections for multiprocessors," in *Proc. 6th Annu. Symp. on Comput. Arch.*, New York, April 1979.

[13] M. Saleh and M. Atiquzzaman, "Performance of shared buffer switches under non-uniform traffic pattern", in *Proc. Australian Telecommunications Networking and Applications Conference, ATNAC'94*, Melbourne, Australia, pp. 283-287, December 1994.

[14] G. Bianchi and J.S. Turner, "Improved queueing analysis of shared buffer switching networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 482-490, August 1993.

[15] S. Gianatti and A. Pattavina, "Performance analysis of shared-buffered Banyan networks under arbitrary traffic patterns," in *Proc. IEEE INFOCOM '93: Conference on Computer Communications*, pp. 943-952, March 1993.

[16] G.F. Pfister and V.A. Norton, "Hot spot contention and combining in multistage interconnection networks," *IEEE Transactions on Computers*, Vol. C-34, no. 10, pp. 943-948, October 1985.