

Exact model for analysis of shared buffer ATM switches with arbitrary traffic distribution

M. Saleh and M. Atiquzzaman

Abstract: The authors have developed an exact model to evaluate the performance of multistage interconnection networks using internal shared buffering. The model assumes a general output distribution which allows the study of the performance of such networks under any desired output distribution. Among many possible output distributions, uniform, hot-spot and favourite distributions are studied. The model is validated by the comparison of some numerical results with event simulation results which are shown to be very close to the model.

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 characterised 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 decentralised 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 [1], 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. The delta network and its equivalent topologies such as omega and inverse cube are among 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 and this results 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.

Owing to the use of dedicated buffers for the inputs or outputs, the networks constructed from input or output buffered SEs have low buffer utilisation for most unbalanced traffics. Shared buffers may be used in the SEs to increase the buffer utilisation 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 are 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 utilisation 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 [2-4]. 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 modelling. Although simulation enables one to closely study the behaviour 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 [5]. In analytical modelling, however, the results are obtained much faster with no special attention to calculation of very small probabilities.

Turner [6] developed a model for the delta network with shared buffer SEs under uniform traffic distribution. His model assumes independence between buffer slots, and uses

© IEE, 2001

IEE Proceedings online no. 20010159

DOI: 10.1049/ip-com:20010159

Paper first received 1st February and in revised form 22nd November 2000

M. Saleh is with the Imam Hussein University, Tehran, Iran

M. Atiquzzaman is with the School of Computer Science, University of Oklahoma, Norman, OK 73019-6151, USA. E-mail: atiq@iecc.org

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_i 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 [7] and Bianchi and Turner [8]. The model in presented in [7] 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, that model allowed packets to be lost inside the network, too. Bianchi and Turner [8] proposed two alternative models to [6] 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 [10] 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 [6] 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 utilisation than local flow control.

The aim of this paper is to study the performance of a delta network with global flow control and operating under an arbitrary traffic pattern.

Our objectives are:

- (i) to develop a model for delta networks using shared buffer SEs and operating under a general traffic pattern
- (ii) to study the behaviour of shared buffer networks under different traffic patterns
- (iii) to study buffer utilisation at different SEs and different stages of a delta network.

2 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 [7] 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.

To realise a general traffic pattern in a delta network, all the SEs in a stage of the network should be distinct. We

number the stages and the SEs in a stage in a delta network as exemplified in Fig. 1.

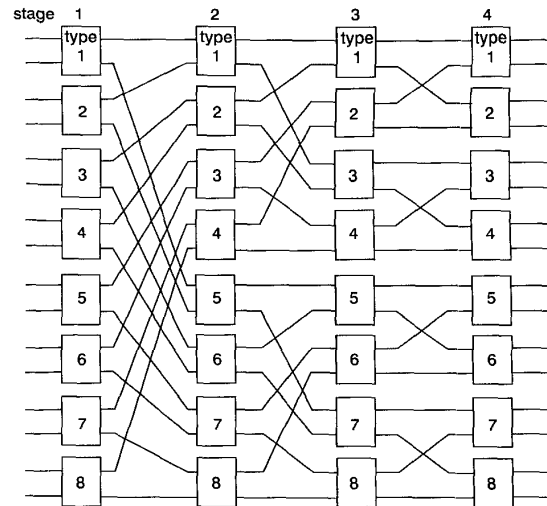


Fig. 1 16×16 delta-2 MIN

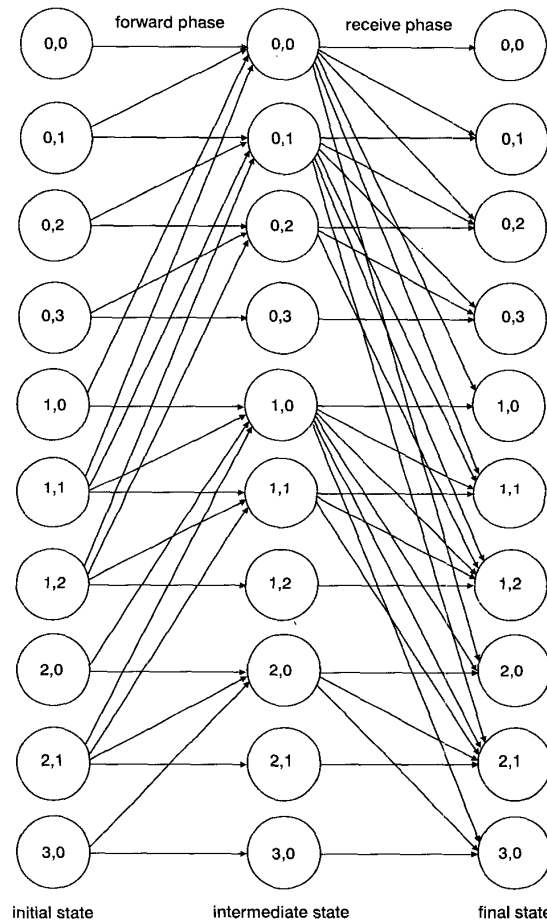


Fig. 2 State diagram of a two-phase network operation in an SE with $d = 2$ and $B = 3$

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.

2.1 Notation

Under a general output distribution, the mixture of the traffic in every SE in a MIN is different; therefore, all SEs in a stage are labelled 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

$\pi_{i,r,t}(s1, \mathbf{V1}, s3, \mathbf{V3})$ = 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 $s3 \leq s2$

$\tilde{\pi}_{i,r,t}(s3, \mathbf{V3})$ = 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 by input m during a cycle is equal to the probability of a packet being offered at that input multiplied by 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 meanings as in eqn. 1.

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

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

$$\mathbf{L} = \rho \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1N} \\ q_{21} & q_{22} & \dots & q_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ q_{N1} & \dots & \dots & q_{NN} \end{bmatrix}$$

$$= \begin{bmatrix} \rho q_{11} & \rho q_{12} & \dots & \rho q_{1N} \\ \rho q_{21} & \rho q_{22} & \dots & \rho q_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \rho q_{N1} & \dots & \dots & \rho q_{NN} \end{bmatrix} \quad (3)$$

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

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

In hot-spot distribution [11], 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

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

All to one distribution is a special case of hot spot distribution where $h = 1$. For favourite 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

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

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

$u_{i,r,j}$ the probability that a packet in $SE_{i,r}$ is destined to the j th outlet of the SE, 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 $SE_{i,r}$.

$$e_{i,r,j} = \sum_{h=\underline{I}_{i,r,j}}^{\bar{I}_{i,r,j}} \sum_{c=\underline{Q}_{i,r,j}}^{\bar{Q}_{i,r,j}} l_{hc}$$

$$\underline{I}_{i,r,j} = \begin{cases} rd + j & , i = 1 \\ \underline{I}_{i-1,\psi,0} & , 1 < i \leq k \end{cases}$$

$$\bar{I}_{i,r,j} = \begin{cases} rd + j & , i = 1 \\ \bar{I}_{i-1,\psi,d-1} & , 1 < i \leq k \end{cases}$$

$$\underline{Q}_{i,r,j} = \begin{cases} rd + j & , i = k \\ \underline{Q}_{i+1,\zeta,0} & , 1 \leq i < k \end{cases}$$

$$\bar{Q}_{i,r,j} = \begin{cases} rd + j & , i = k \\ \bar{Q}_{i+1,\zeta,d-1} & , 1 \leq i < k \end{cases} \quad (8)$$

$\underline{I}_{i,r,j}$, $\bar{I}_{i,r,j}$, $\underline{Q}_{i,r,j}$ and $\bar{Q}_{i,r,j}$ are the lower bound of inputs, upper bound of inputs, lower bound of outputs and upper bound of outputs which are accessible from output j of $SE_{i,r}$, respectively. These limits can be derived from the

permutation function of a delta network [1]. ψ and ζ in eqn. 8 are the types of SEs that are accessible from $SE_{i,r}$ at the previous and next stages, 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 $SE_{i,r}$ 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 $SE_{i,r}$ in the intermediate state $\tilde{\pi}_{i,r,t}(s3, \mathbf{V3})$ and transition $\tilde{\sigma}_{i,r,t}(s3, \mathbf{V3}, s2, \mathbf{V2})$ takes place, summing for all possible intermediate states. The final state of an $SE_{i,r}$ cycle t is equal to the initial state of the $SE_{i,r}$ 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 $a_{i,r}$ (a total of (i^d) elements) in stage $SE_{i,r}$ such that

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

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

$$Y_{i,r}^{4,2} = \left\{ a_{i,r,1} a_{i,r,2} (1 - a_{i,r,3}) (1 - a_{i,r,4}) + a_{i,r,1} (1 - a_{i,r,2}) a_{i,r,3} (1 - a_{i,r,4}) + a_{i,r,1} (1 - a_{i,r,2}) (1 - a_{i,r,3}) a_{i,r,4} + (1 - a_{i,r,1}) a_{i,r,2} a_{i,r,3} (1 - a_{i,r,4}) + (1 - a_{i,r,1}) a_{i,r,2} (1 - a_{i,r,3}) a_{i,r,4} + (1 - a_{i,r,1}) (1 - a_{i,r,2}) a_{i,r,3} a_{i,r,4} \right\} \quad (12)$$

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

2.3.1 $s2 < B$: In this case, every packet which would have wanted to enter $SE_{i,r}$ has actually entered the SE and none is blocked owing 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.

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

2.3.2 $s2 = B$: In this case, it is possible that a packet which was offered to an inlet of $SE_{i,r}$ was not accepted owing to there being fewer 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, whose total cells 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.

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

Subscript ζ in eqn. 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 SE to which $a_{i,r,j}$ is connected. As in eqn. 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 the 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.

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

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

Delay of a packet leaving an output 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 Model validation

The model presented in Section 2 is validated by 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:

- a source tag which denotes the input link at which the packet arrived
- a destination tag denoting the output link to which the packet is destined
- 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 the 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:

- 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.

- 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.

- A new packet is generated at every input of stage 1 with probability ρ , taking into 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 arbitrary output traffic distribution. For any distribution, the only thing that needs to be changed is the load distribution matrix \mathbf{L} discussed in Section 2.2. In this Section we examine the model under uniform, hot-spot and favourite distributions, through analytical and simulation results.

The normalised throughput of a delta network for $N = 64$ and $B/d = 2$ is illustrated in Figs. 3 and 4, for hot-spot and favourite distributions, respectively. The proposed model is quite accurate when the input load is less than 0.7.

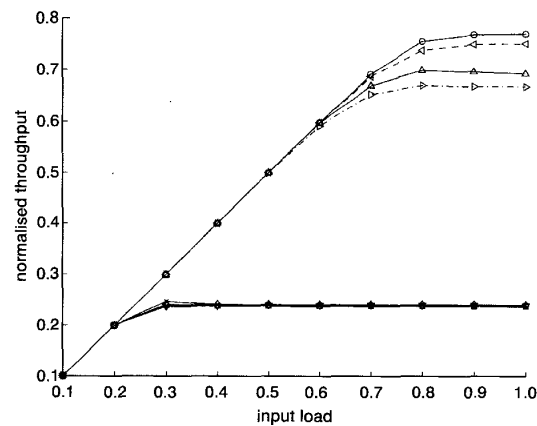


Fig. 3 Throughput against input load for $N = 64$ and $B/d = 2$ (hot-spot)
—○— model, $d = 2, h = 0$ - -◇- - sim, $d = 2, h = 0.05$
—◇— model, $d = 2, h = 0.05$ - -□- - sim, $d = 2, h = 0.05$
- -△- - model, $d = 4, h = 0$ - -△- - sim, $d = 4, h = 0.05$
- -×- - model, $d = 4, h = 0.05$ - -▽- - sim, $d = 4, h = 0.05$

The model gives accurate results for higher input rates under both hot-spot and favourite distributions as the hot or favourite ratios increase.

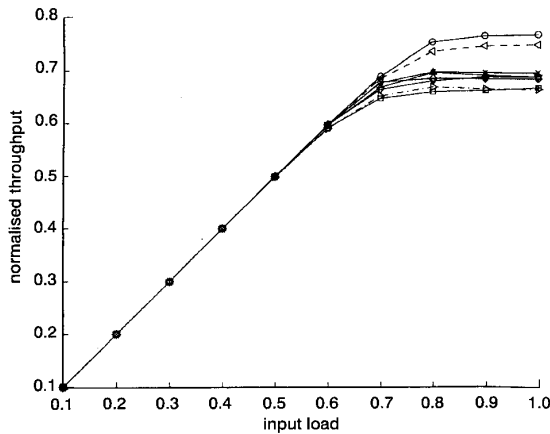


Fig. 4 Throughput against input load for $N = 64$ and $B/d = 2$ (favourite)

- model, $d = 2, h = 0$
- ◇ model, $d = 2, h = 0.05$
- △ model, $d = 4, h = 0$
- × model, $d = 4, h = 0.05$
- sim, $d = 2, h = 0.0$
- ◇ sim, $d = 2, h = 0.05$
- △ sim, $d = 4, h = 0.0$
- ▽ sim, $d = 4, h = 0.05$

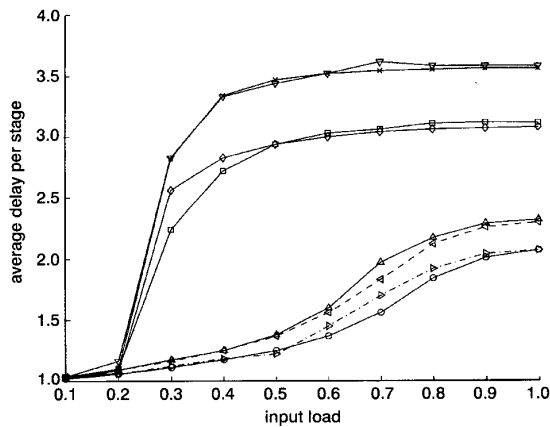


Fig. 5 Average delay per stage against input load for $N = 64$ and $B/d = 2$ (hot-spot)

- model, $d = 2, h = 0$
- ◇ model, $d = 2, h = 0.05$
- △ model, $d = 4, h = 0$
- × model, $d = 4, h = 0.05$
- sim, $d = 2, h = 0.0$
- ◇ sim, $d = 2, h = 0.05$
- △ sim, $d = 4, h = 0.0$
- ▽ sim, $d = 4, h = 0.05$

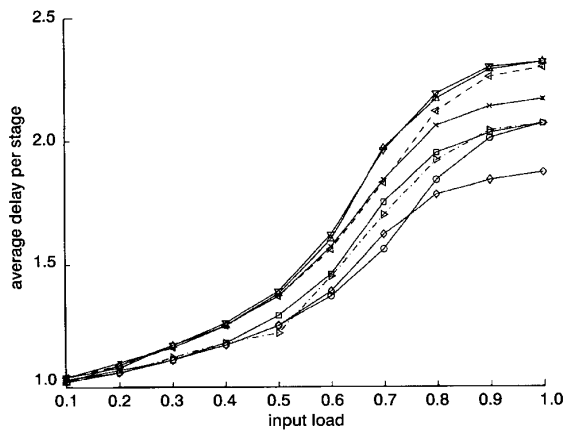


Fig. 6 Average delay per stage against input load for $N = 64$ and $B/d = 2$ (favourite)

- model, $d = 2, h = 0$
- ◇ model, $d = 2, h = 0.05$
- △ model, $d = 4, h = 0$
- × model, $d = 4, h = 0.05$
- sim, $d = 2, h = 0.0$
- ◇ sim, $d = 2, h = 0.05$
- △ sim, $d = 4, h = 0.0$
- ▽ sim, $d = 4, h = 0.05$

Favourite memory distribution has a lesser impact on the overall throughput of delta network than hot-spot. This is owing 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 the favourite distribution, however, there are as many favourite 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 for the hot-spot and favourite distributions confirms the discussion with respect to the difference in the nature of the two distributions.

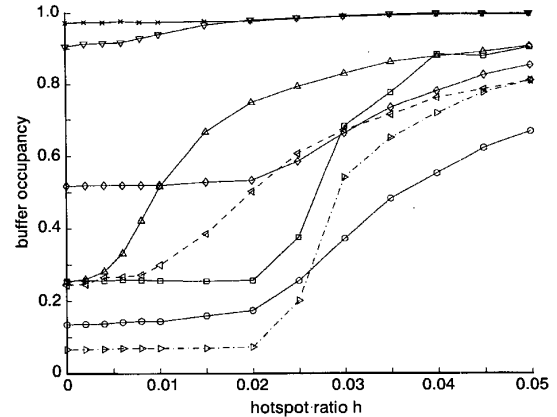


Fig. 7 Buffer occupancy of first type SE at the first stage for $N = 64$, and hot-spot output distribution

- $B/d = 1, p = 0.4, \text{hot occ.}$
- ◇ $B/d = 1, p = 0.4, \text{total occ.}$
- △ $B/d = 1, p = 0.8, \text{hot occ.}$
- × $B/d = 1, p = 0.8, \text{total occ.}$
- $B/d = 2, p = 0.4, \text{hot occ.}$
- ◇ $B/d = 2, p = 0.4, \text{total occ.}$
- △ $B/d = 2, p = 0.8, \text{hot occ.}$
- ▽ $B/d = 2, p = 0.8, \text{total occ.}$

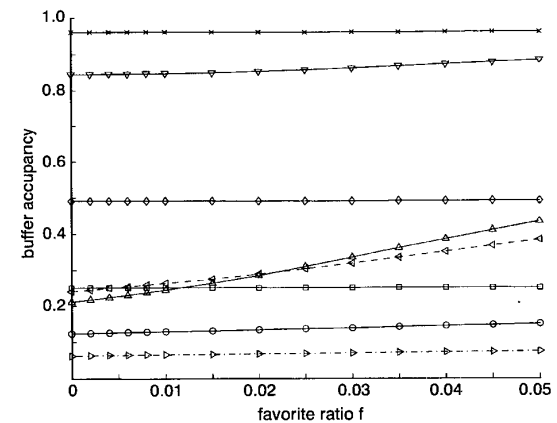


Fig. 8 Buffer occupancy of first type SE at the first stage for $N = 64$, and favourite output distribution

- $B/d = 1, p = 0.4, \text{fav. occ.}$
- ◇ $B/d = 1, p = 0.4, \text{total occ.}$
- △ $B/d = 1, p = 0.8, \text{fav. occ.}$
- × $B/d = 1, p = 0.8, \text{total occ.}$
- $B/d = 2, p = 0.4, \text{fav. occ.}$
- ◇ $B/d = 2, p = 0.4, \text{total occ.}$
- △ $B/d = 2, p = 0.8, \text{fav. occ.}$
- ▽ $B/d = 2, p = 0.8, \text{total occ.}$

Buffer occupancy for hot-spot and favourite distributions is illustrated in Figs. 7 and 8, respectively. For the 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 a 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 the favourite output distribution such as the one shown in Fig. 8, however, although 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 favourite value increases, is linear.

Although the analysis presented in this paper has been based on the delta network, the analysis can be used for other types of ATM switches built from Banyan-type multistage switches.

5 Conclusions

We have developed an analytical model to study the performance of multistage networks constructed from shared buffer switching elements with an arbitrary SE size and buffer size. The proposed model may be used for analysis of an arbitrary traffic pattern. The accuracy of the model has been verified, through various examples, with the results obtained by simulation of the same network for uniform, hot-spot and favourite distributions. Numerical results show close agreement with the results obtained from the simulation.

We have only studied the performance merits of the 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 a small number of inlets and outlets. However, since the number of states in an SE grows exponentially as the SE size increases,

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.

6 References

- 1 PATEL, J.H.: 'Processor-memory interconnections for multiprocessors'. Proceedings of 6th annual symposium on *Computer architecture*, New York, NY, USA, 1979
- 2 SAKURAI, Y., IDO, N., GOHARA, S., and ENDO, N.: 'Large-scale ATM multistage switching network with shared buffer memory switches', *IEEE Commun. Mag.*, 1991, pp. 90-96
- 3 KUWAHARA, H., ENDO, N., OGINO, M., and KOZAKI, T.: 'A shared buffer memory switch for an ATM exchange'. Proceedings of IEEE international conference on *Communications*, 1989, pp. 118-122
- 4 SHOBATAKE, Y., MOTOYAMA, M., SHOBATAKE, E., KAMITAKE, T., SHIMUZU, S., NODA, M., and SAKAUE, K.: 'A one-chip 8*8 ATM switch LSI employing shared buffer architecture', *IEEE J. Sel. Areas Commun.*, 1991, 9, (8), pp. 1248-1253
- 5 FROST, V.S., and MELAMED, B.: 'Traffic modeling for telecommunications networks', *IEEE Commun. Mag.*, 1994, 32, (3), pp. 70-81
- 6 TURNER, J.S.: 'Queueing analysis of buffered switching networks', *IEEE Trans. Commun.*, 1993, 41, (2), pp. 412-420
- 7 MONTEROSSO, A., and PATTAVINA, A.: 'Performance analysis of multistage interconnection networks with shared-buffered switching elements'. Proceedings of INFOCOM'92, 1992, pp. 124-131
- 8 BIANCHI, G., and TURNER, J.: 'Improved queueing analysis of shared buffer switching networks'. Proceedings of INFOCOM'93, 1993, pp. 1392-1399
- 9 ESFAHANI, M.S., and ATIQUZZAMAN, M.: 'Queueing analysis of shared buffer switches for ATM networks'. Proceedings of IEEE GLOBECOM'94, San Francisco, CA, USA, Dec. 1994, pp. 1070-1074
- 10 GIANATTI, S., and PATTAVINA, A.: 'Performance analysis of shared-buffered Banyan networks under arbitrary traffic patterns'. IEEE INFOCOM'93, 1993, pp. 943-952
- 11 PFISTER, G.F., and NORTON, V.A.: 'Hot spot contention and combining in multistage interconnection networks', *IEEE Trans. Comput.*, 1985, C-34, (10), pp. 943-948