A TDM-Based Multibus Packet Switch

Yiu-Wing Leung, *Senior Member, IEEE*, and Tak-Shing Yum, *Senior Member, IEEE*

***Abstract*— A new packet switch architecture using two sets of time-division multiplexed buses is proposed. The horizontal buses collect packets from the input links, while the vertical buses distribute the packets to the output links. The two sets of buses are connected by a set of switching elements which coordinate the connections between the horizontal buses and the vertical buses so that each vertical bus is connected to only one horizontal bus at a time. The switch has the advantages of: 1) adding input and output links without increasing the bus and I/O adaptor speed; 2) being internally unbuffered; 3) having a very simple control circuit; and 4) having 100% throughput under uniform traffic. A combined analytical–simulation method is used to obtain the packet delay and packet loss probability. Numerical results show that for satisfactory performance, the buses need to run about 30% faster than the input line rate. With this speedup, even at a utilization factor of 0.9, each input adaptor requires only 31 buffers for a packet loss rate of 10^{-6}. The output queue behaves essentially as an *M/D/*1 queue.**

***Index Terms*—Packet switches, queueing analysis.**

I. Introduction

THE development of communication networks has reached a point that the switching system rather than the transmission system becomes the bottleneck for the growing volume and varieties of traffic. In Hong Kong, as an example, a large quantity of dark fibers have been laid, but good quality video and image communication is still a rarity because currently available switching facilities cannot accommodate them economically. Many fast packet switches have been proposed in recent years, and they can be classified into three broad types: shared-memory based [1]–[4], shared-medium based [1], [2], [5]–[12], and space-division based [1], [2], [13]–[18].

The shared-medium based switch has among its advocates IBM's PARIS switch designers and NEC's ATOM switch designers. The PARIS switch [5]–[7], [11] is designed for private networks. With the use of automatic network routing, the architecture of the switch can be kept very simple. Variable-size packets can be accommodated, and a very efficient round-robin exhaustive bus-access policy is adopted. On such a single broadcasting medium, multicasting and broadcasting functions can easily be implemented. The ATOM switch [8], [12] uses the bit-slice organization to alleviate the limitation of the bus speed. For still large switches, a multistage organization was proposed. Store and forward of packets, however, is needed at every stage. An alternative to the multistage organization is to use multiple shared media. Nojima *et al*. [9] have developed a switch in which several shared buses are connected in matrix form with memory located at each crosspoint of the buses. Packets contending for access to the same bus are stored in the crosspoint memories connected to this bus. Arbiters scan the crosspoint memories and remove packets from them.

In this paper, we study a new switch architecture using multiple shared buses. This switch has the following advantages: 1) adding input and output links without increasing the bus and I/O adaptor speed, 2) they are internally unbuffered, 3) they have a very simple control circuit, and 4) they have 100% throughput under uniform traffic. We derive the expected delay and the packet loss probability under various bus transfer rate for this switching system.

Paper approved by G. P. O'Reilly, the Editor for Communications Switching of the IEEE Communications Society. Manuscript received July 7, 1993; revised November 1, 1994 and October 13, 1995. This paper was presented in part at IEEE INFOCOM'92, Florence, Italy, 1992.

Y.-W. Leung is with the Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong.

T.-S. Yum is with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: yum@ie.cuhk.hk).

Publisher Item Identifier S 0090-6778(97)05178-7.

II. The TDM-Based Multibus Packet Switch

The multibus packet switch is designed for switching fixed size packets. The packet size can be set to 53 bytes for ATM switching.

A. Architecture

Fig. 1 shows the architecture of an $N \times N$ multibus packet switch. Packets enter the switch through the input links. Each input link is operated synchronously, with time being divided into *link slots*, where each link slot can accommodate one packet. Each input link and each output link are connected to the switch through an input adaptor and an output adaptor, respectively. Fig. 2 shows the internal structure of an input and an output adaptor. The input adaptor receives packets from the input link, performs a serial-to-parallel conversion, and queues the packets in a set of buffers. The output adaptor performs two functions. First, it filters out all packets destined for this particular adaptor and puts them in the output buffer. Second, it performs a parallel-to-serial conversion for the packets for onward transmission.

The N input links (output links) are partitioned in M input groups (output groups) of L links each where $N = ML$. Group i input adaptors are connected to *horizontal bus* HB_i and group i output adaptors are connected to *vertical bus* VB_i. In other words, L input links are sharing a horizontal bus, and L output links are sharing a vertical bus. The group size L here is a design parameter. If we want a smaller packet delay, each horizontal bus should serve a smaller group of input adaptors, or the group size L should be smaller. On the other hand, a larger group size L means a smaller number of groups M (for a fixed N). This means a smaller number of horizontal and vertical buses, and hence a smaller switch complexity. The bus width is another design parameter. A larger bus width gives a

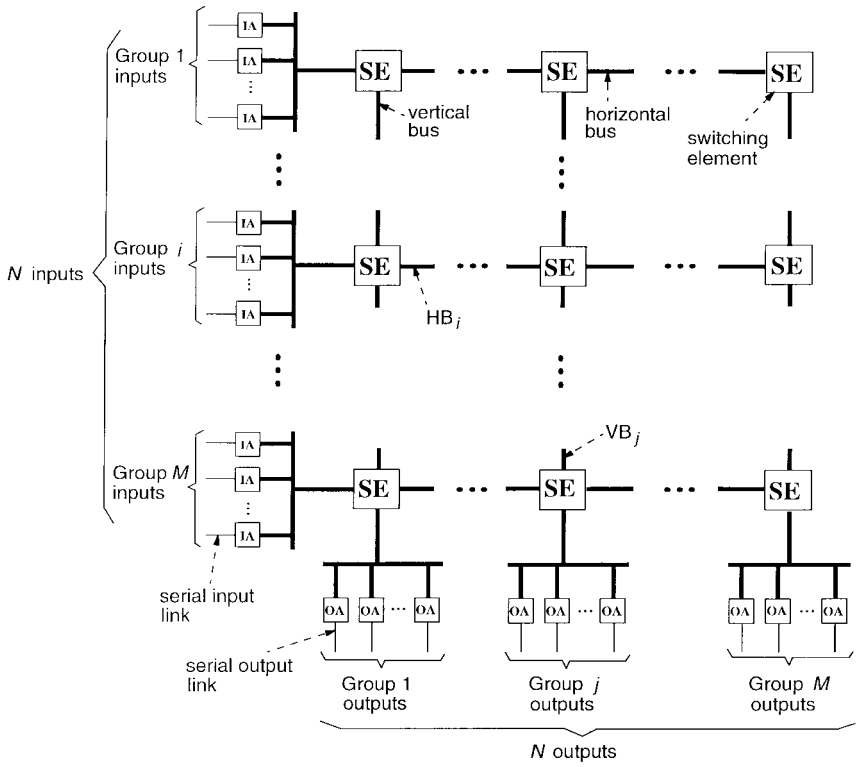

Fig. 1. Multibus packet switch.

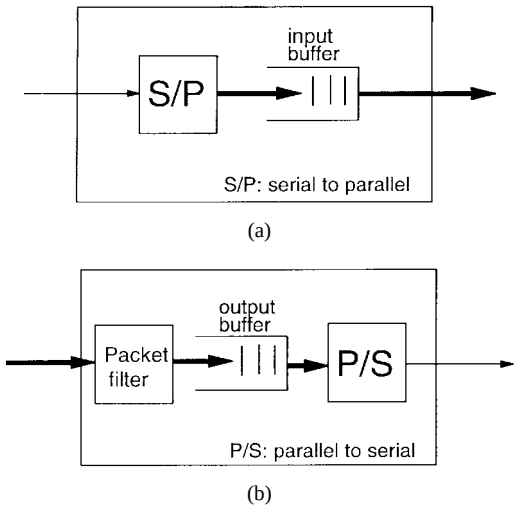

Fig. 2. Input and output adaptors.

higher data transfer rate, and hence, a higher switch throughput at the expense of a higher implementation cost. Based on the current technology, a bus of width 64 bits operating at 100 MHz can provide a bus transfer rate of 6.4 Gbits/s [7].

The M horizontal buses are connected to the M vertical buses in a bus matrix, with a total of M^2 *switching elements* at the crosspoints of the vertical and horizontal buses. The switching element placed at the crosspoint of HB_i and VB_j is identified as $\mathrm{SE}_{i,j}$. Fig. 3(a) shows the schematic of a switching element. It connects the horizontal input bus to either the horizontal output bus or the vertical bus. Fig. 3(b) shows the circuit realization of the switching element, using $2b$ *relays* (b is the bus width), b inverters, and one shift register. The relay is a three-terminal element with one input, one output, and one control line. It connects the input line to the output line whenever there is a "1" on the control line. For prototyping, the set of relays are available as off-the-shelf IC chips (e.g., Motorola's SN54LS). For actual implementation, ASIC chips with multiple switching elements per chip can be used. Since the circuitry in each switching element is very simple, the number of switching elements per chip depends only on the number of available pins per chip. For example, if the bus size is 32 bits and a chip consists of four switching elements, the chip must have 256 pins for inputs/outputs. The shift register in $\mathrm{SE}_{i,j}$ stores a bit pattern which determines when to connect the horizontal input bus to the vertical bus. When a clock pulse arrives, the last bit is shifted out to the relays. If this bit is "1," the horizontal input bus is connected to the horizontal output bus; otherwise, the horizontal input bus is connected to the vertical bus. The connection patterns of the switching elements are chosen such that one vertical bus is connected to only one horizontal bus at a time. Note that the clock rate is equal to the packet rate on the bus (e.g., if the bus is operated at 6 Gbits/s and the packet size is 53 bytes, the clock rate is 14.2 MHz).

B. Operation

The transmission of packets on a bus is divided into *cycles* of equal duration. Each cycle is subdivided into M *subcycles* of equal duration (Fig. 4). In the ith subcycle, group j input

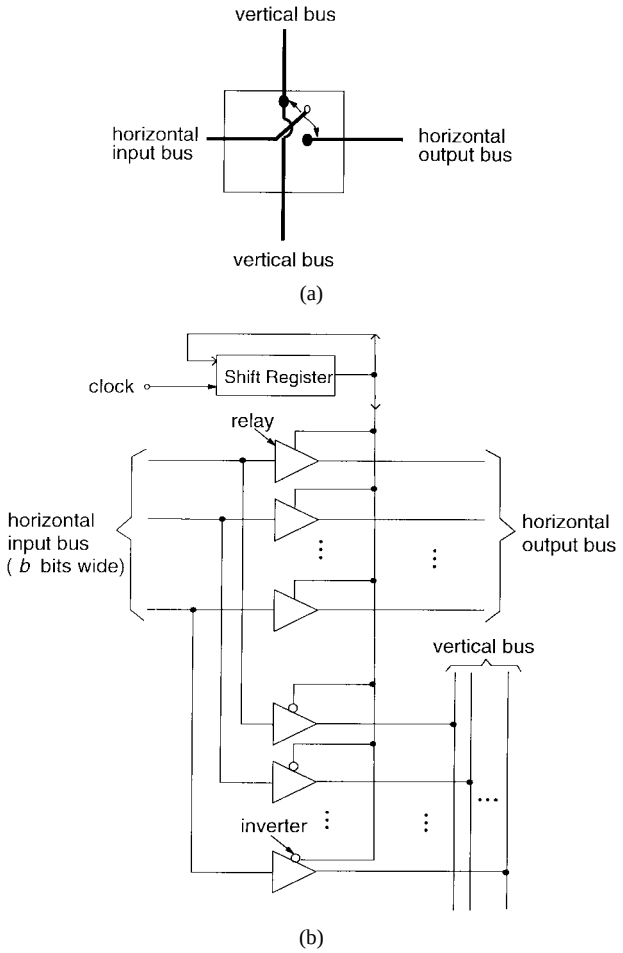

Fig. 3. Switching element. (a) Schematic of the switching element and (b) circuit realization of the switching element.

adaptors are connected to vertical bus $\mathrm{VB}_{f(i,j)}$ where[1]

$$f(i,j)=\begin{cases}(i+j-1) \bmod M & (i+j-1) \bmod M \neq 0\\ M & (i+j-1) \bmod M = 0.\end{cases} \tag{1}$$

Thus, in the ith subcycle, packets from group j input adaptors are switched to group $f(i,j)$ output adaptors. Hence, only the switching elements $\mathrm{SE}_{j,f(i,j)}$ $(j=1,2,\cdots,M)$ connect the horizontal buses HB_j to the vertical buses $\mathrm{VB}_{f(i,j)}$ $(j=1,2,\cdots,M)$, while all of the other switching elements connect the horizontal input buses to the horizontal output buses. Fig. 4 shows an example of this transmission arrangement when $M=3$ and $L=4$. This transmission arrangement ensures that in each subcycle, there is a unique one-to-one connection from every group of input adaptors to every group of output adaptors. This means that the M groups of input adaptors can simultaneously transmit packets to the M groups of output adaptors through the bus matrix.

To resolve the bus contention among the L input adaptors in each group, each subcycle is further divided into L bus slots, where each bus slot can accommodate one packet and is dedicated to one input adaptor. Each adaptor can, therefore, transmit one packet in each subcycle. Note that when the bus transfer rate is fixed, a larger number of inputs N increases the cycle duration.

[1]Note that if we would have labeled the vertical buses as $0, 1, 2, \cdots, M-1$ instead of $1, 2, \cdots, M$, (1) will be simplified to $f(i,j)=(i+j) \bmod M$. But doing so would complicate the subsequent discussion.

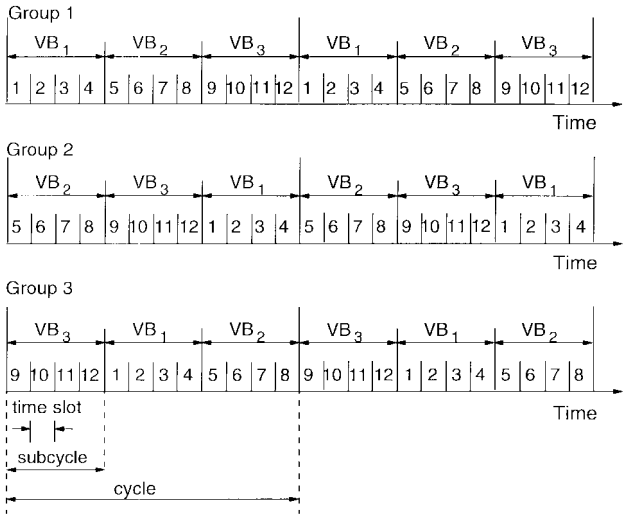

Fig. 4. Transmission cycles, subcycles, and time slots; $M=3$, $L=4$, and $N=12$.

Global timing is used to ensure that all transmissions are properly synchronized. This requires that all of the input adaptors and switching elements are triggered by a common clock.

C. Speedup Factor

We define the *speedup factor* SF of the switch as the ratio of the sum of the data rates of all of the vertical buses to the sum of the data rates of all of the input links. Since there are M vertical buses and N input links, SF can be written as

$$\mathrm{SF}=\frac{M \times \text{data rate of a bus}}{N \times \text{data rate of an input link}}.$$

In the next section, we will analyze the performance of the switch with SF as a parameter. Note that when SF is made larger, the buses can serve the input adaptors at a higher rate and yields a smaller input queueing delay. When $\mathrm{SF}=M$, input queueing is not required, but the implementation cost is high. Our results will indicate that a small SF (say, $\mathrm{SF}=1.3$) can already give satisfactory performance in the sense of very little queueing at the input adaptor. Note also that the cycle length and the link-slot size are given by $N\tau$ and $N\tau(\mathrm{SF}/M)$, respectively, where τ is the duration of a bus slot.

III. Performance Analysis

A. Queueing Model and Decoupling of Queues

In each link slot, there is a packet arrival with probability ρ. Let α_{ij} be the probability that an incoming packet from input link i is destined for output link j. Then, the probability β_{ik} of packet arrival from input link i to group k output links is

$$\beta_{ik}=\rho \sum_{j=(k-1)L+1}^{kL} \alpha_{ij}. \tag{2}$$

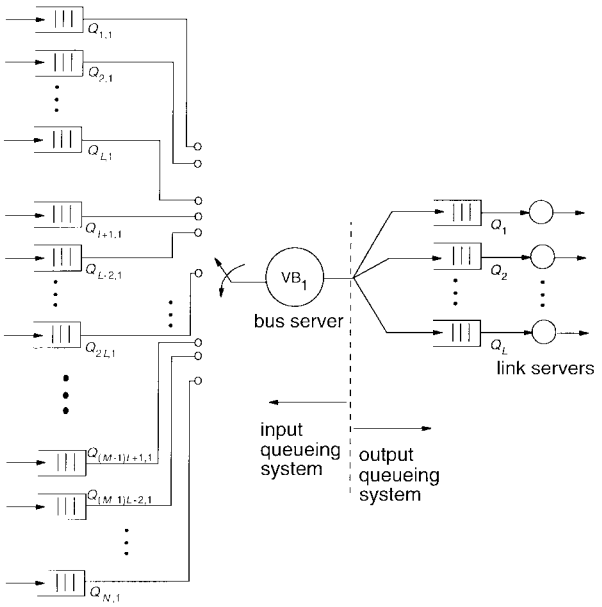

Fig. 5. Queueing model for vertical bus VB_1.

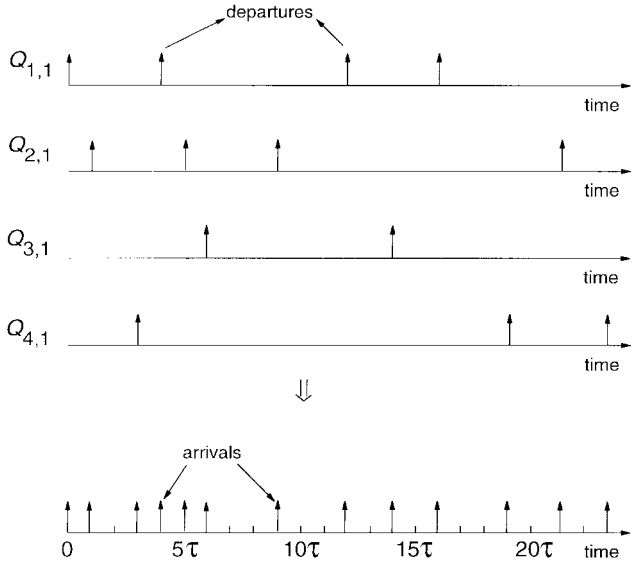

Fig. 6. Arrival process to output queueing system is the superposition of the departure processes ($M = 2$, $L = 2$).

The packets in an input adaptor are logically organized into M queues such that queue j contains all packets destined for group j links. For convenience, we let $Q_{i,j}$ denotes queue j in input adaptor i. Let B_1 and B_2 be the buffer size in each input and output adaptor, respectively. The queues $Q_{i,j}$, $j = 1, 2, \cdots, M$ share these buffers by the complete sharing strategy [19]. Packets queued at the output adaptor are transmitted in a first-in, first-out order.

Without loss of generality, let us consider the delay of the packets departing from the group 1 output links. As group 1 output adaptors only get packets from VB_1, we shall model VB_1 as a *bus server*. For convenience, we shall call the subsystem up to and after the bus server the *input queueing system* and *output queueing system*, respectively. As seen from Fig. 5, there are altogether N input queues $Q_{1,1}, Q_{2,1}, \cdots, Q_{N,1}$ feeding packets to the bus server. The output queueing system consists of L queues corresponding to the L output adaptors in group 1.

The switching elements connecting the horizontal buses and the vertical buses are operated in such a way that each input queue has a fixed dedicated bus slot for transmitting a packet in every cycle. All input queues are therefore independent. Recall that the duration of each bus slot and each cycle are τ and $N\tau$, respectively (see Section II-C). Therefore, all queues are served once every $N\tau$ seconds with service time τ. Analysis of the input queues is given in the next subsection.

The arrival process to the output queueing system is the superposition of the departure processes of all of the input queues in the input queueing system. To characterize this arrival process, we must first characterize the departure process of each of the input queues. The bus server visits an input queue every $N\tau$ seconds, and removes one packet from the queue when the queue is not empty. As far as the characterization of the departure process is concerned, the service time in the input queue can be considered as equal to $N\tau$ seconds. For input queue $Q_{1,1}$, packet departure occurs at time epochs that are integer multiples of $N\tau$. Therefore, the durations of busy and idle periods are both integer multiples of $N\tau$. The input queues $Q_{2,1}, Q_{3,1}, \cdots$ are served by the bus server in a similar manner as for $Q_{1,1}$, except for a time lag of $\tau, 2\tau, \cdots$ seconds, respectively. The probability mass functions of the idle period and busy period durations for input queue $Q_{m,n}$ are derived in Section III-D. In general, the departure epoch for the input queue $Q_{i,1}$ occurs at $(kN + i - 1)\tau$ for $k = 1, 2, \cdots$. The departure process of each of the input queues is characterized by these time epochs and the distributions of the idle and busy periods. The arrival process to the output queueing system is the superposition of the departure processes from all of the input queues. Fig. 6 shows an example of the departure process from the input queueing system with $M = L = 2$.

Departing packets from $Q_{i,1}$ join output queue j with probability α_{ij}. So the arrival to output queue j due to $Q_{i,1}$ is just the α_{ij} bifurcation of the departure process from $Q_{i,1}$. The composite arrival process to output queue j is the superposition of the N bifurcated departure processes from $Q_{i,1}$, $i = 1, 2, \cdots, N$ to output queue j. With such a complicated arrival process, we have to resort to computer simulation to obtain the queueing delay. As only a single server queue needs to be simulated, very accurate delay and packet loss results can be obtained.

B. Expected Delay in the Input Queue

We assume that the packet loss probability at the input queue with a finite buffer size B_1 is very small, and we approximate the expected delay with a finite buffer by the delay with an infinite buffer. As all queues are similar, we choose to analyze a particular one with packet arrival probability β. We use the standard imbedded Markov chain analysis. The input queue is served once every cycle for τ seconds. The imbedded points are chosen at the time instances at which the bus server has just visited the input queue. We let there be K link slots in the time interval $[0, N\tau]$, and let q_i be the number of customers in the queue at the ith imbedded point. Then q_{i+1} is related

to q_i by

$$q_{i+1} = \begin{cases} q_i + a - 1, & q_i > 0 \\ a - 1, & a > 0,\ q_i = 0 \\ 0, & a = 0,\ q_i = 0 \end{cases} \quad (3)$$

where a is a random variable denoting the number of arrivals in K link slots (i.e., in one cycle). Taking the z transform, we have

$$\begin{aligned} E[z^{q_{i+1}}] &= E[z^{q_{i+1}}|q_i > 0]\,\text{Prob}\,[q_i > 0] \\ &\quad + E[z^{q_{i+1}}|q_i = 0]\,\text{Prob}\,[q_i = 0] \\ &= E[z^{q_i+a-1}|q_i > 0]\,\text{Prob}\,[q_i > 0] \\ &\quad + \{E[z^0|a = 0]\,\text{Prob}\,[a = 0] \\ &\quad + E[z^{a-1}|a > 0]\,\text{Prob}\,[a > 0]\}\,\text{Prob}\,[q_i = 0]. \quad (4) \end{aligned}$$

In steady state, $E[z^{q_{i+1}}] = E[z^{q_i}] = E[z^q]$. The generating function $G_Q(z)$ of the number of customers in the input queue can be obtained from (4) as

$$G_Q(z) = E[z^q] = \frac{(1 - K\beta)(1 - z)}{(1 - \beta + \beta z)^K - z}. \quad (5)$$

The expected number of customers $E[q]$ at the imbedded points is given by

$$E[q] = \left.\frac{dG_Q(z)}{dz}\right|_{z=1} = \frac{K(K-1)\beta^2}{2(1 - K\beta)}. \quad (6)$$

Consider the arrival of a tagged packet. Since the arrival of packets is a Bernoulli process, each link slot is equally likely to contain an arrival. Then, the arrival time of the tagged packet is equally probable in any of the K link slots. Let A_i be the event that the tagged packet arrives in link slot i. Then the expected number of packets $E[L]$ arriving from the last imbedded point until the arrival of the tagged packet is

$$\begin{aligned} E[L] &= \sum_{i=1}^{K} E[L|A_i]\,\text{Prob}\,[A_i] \\ &= \frac{1}{K}\sum_{i=1}^{K}(i-1)\beta \\ &= \frac{(K-1)\beta}{2}. \end{aligned} \quad (7)$$

The number of packets in the queue averaged over a cycle, denoted as $E[m]$, is

$$E[m] = E[q] + \frac{K\beta}{2}. \quad (8)$$

The probability of packet arrival to an input adaptor is $K\beta/N\tau$. Therefore, by Little's formula, the expected delay D is

$$D = \frac{E[m]}{\lambda} = \frac{(1-\beta)N\tau}{2(1 - K\beta)}. \quad (9)$$

Consider the following numerical example. Let there be $N = 50$ inputs, and they are divided into $M = 5$ groups. Let the link utilization be 0.9. Then, $\beta = 0.9/M = 0.18$. If the packet size is 53 bytes and the bus is operated at 1.25 Gbits/s, then $\tau = 0.34$ μs. If the input link rate is 100 Mbits/s, then $K = 4$, and hence, the delay D from (9) is 25 μs.

C. Packet Loss Probability at the Input Adaptor

The M logical queues in each input adaptor share the B_1 buffers by the complete sharing strategy which has the best blocking performance [19]. When all of the buffers are occupied, incoming packets are lost. The buffer size B_1 must be chosen such that the probability of packet loss P_L is very small. In this section, we derive an approximate expression of P_L as a function of the buffer size B_1. This expression is an upper bound of P_L. First, we derive the probability mass function of the number of packets in the input queue with infinite buffers. Let π_i be the probability that there are i packets in an input queue at the imbedded points. It is given by the coefficient of z^i in the power series expansion of $G_Q(z)$.

Consider a tagged packet that arrives at the jth link slot after an imbedded point. The probability $\Lambda(i, j)$ that this tagged packet sees i packets in queue is given by

$$\begin{aligned} \Lambda(i, j) &= \sum_{p=0}^{\min\{i, j-1\}} \text{Prob}[\,p \text{ packet arrivals in the first} \\ &\qquad j-1 \text{ link slots}] \cdot \text{Prob}[\text{there are } i-p \text{ packets} \\ &\qquad \text{in the queue at the last imbedded point}] \\ &= \sum_{p=0}^{\min\{i, j-1\}} \left[\binom{j-1}{p}\beta^p(1-\beta)^{j-1-p}\right]\pi_{i-p}. \end{aligned} \quad (10)$$

The probability Π_i that there are i packets in an input queue with infinite buffer is given by

$$\begin{aligned} \Pi_i &= \sum_{j=1}^{K} \text{Prob}[\text{the tagged packet sees } i \text{ packets} \\ &\qquad \text{in the input queue}|A_j] \cdot \text{Prob}\,[A_j] \\ &= \frac{1}{K}\sum_{j=1}^{K}\Lambda(i, j). \end{aligned} \quad (11)$$

When the buffer size B_1 is finite and the complete sharing strategy is employed, the M input queues in an input adaptor are not independent. However, for a well-designed fast packet switch, the buffer size B_1 can be chosen such that the probability of packet loss is very small (say, less than 10^{-6}). In this case, the input queues can be regarded as independent. The probability of packet loss P_L in an input adaptor is given as

$$P_L = 1 - \sum_{\sum_{j=1}^{M} i_j \le B_1} \Pi_1(i_1)\Pi_2(i_2)\cdots\Pi_M(i_M). \quad (12)$$

D. Probability Mass Functions of the Idle and Busy Periods

In this section, we derive the probability mass functions of the length of the idle and busy periods from an input queue. These functions characterize the departure processes from the input queues, and are used to generate the arrival process to the output queue in the simulation experiments. Let X and Y be the duration of the idle and busy periods in unit of cycles. Then, $x_i = \text{Prob}[X = iN\tau]$ is given by

$$\begin{aligned} x_i &= \text{Prob}[\text{no packet arrival for consecutive } i \text{ cycles} \\ &\qquad \text{and a packet arrives at cycle } i+1 | i \ge 1] \\ &= \frac{[(1-\beta)^K]^i[1 - (1-\beta)^K]}{(1-\beta)^K}. \end{aligned} \quad (13)$$

Let $G_Y(z)$ be the probability generating function of Y. $G_Y(z)$ can be found using the method in [20] as

$$G_Y(z) = \left[\frac{z^{-1}G_R(z) - (1-\beta)^K}{1-(1-\beta)^K}\right] \tag{14}$$

where

$$G_R(z) = \sum_{i=1}^{\infty} r_i z^i = z[1-\beta+\beta G_R(z)]^K. \tag{15}$$

From (14) and (15), $y_i = \text{Prob}\,[Y = iN\tau]$ is obtained as

$$\begin{aligned} y_i &= \frac{r_{i+1}}{1-(1-\beta)^K} \\ &= \frac{\dfrac{1}{(i+1)!}\dfrac{d^{i+1}G_R(z)}{dz^{i+1}}\Big|_{z=0}}{1-(1-\beta)^K} \end{aligned} \tag{16}$$

where

$$\begin{aligned} \frac{d^n G_R(z)}{dz^n}\Big|_{z=0} = &\sum_{\substack{1+l_2+l_3+\cdots+l_{K+1}=n \\ 0\le l_2, l_3, \cdots, l_{K+1}\le n-1}} \binom{n}{1, l_2, l_3, \cdots, l_{K+1}} \\ &\cdot \prod_{i=2}^{K+1} \frac{d^{l_i}}{dz^{l_i}}[1-\beta+\beta G_R(z)]\Big|_{z=0} \end{aligned} \tag{17}$$

which can be evaluated recursively.

IV. Numerical Results and Discussion

Consider a 1024 $\times$ 1024 switch ($N = 1024$) with inputs divided into eight ($M = 8$) equal groups. There are 128 links per group, and a total of $8 \times 8 = 64$ switching elements. (If a single chip can contain four switching elements, then the switch fabrics requires only 16 chips.) Let the packet transmission time in any input link be normalized to one time unit.

Fig. 7 shows the average queueing delay in the input adaptor. As SF increases, the queueing delay becomes smaller because the input queues are served at a faster rate. At 30% speedup, very small delay is obtained even at $\rho = 0.9$. Fig. 8 shows the packet loss probability at the input adaptor for various buffer sizes at $\rho = 0.9$. When SF $= 1$, the required buffer size to achieve a packet loss probability of 10^{-6} is found to be about 150. However, with SF $= 1.3$, the required buffer size is reduced to only 31. In the input queue, only packets to a certain destination can be served at a certain time. As all packet destinations are assumed to be independent and uniformly distributed, this extra "randomness" makes the speeding up of the bus rate necessary for satisfactory performance.

Fig. 9 shows the average queueing delay in the output adaptor. Here, on the contrary, a larger SF gives a larger delay at the output adaptor. The difference, however, is only apparent at ρ very large (a difference of one time unit at $\rho = 0.9$). Moreover, all SF ≥ 1.3 cases give almost identical delay characteristics. This phenomenon can be explained as follow. When SF $\gg$1, there is essentially no queueing at the input adaptor. All packets to a certain output link will immediately appear at the output queue. The input process is, therefore, a superposition of N Bernoulli processes. For $N = 1024$, that process should be indistinguishable from a Poisson process. Thus, the output queue is just a simple *M/D*/1 queue. In fact,

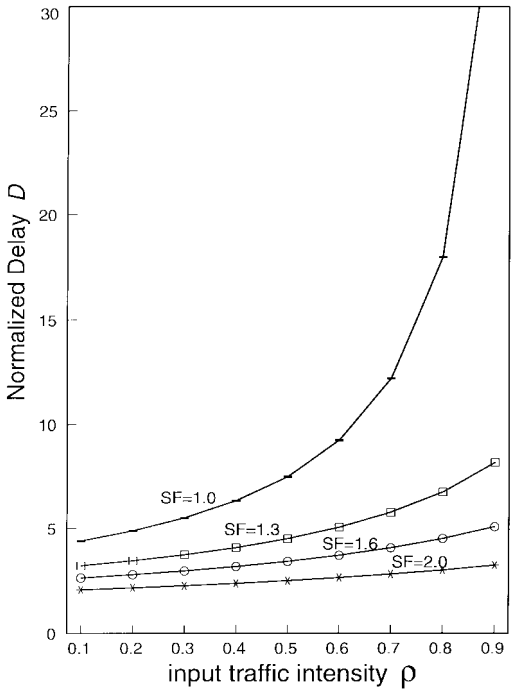

Fig. 7. Input queueing delay versus $\rho(B_1 = \infty)$.

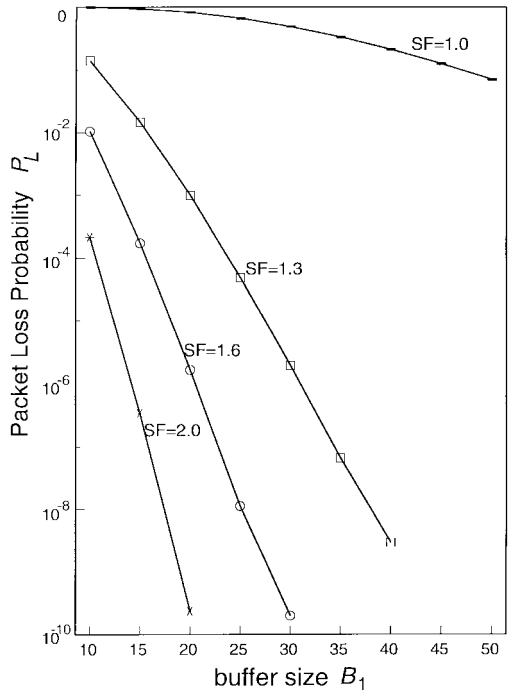

Fig. 8. Packet loss probability versus buffer size in input adaptor ($\rho = 0.9$).

the *M/D*/1 delay characteristics coincide with the SF $= 8$ curve in Fig. 9. What is interesting to note is that for SF $= 1.0$, the delay at the output queue is smaller than that of the *M/D*/1 queue, and for SF $= 1.3$, the delay is essentially that of the *M/D*/1 queue. Fig. 10 shows the packet loss probability versus the buffer size B_2 in the output adaptor when SF $= 1.3$. As can be seen, at $\rho = 0.9$, a packet loss probability of 10^{-4} can be achieved with a buffer size of 30, and 10^{-5} can be achieved with a buffer size of 43.

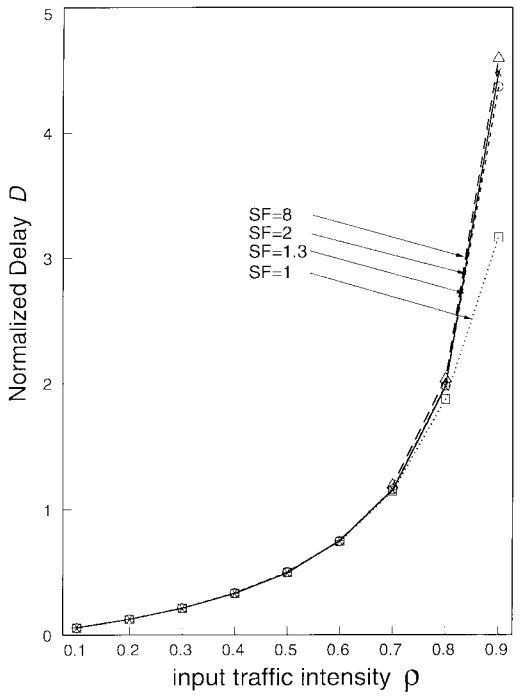

Fig. 9. Output queueing delay versus $\rho(b_2 = 40)$.

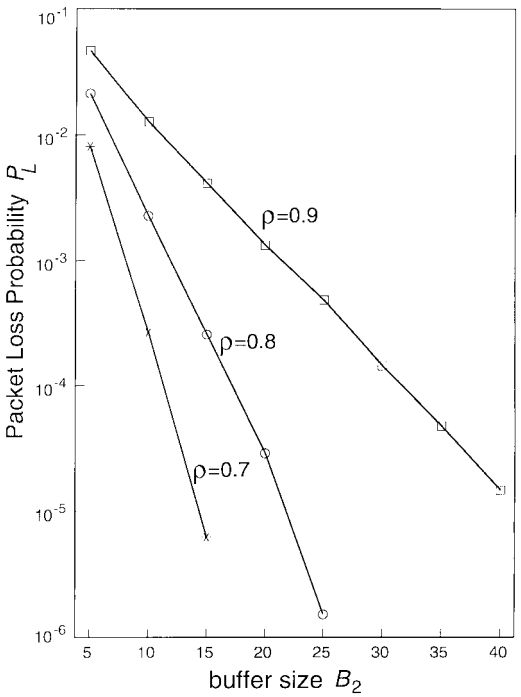

Fig. 10. Packet loss probability in output adaptor versus buffer size (SF = 1.3).

V. Conclusions

There are various approaches to the design of fast packet switch. Using a high-speed bus, packet switching can be done very simply by individual stations through the filtering of unwanted packets. The bus width and bus speed, however, limit the total throughput of the switch. To bypass this limitation, we have designed a TDM-based multibus switch. In addition to modular growth, it preserved the advantages of an internally unbuffered, simple control circuit and 100% throughput under uniform traffic. We have analyzed the performance of the switch in terms of the speedup factor, and have found that for satisfactory performance, the buses need to speed up 30% relative to the individual input line rate.

Since the bus bandwidth is allocated to the input adaptors in a fixed cyclic order, the performance of the switch might not be satisfactory under highly asymmetric traffic conditions. We are currently investigating dynamic bandwidth allocation policies for use in such conditions.

Acknowledgment

Dr. S. C. Liew and an anonymous reviewer made a number of suggestions that resulted in a significant improvement of the technical content.

References

[1] F. A. Tobagi, "Fast packet switch architectures for broadband integrated services digital networks," *Proc. IEEE*, vol. 78, pp. 133–167, Jan. 1990.
[2] H. Ahmadi and W. E. Denzel, "A survey of modern high-performance switching techniques," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1091–1103, Sept. 1989.
[3] J. Garcia-Haro and A. Jajszczyk, "ATM shared-memory switching architectures," *IEEE Network*, pp. 18–26, July/Aug. 1994.
[4] N. Endo, T. Kozaki, T. Ohuchi, H. Kuwahara, and S. Gohara, "Shared buffer memory switch for an ATM exchange," *IEEE Trans. Commun.* vol. 41, pp. 237–245, Jan. 1993.
[5] I. Cidon and I. Gopal, "PARIS: An approach to integrated high speed private networks," *Int. J. Digital Analog Cabled Syst.*, vol. 1, pp. 77–85, 1988.
[6] I. Cidon, I. Gopal, G. Grover, and M. Sidi, "Real time packet switching: A performance analysis," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 1576–1586, Dec. 1988.
[7] I. Gopal, R. Guerin, J. Janniello, and V. Theoharakis, "ATM support in a transparent network," *IEEE Network*, pp. 62–68, Nov. 1992.
[8] H. Suzuki, H. Nagano, T. Suzuki, T. Takeuchi, and S. Iwasaki, "Output buffer switch architecture for asynchronous transfer mode," in *Proc. IEEE ICC'89*, pp. 99–103.
[9] S. Nojima, E. Tsutsui, H. Fukuda, and M. Hashimoto, "Integrated services packet network using bus matrix switch," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1284–1292, Oct. 1987.
[10] C. Fayet, A. Jacques, and G. Pujolle, "High speed switching for ATM: The BSS," *Comput. Networks ISDN Syst.*, vol. 26, pp. 1225–1234, 1994.
[11] I. Cidon, I. Gopal, M. A. Kaplan, and S. Kutten, "A distributed control architecture of high speed networks," *IEEE Trans. Commun.*, vol. 43, pp. 1950–1960, May 1995.
[12] R. Fan, H. Suzuki, K. Yamada, and N. Matsuura, "Expandable ATOM switch architecture (XATOM) for ATM LAN's," in *Proc. IEEE ICC'94*, pp. 402–409.
[13] Y. M. Kim and K. Y. Lee, "PR-Banyan: A packet switch with a pseudo-randomizer for nonuniform traffic," *IEEE Trans. Commun.*, vol. 41, pp. 1039–1042, July 1993.
[14] I. Widjaja and A. Leon-Garcia, "The Helical switch: A multipath ATM switch which preserves cell sequence," *IEEE Trans. Commun.*, vol. 42, pp. 2618–2629, Aug. 1994.
[15] D. X. Chen and J. W. Mark, "SCOQ: A fast packet switch with shared concentration and output queueing," *IEEE/ACM Trans. Networking*, vol. 1, pp. 142–151, Feb. 1993.
[16] T. T. Lee and S. C. Liew, "Broadband packet switches based on dilated interconnection networks," *IEEE Trans. Commun.*, vol. 42, pp. 732–744, Feb. 1994.
[17] S. C. Liew and T. T. Lee, "$N \log N$ dual shuffle-exchange network with error-correcting routing," *IEEE Trans. Commun.*, vol. 42, pp. 754–766, Feb. 1994.
[18] A. Pattavina, "Nonblocking architectures for ATM switching," *IEEE Commun. Mag.*, pp. 38–48, Feb. 1993.
[19] F. Kamoun and L. Kleinrock, "Analysis of shared finite storage in a computer network node environment under general traffic conditions," *IEEE Trans. Commun.*, vol. COM-28, pp. 992–1003, July 1980.
[20] T. S. Yum, "Measuring the utilization of a synchronous data link: An application of busy period analysis," *Bell Syst. Tech. J.*, vol. 59, pp. 731–744, May 1980.

Yiu-Wing Leung (M'92–SM'96) received the B.Sc. and Ph.D. degrees from the Chinese University of Hong Kong in 1989 and 1992, respectively.

He is currently an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University. His current research interests include information networks, design and analysis of algorithms, and software technology.

Tak-Shing Yum (S'76–M'78–SM'86) worked at Bell Telephone Laboratories in the U.S. for two and a half years and taught at National Chiao Tung University, Hsinchu, Taiwan, R.O.C., for two years before joining the Chinese University of Hong Kong in 1982. He has published original research on packet switched networks with contributions in routing algorithms, buffer management, deadlock detection algorithms, message resequencing analysis, and multiaccess protocols. In recent years, he branched out to work on the design and analysis of cellular network, lightwave networks, and video distribution networks.