Round-Robin Selection with Adaptable-Size Frame in a Combined Input-Crosspoint Buffered Switch

Roberto Rojas-Cessa and Eiji Oki

***Abstract*— Combined input-crosspoint buffered switches are an alternative to relax arbitration timing and to provide high-performance switching for packet switches with high-speed ports. It has been shown that these switches, with one-cell crosspoint buffer and round-robin arbitration at input and output ports, provide 100% throughput under uniform traffic. However, under admissible traffic patterns with non-uniform distributions, only weight-based selection schemes are reported to provide high throughput. In this paper, we propose a round-robin based arbitration scheme for a combined input-crosspoint buffered packet switch that provides nearly 100% throughput for several admissible traffic patterns, including uniform and unbalanced traffic, using one-cell crosspoint buffers. The presented scheme uses adaptable-size frames, so that the frame size adapts to the traffic pattern.**

***Index Terms*— crosspoint-buffered switch, packet scheduling arbitration, virtual output queue, credit-based flow control, adaptable-size frame.**

I. Introduction

Combined input-crosspoint buffered packet switches are an alternative to input-buffered switches to relax arbitration timing and to provide high-performance switching for packet switches with high-speed ports [1], [2]. These switches use time efficiently as input and output port selections are independent. As an example, a switch with 40-Gbps (OC-768) ports transferring 64-byte cells must perform input (or output) arbitration within 12.8 ns. However, the number of buffers in a crossbar grows in the same order as the number of crosspoints ($O(N^2)$), where N is the number of input/output ports, making implementation costly for a large buffer size or large N. One way to keep the buffer complexity within a feasible state is by using crosspoint buffers that are smaller in size.

It is a common practice to segment incoming variable-size packets at the ingress side of a switch to perform internal switching with fixed-size packets, or cells, and to re-assemble the packets at the egress side before they depart from the switch. We follow that practice in this paper. Moreover, switches can use separate queues at the inputs, one for each output, called virtual output queueing (VOQ), to avoid Head-of-Line (HOL) blocking [3]. Crossbar fabrics are very popular for implementation of packet switches because of their non-blocking capability, simplicity, and market availability.

In general, packet switches are required to provide: (a) low complexity, (b) fast contention resolution, (c) fairness, and, (d) high matching efficiency.

This project has been partially funded by the SBR Program at NJIT

Roberto Rojas-Cessa is with the Department of Computer and Electrical Engineering, New Jersey Institute of Technology. University Heights, Newark NJ 07102 USA. Tel: 973-596-3508, Fax:973-596-5680. Email: rojasces@njit.edu.

Eiji Oki is with NTT Network Innovation Laboratories, 3-9-11 Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan, Tel:+81-422-59-3441, Fax:+81-422-59-6387. Email: oki.eiji@lab.ntt.co.jp.

As actual traffic may present non-uniform distributions, it is necessary to provide arbitration schemes such that switches provide 100% throughput for admissible traffic. Admissible traffic is defined in [4].

Previously, we showed that a switch using one-cell crosspoint buffers in a buffered crossbar with VOQs at the inputs, a simple round-robin arbitration scheme for input and output arbitration, and a credit-based flow control provides 100% throughput under uniform traffic [2]. We also showed that such a switch needs a large crosspoint buffer to provide high throughput under admissible unbalanced traffic.

One way to provide near 100% throughput under unbalanced traffic is by using a weight-based selection scheme, where weights are assigned to input queues proportionally to their length or HOL cell age. [5] showed that weight-based schemes in buffered crossbars can provide high throughput under various traffic patterns. From the two presented schemes, one is based on the selection the VOQ occupancy at inputs and using a round-robin based arbitration scheme at the output, and the other scheme is based on the oldest cell first (OCF) instead the VOQ occupancy. However, weight-based schemes need to perform comparisons among all contending queues, which number can be large. Furthermore, as the queuing structures tend to be flow-based, the number of comparisons is expected to increase. Moreover, weight-based schemes may starve some queues to provide more service to the congested ones [6], presenting unfairness. On the other hand, many round-robin algorithms have been shown to provide fairness and implementation simplicity as no comparisons are needed among queues [6].

It is necessary to provide a feasible arbitration scheme based in round-robin selection for buffered crossbars such that a switch can obtain high throughput under admissible traffic [6], including unbalanced traffic, with a small crosspoint buffer size.

In this paper, we propose an arbitration scheme for buffered crossbars based on round-robin selection that also uses the concept of adaptable-size frame. We show that this selection scheme can achieve nearly 100% throughput under a non-uniform traffic pattern, the unbalanced traffic model, with one-cell crosspoint buffers. We show via simulation that this scheme offers a high performance.

This paper is organized as follows. In Section II, we present the switch model. In Section III, we present a simulation study of the delay performance of this architecture under uniform and non-uniform traffic patterns. In Section IV, we present our conclusions.

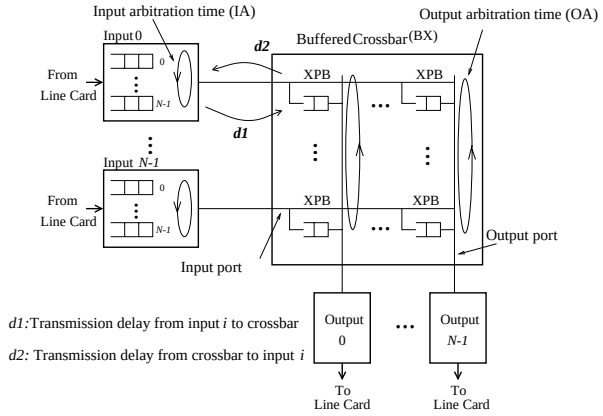

Fig. 1. $N \times N$ buffered crossbar with VOQ structure

II. Combined Input-Crosspoint Buffered Switch Model

Figure 1 shows a buffered crossbar (BX) with N input and output ports. In our switch model, there are N VOQs at each input. A VOQ at input i that stores cells for output j is denoted as $VOQ(i,j)$. A crosspoint (XP) element in the BX that connects input port i, where $0 \leq i \leq N-1$, to output port j, where $0 \leq j \leq N-1$, is denoted as $XP(i,j)$. The buffer at $XP(i,j)$ is denoted as $XPB(i,j)$. The size of $XPB(i,j)$, k, is given in number of cells that can be stored. A credit-based flow-control mechanism indicates to input i whether $XPB(i,j)$ has room available for a cell or not, as described in [2]. A VOQ is said to be eligible if the VOQ is non-empty and its corresponding XPB at BX has room to store a cell.

Arbitration scheme. Our novel arbitration scheme is round-robin based. Each time that a VOQ (or an XPB at an output) is selected by the arbiter, the VOQ gets the right to forward a frame, where a frame is formed by one or more cells. Each cell of a frame is dispatched in a time slot. The frame size is adaptively determined by the serviced and unserviced traffic, such that no intervention is needed to select the frame size. We call this arbitration scheme round-robin with adaptable-size frame (RR-AF).

In each VOQ (and XPB) there are two counters: a frame-size counter, $FSC_{i,j}(t)$, and a current service counter, $CSC_{i,j}(t)$. The value of $FSC_{i,j}(t)$, $|FSC_{i,j}(t)|$, indicates the frame size, that is, the maximum number of cells at time slot t that a $VOQ(i,j)$ can send in back-to-back time slots to the buffered crossbar, one cell per time slot. The initial value of $|FSC_{i,j}(t)|$ is one cell (i.e., its minimum value).[1] $CSC_{i,j}(t)$ counts the number of serviced cells at time slot t in a frame corresponding to a VOQ, where the frame size is indicated by FSC, in a regressive fashion.[2] The initial value of $CSC_{i,j}(t)$, $|CSC_{i,j}(t)|$ (as is its minimum value).

For the sake of clarity, we explain the input arbitration scheme by using the following pseudo-code, as seen at an input:

-At time slot t, starting from the pointer position j, find the nearest $VOQ(i,j')$ that has a cell and it is not inhibited by the flow-control mechanism, in a round-robin fashion.
-Send a cell from $VOQ(i,j')$ to $XPB(i,j')$.
-If $CSC_{i,j'}(t) > 1$ then
$CSC_{i,j'}(t+1) = CSC_{i,j}(t) - 1$,
the pointer points to j'.
-else $FSC_{i,j'}(t+1) = FSC_{i,j'}(t) + N$,
$CSC_{i,j'}(t+1) = FSC_{i,j'}(t+1)$,
the pointer points to (j'+1) module N.
-For VOQ(i, h), where $j \leq h < j'$ for $j < j'$, or $0 \leq h < j'$ and $j \leq h \leq N-1$ for $j > j'$:[3]
$FSC_{i,h}(t+1) = FSC_{i,h}(t) - 1$.
- Go to the next time slot.

The output arbitration works in a similar way to that at the inputs, considering $XPB(i,j)$ instead of VOQs. Fig. 2 shows an example of round-robin with adaptable-size frame. The queues are VOQs in a 3×3 switch. Let us assume that all queues have three cells each, as shown in Fig. 2.a. The cells will be leaving the input as shown in Fig. 2.b. Assuming that the FSC counter for each queue has the initial value of one, a cell from each queue is served in a round-robin fashion. Then, each frame size will be increased by three cells. As a result, the remaining two cells in each queue are served back-to-back.

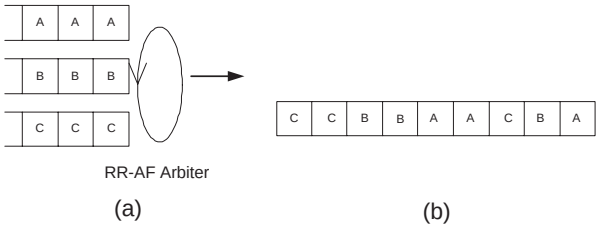

Fig. 2. Example of RR-AF among three queues

III. Performance Study

We studied the performance of two combined input-crosspoint buffered crossbar switches, one using RR-AF arbitration and the other using a simple round-robin (RR) arbitration. In addition, we included an output buffered (OB) switch in our study. The traffic models considered have destinations with uniform and a non-uniform distributions, the latter called unbalanced, with Bernoulli arrivals. In our switch model, we do not consider the segmentation and re-assembly delays. Our simulation results are obtained with a 95% confidence interval, not greater than 5% for the average cell delay.

A. Uniform Traffic

Fig. 3 shows simulation results of 32×32 switches of buffered crossbars with RR-AF, simple round-robin (RR), and an OB switch under uniform traffic with Bernoulli arrivals ($l = 1$) and bursts with average lengths of 10 and 100 cells ($l = 10$ and $l = 100$). The buffered crossbars have crosspoint buffers with size of one cell each. The burst length is exponentially distributed. The simulation shows that the

[1] We consider that $|FSC_{i,j}(t)|$ can be large as needed, although practical results have shown that its value does not reach large numbers.

[2] We used regressive fashion in CSC as CSC only considers FSC at the end of a serviced frame.

[3] Note that when $j' = j$, there is no $VOQ(i,h)$.

RR-AF arbitration scheme provides 100% throughput under uniform traffic.

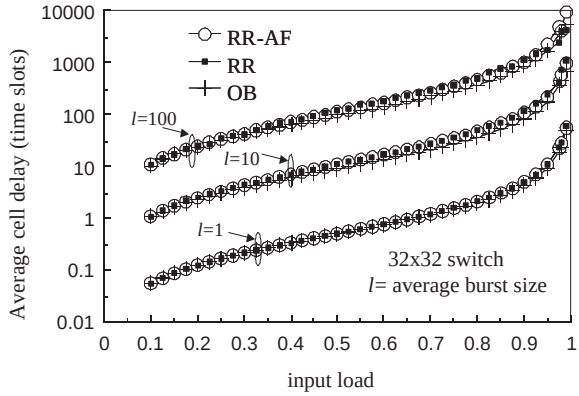

Fig. 3. Average delay of RR-AF under bernoulli and bursty traffic

Fig. 3 shows that the average delay performance of RR-AF under Bernoulli arrivals is close to that of RR, and therefore, to that of an OB switch. The adaptable frame-size condition in the arbitration does not degrade the performance neither does it increase the average delay significantly under this traffic model. RR-AF can be considered as a more general model of RR, where RR can be obtained by fixing the maximum frame size to one cell. As the RR-AF uses the history of serviced and unserviced traffic from each buffer (i.e., VOQ and XPB), the switch practically adapts itself to uniform traffic. In addition, Fig. 3 shows that RR-AF arbitration offers similar performance to that of an OB switch under bursty traffic. The average delay is then proportional to the burst length and the throughput is unaffected.

B. Non-uniform Traffic

We simulated the RR-AF and the RR arbitrations under non-uniform traffic models, such as the unbalanced traffic used in [2]. This traffic model uses a probability w as the fraction of input load directed to a single predetermined output, while the rest of the input load is directed to all outputs with uniform distribution. When $w = 0$, the offered traffic is uniform. On the other hand, when $w = 1$, the traffic is completely directional, from input i to output j, i.e., $i = j$. RR arbitration is used with a $k = 1$ and $k = N = 32$. Fig. 4 shows that RR-AF, with $k = 1$, provides near 100% throughput. This shows that RR-AF with $k = 1$ outperforms RR with $k = 32$. This results in a feasible implementation of buffered crossbars.

The high throughput of RR-AF is the product of providing service to a queue in proportion to its cell occupancy and to its previously received service, without starving the other queues. RR-AF ensures service to the queues with high load by increasing the frame size, and to the queues with low load by using round-robin selection. In addition, the decreasing policy for the frame-size counter ensures that the counter does not increase infinitely.

Under uniform traffic, the frame counters of the queues are not expected to increase largely as the cell arrival is uniformly distributed. The frame's increase and the decrease processes are balanced. This results in having an arbitration that behaves

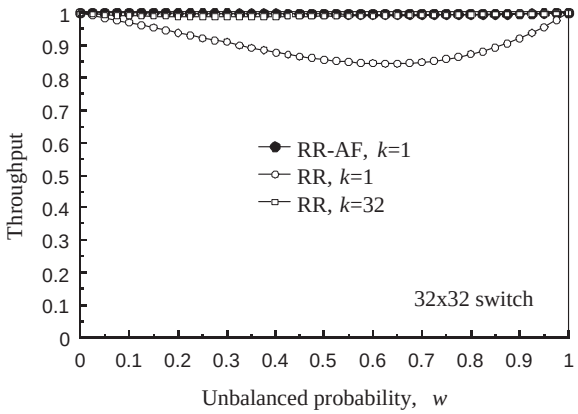

Fig. 4. Throughput performance of RR-AF under unbalanced traffic

as pure round-robin. Under unbalanced traffic, some queues are expected to have heavier loads than other queues. The queues with large occupancy will have a higher probability of finishing servicing a complete frame in each opportunity of service, and their frame size will increase consequently. The queues with low occupancy will tend to have a frame size rather small because they will miss service opportunities. This different behavior of frame sizes results in having higher service rate for queues with larger number of arrivals than those queues with a small number of arrivals. Moreover, the round-robin policy ensures that all queues receive service.

IV. Conclusions

We presented a novel arbitration scheme for input-crosspoint buffered crossbars. In our simulation, we observed that the presented round-robin scheme with adaptable-size frame provides nearly 100% throughput under uniform and unbalanced traffic models. With this arbitration scheme, we showed that a buffered crossbar with one-cell crosspoint buffers is sufficient to provide such throughput. Furthermore, our arbitration scheme does not need to compare the status of other queues or weights as it is based in simple round-robin. In addition to high throughput, this switch provides timing relaxation that allows high-speed arbitration and scalability. This results in a simplified and scalable arbitration scheme.

References

[1] E. Oki, N. Yamanaka, Y. Ohtomo, K. Okazaki, and R. Kawano, "A 10-Gb/s (1.25 Gb/s x8) 4 x 0.25-μm CMOS/SIMOX ATM Switch Based on Scalable Distributed Arbitration," *IEEE J. Solid-State Circuits*, vol. 34, no 12, pp. 1921-1934, December 1999.

[2] R. Rojas-Cessa, E. Oki, Z. Jing, and H. Jonathan Chao, "CIXB-1: Combined Input-One-cell-crosspoint Buffered Switch," *IEEE HPSR 2001*, pp. 324-329, May 2001.

[3] M. Karol, M. Hluchyj, "Queuing in High-performance Packet-switching," *IEEE J. Select. Area Commun.*, vol. 6, pp. 1587-1597, December 1988.

[4] N. McKeown, A. Mekkittikul, V. Anantharam, J. Walrand, "Achieving 100% Throughput in an Input-queued Switch," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1260-1267, August 1999.

[5] T. Javadi, R. Magill, and T. Hrabik, "A High-Throughput Algorithm for Buffered Crossbar Switch Fabric," *IEEE ICC 2001*, pp.1581-1591, June 2001.

[6] N. McKeown, "Scheduling algorithm for input-queued cell switches," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California at Berkeley, Berkeley, CA, 1995.