Framed Round-Robin Arbitration with Explicit Feedback Control for Combined Input-Crosspoint Buffered Packet Switches

Zhen Guo and Roberto Rojas-Cessa

Abstract— **This paper introduces a frame-based round-robin arbitration scheme with explicit feedback control (FRE) for combined input-crosspoint buffered packet switches. We consider a CICB switch with fixed-length packets, called cells. The proposed scheme dynamically sets the frame size according to the amount of cell accumulation at the input queues. We study FRE when applied to a continuous system. We transport the FRE concept into a discrete system as an arbitration scheme for a packet switch, and study the switching performance. We combine FRE, which is used as the input arbitration scheme, with other round-robin based schemes, used as output arbitration schemes. The resulting combined schemes provide high throughput under several admissible traffic patterns when used in a switch with one-cell crosspoint buffers and no speedup.**

Index Terms— **Round-robin, buffered crossbar, queued crosspoint, stability, explicit feedback.**

I. Introduction

Combined input-crosspoint buffered (CICB) switches use time efficiently as input and output port selections are performed separately. At each input, there is one input arbiter that selects which virtual output queue (VOQ) in this input is allowed to transfer a cell to the buffered crossbar.[1] In a similar way, an output arbiter independently selects which crosspoint buffer is allowed to transfer a cell out of the buffered crossbar. The input and output arbiters are coupled by a credit-based flow control [1]. Here, we consider that switches work with fixed-length packets, or cells, and variable-length packets are segmented and reassembled at entering and before leaving a switch, respectively.

CICB switches are also attractive because they provide high throughput under admissible traffic using simple arbitration schemes [2], [3], [4], [5]. Several schemes that provide 100% throughput under uniform traffic have been proposed, and the search for such throughput under admissible nonuniform traffic patterns by switches with no speedup is still under way. One way to provide 100% throughput under nonuniform traffic patterns is by using weight-based arbitration schemes, where weights are assigned to input queues proportionally to their occupancy or HOL cell age. It has been shown that weight-based [5] schemes in buffered crossbars can provide high throughput under various traffic patterns. Two schemes were presented in [3] and [5]: one is based on the selection of the longest VOQ occupancy at inputs and round-robin selection at the outputs; the other scheme is based on the selection of the oldest cell first (OCF) instead of VOQ occupancy. These schemes provide 100% throughput under uniform traffic. Other schemes that use the crosspoint buffer occupancy instead of the VOQ occupancy have been proposed [6].

However, weight-based schemes need to perform comparisons among all contending queues, which can be a large number, thus increasing the implementation complexity. Moreover, weight-based schemes (e.g., queue-occupancy based) may starve some queues for long time to provide more service to the congested ones, presenting unfairness.

Schemes based on round-robin selection have been shown to provide higher level of fairness. Furthermore, it has been shown that a CICB switch using one-cell crosspoint buffers (CIXB-1), a round-robin arbitration (RR) scheme for input and output arbitration, and a credit-based flow control provide 100% throughput for uniform traffic [4]. The input arbiter selects a VOQ if there is at least a single eligible VOQ. The eligibility of VOQs is determined by the flow control mechanism. However, the throughput of this scheme under nonuniform traffic is not 100% when the crosspoint buffer size is equal to one cell.

In order to improve the throughput of CIXB-1 for nonuniform traffic, a frame-based round-robin scheme with adaptable frame size, RR-AF, was proposed [7]. This scheme sends cells from a VOQ in a back-to-back fashion, where the number of cells sent continuously depends on the frame size. In this scheme, the frame size increases by a constant number, independently of the actual size needed, each time a frame is completely served. The frame size decreases each time a VOQ misses an opportunity to be served. This scheme provides high performance, but it might be difficult to analyze.

To give a better insight of the feedback nature of arbitration schemes for CICB switches, this paper introduces a frame-based round-robin with explicit feedback control (FRE) arbitration scheme for CICB packet switches. This scheme also adopts a frame-based arbitration scheme and dynamically sets the frame size according to the input load and to the number of cells accumulated in the input queues. The stability of this scheme is analyzed by using control theory and the switching performance is evaluated by computer simulations. We combine FRE as the input arbitration scheme with other weightless arbitration schemes as output arbitration schemes. We show that FRE provides high throughput under several admissible traffic patterns when using a CICB switch with one-cell crosspoint buffers and no speedup.

This work is supported in part by National Science Foundation under Grant Awards 0435250 and 0423305.

The authors are with th Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102. Roberto Rojas-Cessa is the corresponding author. Email:rrojas@njit.edu.

[1] A VOQ is a queue that stores cells going to a specific output.

The remainder of this paper is organized as follows. Section II describes the switch architecture and the notations used in this paper, and shows the control theoretic analysis of the proposed scheme. Section III describes the FRE concept as an arbitration scheme. Section IV presents the performance study of a 32×32 CICB switch using FRE. Section V presents the conclusions.

II. CICB Switch Model

Figure 1 shows a buffered crossbar switch with N inputs and outputs. In this 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 element in the buffered crossbar 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 $CP_{i,j}$. The buffer at $CP_{i,j}$ is denoted as $CPB_{i,j}$, and it is considered of k-cell size. In this paper we use $k = 1$.

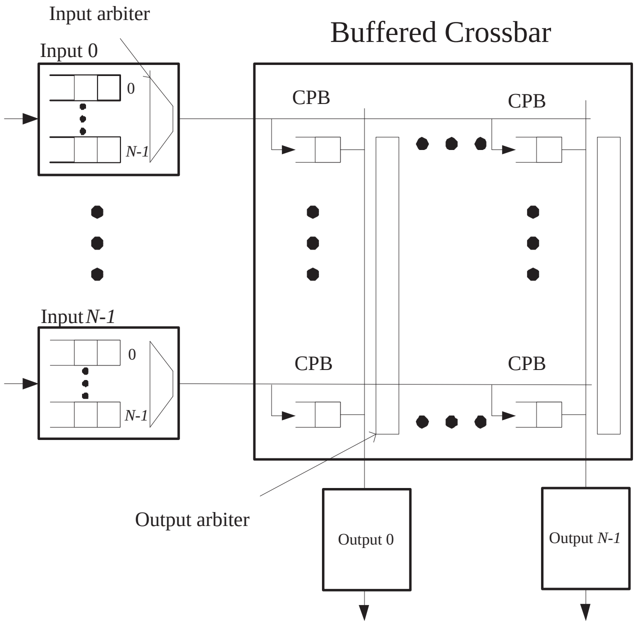

Fig. 1. Combine input-crosspoint queued crossbar switch.

Different from previous approaches, the design of the FRE scheme is started by defining explicitly the feedback system needed to control the service rate to a VOQ. In a CICB switch, the VOQ's sending rate is controlled by a frame size. The frame size is the number of packets transferred into the buffered crossbar.

The FRE scheme is analyzed by making an analogy to a continuous level-control system for a fluid container where the input arbiters are represented as the controller of an actuator system that drains fluid from the container.

A. Controlling the Service Rate by Explicit Feedback

Let's consider that the minimum unit of data can be divided into an infinitesimal amount and packetization is not strictly required. Therefore, the switching of data can be performed at any part of a packet. This is, instead of switching packets by their boundaries, it is assumed that fluid data can be switched at any time and that the packets can be re-assembled at the outputs.

Consider a fluid container with a inflow rate V_i, the fluid level h is required to be kept constant by changing the fluid outflow rate through changing the drain area. The inflow rate V_i is used to emulate the input traffic rate in a switch and V_o is the outflow rate, which is the serving rate provided by the arbitration scheme.

The equation governing the change in fluid level is that the rate of change of fluid level is equal to the inflow rate minus the outflow rate.

$$\frac{dh(t)}{dt} = V_i(t) - V_o(t) \tag{1}$$

The Laplace transform is derived:

$$H(s) = \frac{V_i(s) - V_o(s)}{s} \tag{2}$$

Therefore, the plant is a first-order system.

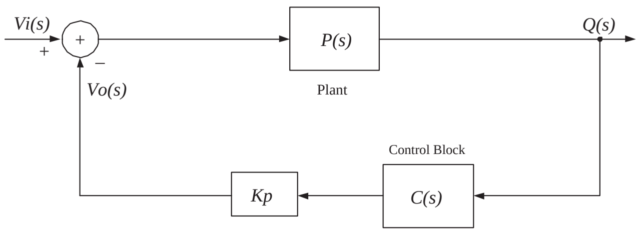

Fig. 2. Block diagram of a feedback control system.

Figure 2 shows the block diagram of a fluid system with a feedback control that makes the outflow rate follow the inflow rate to keep the fluid level of the container from raising indefinitely.

It is assumed that the container's height is large enough to avoid overflow (this is analogous to a packet queue of large capacity) so that the measuring of the differences of fluid accumulation can be performed. If the inflow rate is greater than the outflow rate, the feedback controller is expected to increase the outflow rate in order to bring down the level. The integral plant $P(s)$ reflects the dependency of the level on the difference between the inflow and outflow rates.

The stability behavior of this feedback system is studied by control analysis and by using computer simulation (using Simulink). Figure 3 shows the block diagram in Simulink of a fluid system with a feedback control that permits the outflow rate to effectively follow the inflow rate. The control block $C_1(s)$ is used to generate Δ_q, which is the difference between the actual current fluid level $q_{current}$ and the previous fluid level q_{pre}. The control block $C_2(s)$ is used to generate FC, which is regarded as the amount of fluid to be drained out. FC is adjusted by the previous FC plus the updated Δ_q. With the block $\frac{1}{\tau}$ and proportional gain K_p, the outflow rate is obtained.

Figure 4 shows the simulation results. In this example, we use a random number generator as the source of the inflow rate with mean of 100 and variance of 10. In this figure, the top

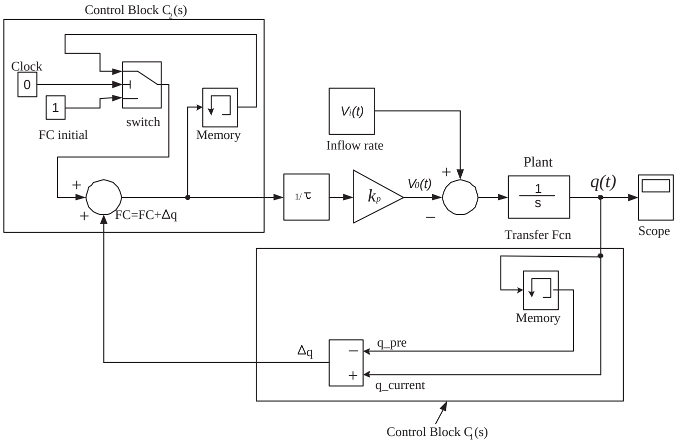

Fig. 3. Block diagram of the explicit feedback control system.

graph shows the inflow rate, and the bottom graph shows the outflow rate. This figure shows that the outflow rate follows closely the inflow rate.

Figure 5 indicates that the fluid level is stable around the constant value of 9. Because the outflow rate effectively follows the inflow rate, the fluid does not accumulate so as to cause fluid overflow. For the sake of brevity, we show the stability analysis in the Appendix.

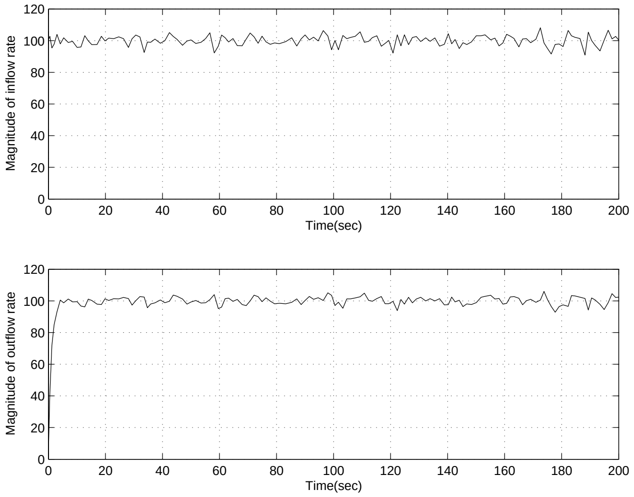

Fig. 4. Simulink results of the outflow rate tracking the inflow rate.

As the control system for a continuous system shows stability, the explicit feedback concept is transported into a discrete system, such as a cell-based arbitration scheme. The objective of the new scheme is to make a VOQ get a service rate according to the accumulation of cells in it. This cell occupancy is an indicator of the magnitude of the input load. If the input traffic is heavy in one queue and the service rate is such that accumulation of cells occurs, the queue requests more service. Otherwise, the queue requests less service. Based on the analysis above, the difference in the VOQ occupancy between two consecutive service cycles is used to adaptively update the frame size.

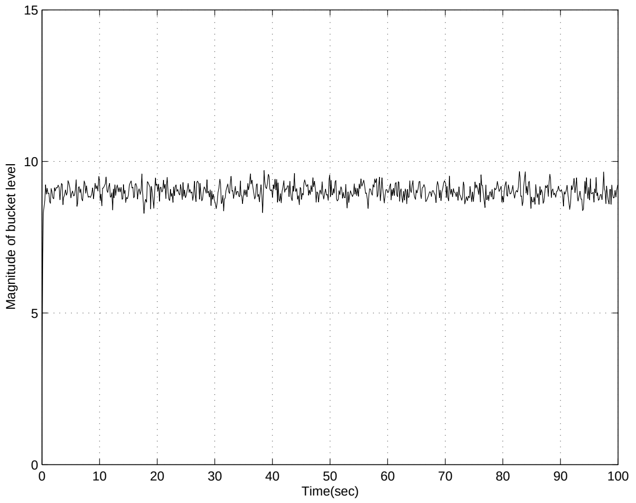

Fig. 5. Simulation result on fluid level of the example.

III. FRE Arbitration Scheme

The proposed arbitration scheme is round-robin based. In each VOQ (and CPB), there are two counters: a frame-size counter, $FC_{i,j}(T)$, and a current service counter, $CC_{i,j}(t)$, where T is the service cycle that starts after the time slot when a frame is completely served, and t is any time slot when $VOQ_{i,j}$ receives service. The value of $FC_{i,j}(T)$, $|FC_{i,j}(T)|$, indicates the frame size; that is, the maximum number of cells that $VOQ_{i,j}$ can send in consecutive time slots to the CPB, one cell per time slot. The initial (an minimum) value of FC, $|FC_{i,j}(0)|$, is one cell. The value of $CC_{i,j}(t)$, $|CC_{i,j}(t)|$, indicates the number of serviced cells of a given frame of $VOQ_{i,j}$ at time slot t. A regressive-fashion count is used in CC as CC only considers FC at the end of a serviced frame. The minimum value of $CC_{i,j}(t)$ is one cell.

The input arbitration process is as follows. An input arbiter selects an eligible $VOQ(i,j')$ in round-robin fashion, starting from the pointer position, j. A VOQ is considered eligible if the VOQ is not empty and the corresponding CPB is not full.

For the selected $VOQ(i,j')$, if $|CC_{i,j'}(t)| > 1$, $|CC_{i,j'}(t+1)| = |CC_{i,j'}(t)| - 1$, and the input pointer remains at $VOQ(i,j')$, so that this VOQ has the higher priority for service in the next time slot and the frame transmission can continue.

If $|CC_{i,j'}(t)| = 1$, a new service cycle is defined, the input pointer is updated to $(j'+1) \quad \text{modulo} \quad N$, $|FC_{i,j'}(T+1)|$ is increased by $\Delta_{i,j}$ cells, and $|CC_{i,j'}(t+1)| = |FC_{i,j'}(T+1)|$, where $\Delta_{i,j}$ is defined as the current actual queue occupancy $Q(T+1)$ minus the previous queue occupancy $Q(T)$. Note that the number of time slots in each service cycle is not necessarily constant.

For the sake of clarity, the following pseudo-code describes the input arbitration scheme, as seen at an input:

-At time slot t, starting from the pointer position j, find the nearest eligible $VOQ(i,j')$ in a round-robin fashion.

-Send the HOL cell from $VOQ(i,j')$ to $CPB(i,j')$ time slot $t+1$.

```
    -If $|CC_{i,j'}(t)| > 1$ then
        $|CC_{i,j'}(t+1)|$=$|CC_{i,j'}(t)| - 1$,
        the pointer points to j'.
    -else (i.e., $|CC_{i,j'}(t)| = 1$)
        Set the new serving cycle as $T + 1$,
        $\Delta_{i,j'} = Q_{i,j'}(T) - Q_{i,j'}(T-1)$,
        $|FC_{i,j'}(T+1)| = |FC_{i,j'}(T)| + \Delta_{i,j'}$,
        $|CC_{i,j'}(t+1)| = |FC_{i,j'}(T+1)|$,
        the pointer points to (j'+1) modulo N.
- Go to the next time slot.
```

Figure 6 shows an example of FRE. The figure depicts the evolution of FC and Q of a VOQ, where Q shows the occupancy of this VOQ and FC shows how the frame size is set each time the VOQ gets a frame service completed. As the figure shows, each time FC is updated, the addition to FC is the difference between the last frame size and the positive (or negative) change of the VOQ occupancy.

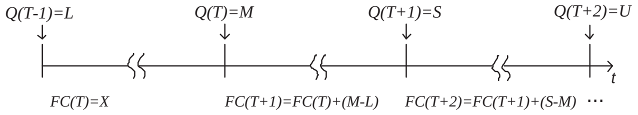

Fig. 6. Example of FRE arbitration.

Here, we consider three different schemes for output arbitration: a) round-robin (RR) selection, where the output pointer moves to one position beyond the selected one, or $(i+1)$ modulo N, when $CPB(i,j)$ is selected, b) persistent round-robin (PRR) selection, where the pointer moves to the selected input, or i when $CPB(i,j)$ is selected, and c) FRE, where the values of $Q(T)$ are the occupancy of the VOQs at time T (note that the CPB size considered in this paper has a size of one cell, therefore, we do not consider the CPB occupancy). The combination of FRE as input arbitration with RR, PRR, and FRE are denoted as FRE-RR, FRE-PRR, and FRE-FRE, respectively.

IV. Performance Evaluation

The performance evaluations are produced by computer simulation. We consider traffic models with a uniform distribution with Bernoulli and bursty arrivals (two-state modulated Markov model), and with nonuniform distributions with Bernoulli arrivals. The simulation does not consider the segmentation and re-assembly delays. Simulation results are obtained with a 95% confidence interval, not greater than 5% for the average cell delay.

A. Uniform Traffic

We assume cells arrive at each input in a slot by slot manner. Under a Bernoulli arrival process, the probability that there is a cell arriving in each time slot is identical and independent of any other slot. This probability is referred as the offered load to the input. If each cell is equally likely to be destined for any output, the traffic becomes uniformly distributed over the switch.

In the bursty traffic model, each input alternates between active and idle periods of geometrically distributed duration. During an active period, cells destined for the same output arrive continuously in consecutive time slots.

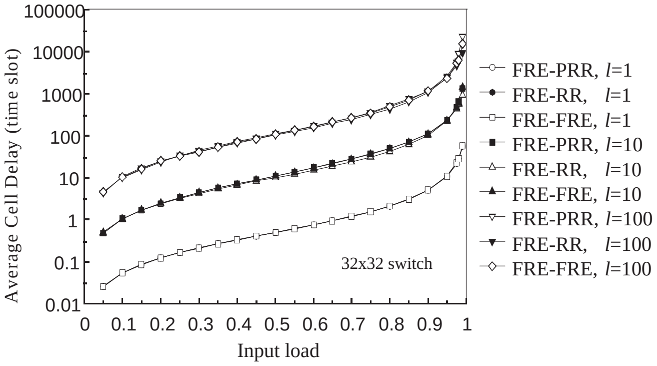

Fig. 7. Performance with Bernoulli and bursty arrivals.

Figure 7 shows simulation results of 32×32 CICB switch with FRE-PRR, FRE-RR, and FRE-FRE under uniform traffic with Bernoulli arrivals ($l = 1$) and bursts with average lengths of 10 and 100 cells ($l = 10$ and $l = 100$). The simulation shows that the average delay is proportional to the burst length and that the throughput is unaffected at any load. The simulation shows that all the three arbitration schemes provide nearly 100% throughput under uniform traffic.

B. Nonuniform Traffic: Unbalanced

The unbalanced 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. Let us consider input port s, output port d, and the offered input load for each input port ρ. The traffic load from input port s to output port d, $\rho_{s,d}$ is given by,

$$\rho_{s,d} = \begin{cases} \rho\left(w + \frac{1-w}{N}\right) & \text{if } s = d \\ \rho\frac{1-w}{N} & \text{otherwise.} \end{cases} \tag{3}$$

When $w = 0$, the offered traffic is uniform. On the other hand, when $w = 1$, it is completely directional, from input s to output d, where $s = d$. Figure 8 shows simulation results of

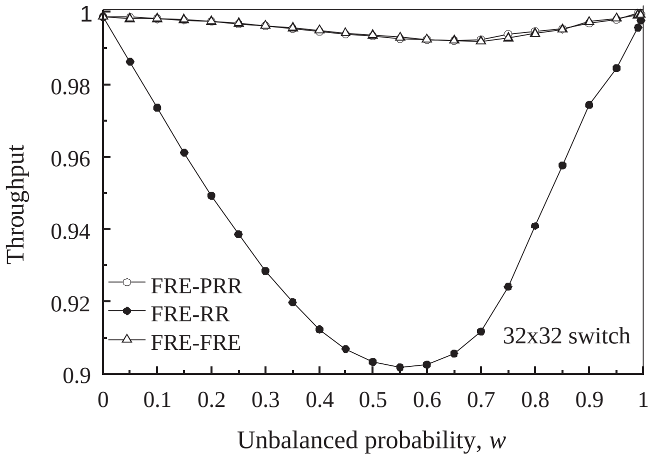

Fig. 8. Performance under unbalanced traffic.

32×32 CICB switch with FRE-PRR, FRE-RR, and FRE-FRE under unbalanced traffic. The simulation shows that FRE-PRR and FRE-FRE provide high throughput (more than 99%) under unbalanced traffic. However, the throughput of FRE-RR when $w = 0.55$ is 90.19%. This is because RR selection tries to assign the same service rate to all queues, independently of the individual input loads. The other two schemes are more sensitive to the individual input loads than RR.

C. Nonuniform Traffic: Diagonal

The diagonal traffic can be represented as $d\rho(i,j) = d\rho$ for $i = j$, $(1-d)\rho$ for $j = (i+1)\ modulo\ N$, where d is the diagonal degree probability. This traffic model presents load distributions among two outputs per each input, instead of one output as the unbalanced traffic models does. Figure 9 shows the performance of the different arbitration schemes under diagonal traffic with $d = 0.75$. The figure shows that the throughput of FRE-RR is the highest, nearly 100%, and that the average cell delay of FRE-RR is the smaller than that of the other two.

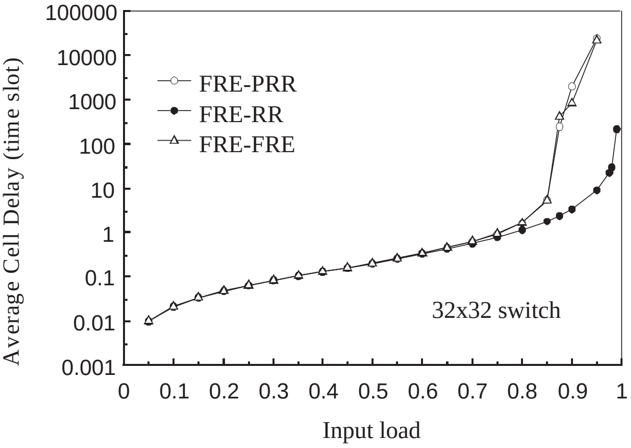

Fig. 9. Performance under diagonal traffic.

D. Nonuniform Traffic: Chang's and Asymmetric

The schemes are also tested under other nonuniform traffic models: Chang's [9] and asymmetric [10].

Chang's traffic model can be defined as $\rho = 0$ for $i = j$ and $\rho = \frac{1}{N-1}$, otherwise. Figure 10 shows the average cell delay experienced by a 32×32 switch under this traffic model. As the figure shows, there is no significant difference on the average cell delay of FRE-PRR, FRE-RR, and FRE-FRE under this traffic model. Figure 11 shows the average cell delay experienced by a 32×32 switch under asymmetric traffic model. The average cell delay of FRE-FRE is the largest among all three.

In summary, based on the evaluations of the performance of FRE-PRR, FRE-RR, and FRE-FRE under all the traffic models above and while considering the implementation cost and complexity, FRE-PRR shows an advantage over other schemes as this scheme provides the highest throughput under unbalanced traffic and it is easy to implement.

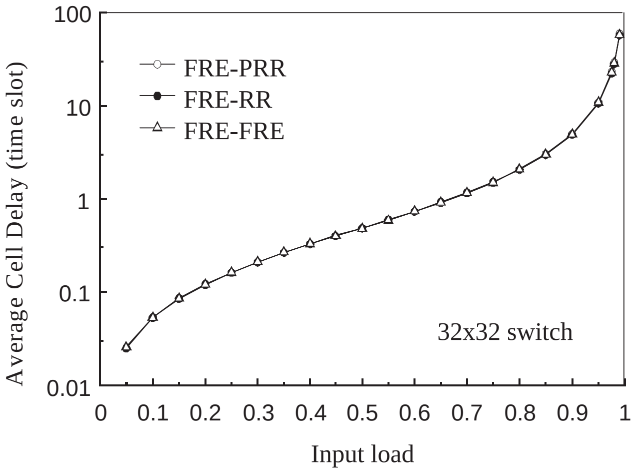

Fig. 10. Performance under Chang's traffic.

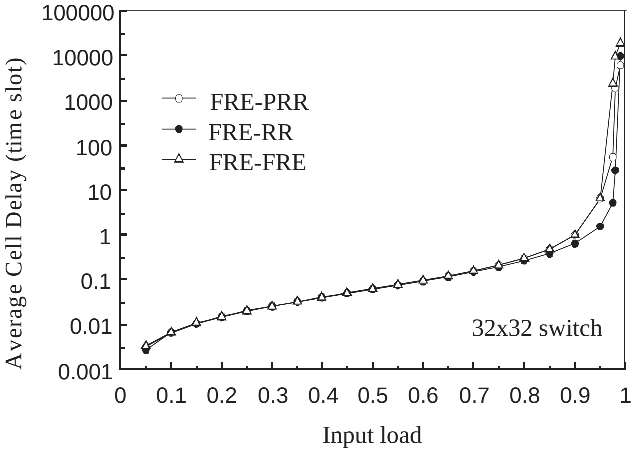

Fig. 11. Performance under asymmetric traffic.

V. Conclusions

This paper introduced a frame-based round-robin arbitration scheme with explicit feedback control for CICB packet switches. The proposed arbitration scheme dynamically sets the frame size according to the input load of a queue by considering the accumulation of cells in it. We showed that the concept of explicitly feedback in a continuous system is stable, and provided an interpretation of the explicit feedback approach into a discrete system for a cell-based switch. We tested FRE as the input arbitration in combination with RR, PRR, and FRE as output arbitration schemes. These combined schemes deliver high performance under uniform and nonuniform traffic models when used in a buffered crossbar with one-cell crosspoint buffers.

Among all three combined schemes, FRE-FRE is probably the scheme with the highest implementation cost as the output arbitration needs the VOQ occupancy, and this may increase the transmission overhead.

Because the proposed FRE scheme is based on round-robin selection, the arbitration needs not to compare the status of different VOQs as weight-based schemes do. The amount of service that FRE designates for a VOQ is based on a frame size whose value changes according to the accumulation of cells resulted from the incoming load and received service in previous service cycles. In this way, the scheme explicitly controls the service rate provided to a VOQ.

References

[1] H.T. Kung, K. Chang, "Receiver-Oriented Adaptive Buffer Allocation in Credit-Based Flow Control for ATM Networks," Proc. *IEEE INFOCOM'95*, Vol. 1, pp. 239-252, 1995.

[2] Y. Doi and N. Yamanaka, "A High-Speed ATM Switch with Input and Cross-Point Buffers," *IEICE Trans. Commun.*, vol. E76, no.3, pp. 310-314, March 1993.

[3] M. Nabeshima, "Performance Evaluation of Combined Input and Crosspoint-Queued Switch," *IEICE Trans. Commun.*, Vol. E83-B, No. 3, pp.742-745, March 2000.

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

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

[6] L. Mhamdi and M. Hamdi, "MCBF: a High-Performance Scheduling Algorithm for Buffered Crossbar Switches," *IEEE Commun. Letters*, Vol. 7, Issue 9, pp. 451-453, September 2003.

[7] R. Rojas-Cessa and E. Oki, "Round-Robin Selection with Adaptable-Size Frame in a Combined Input-Crosspoint Buffered Switch," *IEEE Communications Letters*, Vol. 7, No. 11, pp. 555-557, Nov. 2003.

[8] A. Bianco, M. Franceschinis, et. al.,"Frame-Based Matching Algorithms for Input-Queued Switches," Proc. *IEEE HPSR 2002*, pp. 69-76, May 2002.

[9] C-S. Chang, W-J. Chen, and H-Y. Huang "Birkhoff-von Neumann Input Buffered Crossbar Switches," Proc. *IEEE INFOCOM 2000*, pp.1614-1623, March 2000.

[10] R. Schoenen, G. Post, and G. Sander, "Weighted Arbitration Algorithms with Priorities for Input-Queued Switches with 100% Throughput," Proc. *Broadband Switching Symposium'99*, 1999. http://www.schoenen-service.de/assets/papers/Schoenen99bssw.pdf

[11] Q. Chen and Q.W. Yang, "AQM Controller Design for IP routers Supporting TCP Flows Based on Pole Placement," *IEE Proc. Commun.*, Vol. 151, No. 4, pp. 347-354, 2004.

Appendix

Here, we show the stability analysis on the feedback control system in Section II.

We redraw the block diagram in Figure 3 as Figure 12 shows.

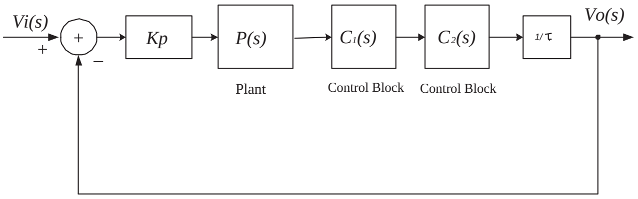

Fig. 12. Block diagram of outflow rate tracking the inflow rate with a proportional control.

$C_1(s)$ is used to generate Δ_q.

$$\Delta_q(t) = q(t) - q(t-\alpha_1). \tag{4}$$

Here, α_1 is period time for generating Δ_q.

The Laplace transform is derived:

$$\Delta_q(s) = q(s)(1-e^{-\alpha_1 s}). \tag{5}$$

The transfer function for control block $C_1(s)$ is

$$C_1(s) = \frac{\Delta_q(s)}{q(s)} = 1 - e^{-\alpha_1 s}. \tag{6}$$

$C_2(s)$ is used to generate FC

$$FC(t) = FC(t-\alpha_2) + \Delta_q(t). \tag{7}$$

Here, α_2 is period time for updating FC. The Laplace transform is derived:

$$FC(s) = FC(s)e^{-\alpha_2 s} + \Delta_q(s). \tag{8}$$

The transfer function for control block $C_2(s)$ is

$$C_2(s) = \frac{FC(s)}{\Delta_q(s)} = \frac{1}{1-e^{-\alpha_2 s}}. \tag{9}$$

We obtain the transfer function of the system from the block diagram in Figure 12:

$$H(s) = \frac{V_o(s)}{V_i(s)} = \frac{K_p R(s)}{1+K_p R(s)}, \tag{10}$$

where

$$R(s) = \frac{C_1(s)C_2(s)}{s\tau} = \frac{1-e^{-\alpha_1 s}}{1-e^{-\alpha_2 s}}\frac{1}{s\tau} \tag{11}$$

If the pole of the system is in left half plane (LHP), the feedback control system is stable. The denominator of the transfer function $H(s)$ is then:

$$1 + \frac{K_p}{s\tau}\frac{1-e^{-\alpha_1 s}}{1-e^{-\alpha_2 s}}.$$

There are two cases:

(1) when $\alpha_1 = \alpha_2$, the denominator of the transfer function $H(s)$ is $1+\frac{K_p}{s\tau}$, then as long as $\frac{K_p}{\tau}$ is positive, the pole of the system is in LHP, the feedback control system is stable.

(2) when $\alpha_1 \neq \alpha_2$, the term $e^{-\alpha s}$ can be approximated by Taylor series expansion of exponential function or first order $Pad\acute{e}$ approximation [11]. We also show that the poles of the system are in LHP. Therefore, the feedback control system is stable.

□