

# Captured-Frame Eligibility and Round-Robin Matching for Input-Queued Packet Switches

Roberto Rojas-Cessa and Chuan-bi Lin  
 Department of Electrical and Computer Engineering  
 New Jersey Institute of Technology  
 Newark, NJ 07102  
 Email: {rrojas, cl23}@njit.edu

**Abstract**—A variety of matching schemes that deliver high throughput under traffic with uniform distributions has been proposed. However, there is a need of matching schemes that provide high throughput under several admissible traffic patterns, including those with nonuniform distributions, while keeping implementation complexity low. In this letter, first, we introduce the captured frame concept for matching schemes in IQ switches. Second, we propose a round-robin based matching scheme, uFORM, which uses the proposed concept for cell matching eligibility. We show via simulation that our matching scheme delivers 99% throughput under several nonuniform traffic patterns, and retains the high performance under uniform traffic that round-robin matching schemes are known to offer.

**Index Terms**—input-queued switch, virtual output queue, captured frame, frame eligibility, unbalanced traffic

## I. INTRODUCTION

The introduction of virtual output queues (VOQs), where one queue per output port is allocated at an input port of an input queued (IQ) packet switch, is known to remove the head-of-line (HOL) blocking problem. HOL blocking causes idle outputs to remain so, even in the existence of traffic for them at an idle input [1], thus impeding high throughput delivery.

This paper considers the following common practices in packet switch design: 1) segmentation of incoming variable-size packets at the ingress side of a switch to perform internal switching with fixed-size packets, or cells, and re-assembling the packets at the egress side before they depart from the switch; 2) use of VOQs; and 3) use of crossbar fabrics.

One major requirement for an input-queued switch is the delivery of high throughput under admissible traffic conditions [2]. The matching scheme used in crossbar-based IQ switch determines in large measure the achievable throughput.

Maximum weight matching (MWM) schemes can provide 100% throughput under admissible traffic [2], but their complexity is too high to be considered practical. Maximal-weight matching schemes can be considered as an alternative [4]. However, they may need a large number of iterations to achieve satisfactory matching results. Therefore, these schemes may have a high complexity. Maximal-size matching schemes are an attractive solution [3]. Schemes based on round-robin matching, such as *i*SLIP [4], DRRM [5], and

SRR [7] have been shown to deliver 100% throughput under uniform traffic. Other schemes with improved features can deliver smaller average delays [5], [7], [8], [9]. However, under nonuniform traffic patterns, only switches with an extra load-balancing stage [10] have been shown to offer 100% throughput. With a single-stage switch, the exhaustive dual round-robin matching (EDRRM) has been shown to achieve a throughput higher than *i*SLIP and DRRM under nonuniform traffic patterns. Furthermore, frame<sup>1</sup> based matching have been shown to improve switching performance under different traffic scenarios [11], [12].

This letter introduces the captured-frame concept and proposes the unlimited frame-size occupancy-based round-robin matching (uFORM) scheme for IQ switches, based on round-robin selection and on the captured-frame concept. This paper shows that the capture-frame concept, used for cell matching eligibility, improves the performance of arbitration schemes. As an example, this concept is tested with random and round-robin selections. The result is that the framed schemes achieve 99% throughput under several nonuniform traffic patterns. In addition, this paper shows that uFORM retains the high performance that round-robin schemes are known to offer under uniform traffic.

This letter is organized as follows. Section II presents the switch model under study and definitions used in this paper. Section III introduces the proposed arbitration scheme. Section IV presents a simulation study of the throughput and delay performance of the resulting switch under uniform and nonuniform traffic patterns. Section V presents the conclusions.

## II. SWITCH MODEL AND PRELIMINARY DEFINITIONS

We consider an IQ switch with  $N$  input and output ports. There are  $N$  VOQs at each input port. A VOQ at input port  $i$  that stores cells for output port  $j$  is denoted as  $VOQ(i, j)$ .

The following definitions are used in the description of the proposed matching scheme. A frame is related to a VOQ. A frame is the set of one or more cells in a VOQ that are eligible for matching in successive time slots. Only the HOL cell of the VOQ is eligible per time slot. A VOQ is said to be in **on-service** status if the VOQ has a frame size of two or more cells and the first cell of the frame has been matched. An input

This work was supported in part by New Jersey Institute of Technology under Grant 421070.

Roberto Rojas-Cessa is the corresponding author. Email: rrojas@njit.edu

<sup>1</sup>The frame term used in this letter is different from the one used in layer 2 of the Open Systems Interconnection model.

is said to be in on-service status if there is at least one on-service VOQ. A VOQ is said to be in **off-service** status if the last cell of the VOQ's frame has been matched or no cell of the frame has been matched. Note that for frame sizes of one cell, the associated VOQ is off-service during the matching of its one-cell frame. An input is said to be off-service if all VOQs are off-service.

At the time  $t_c$  of matching the last cell of the frame associated to  $VOQ(i, j)$ , the next frame is assigned a size equal to the cell occupancy at  $VOQ(i, j)$ . We define this as as captured frame size.<sup>2</sup> Cells arriving to  $VOQ(i, j)$  at time  $t_d$ , where  $t_d > t_c$ , are not considered for matching until the current frame is totally served and a new frame is captured.

### III. UNLIMITED FRAME-SIZE OCCUPANCY-BASED ROUND-ROBIN MATCHING SCHEME

For each output, there is an output arbiter  $a_j$ , for each input, there is an input arbiter  $a_i$ . Each arbiter has a pointer that indicates the counter-part port with the highest priority position in a round-robin schedule.

For each VOQ there is a captured frame-size counter,  $CF_{i,j}(t)$ . The value of this counter, also represented by  $CF_{i,j}(t)$ , indicates the frame size; that is, the maximum number of cells that a  $VOQ(i, j)$  can have as candidates in the immediate and future time slots.  $CF_{i,j}(t)$  takes a new frame-size value when the last cell of the current frame of  $VOQ(i, j)$  is matched.  $CF_{i,j}(t)$  decreases its count each time a cell of the frame is matched, other than the last cell. All VOQs are initially considered with a frame size of one and in off-service status.

This scheme follows the request-grant-accept steps, as in the  $i$ SLIP algorithm, and these are:

**Step 1: Request.** Non-empty on-service VOQs send a request to their destined outputs. Non-empty off-service VOQs send a request to their destined outputs only if input  $i$  is off-service.

**Step 2: Grant.** If output arbiter  $a_j$  receives any requests, it chooses the request from the on-service VOQ (also called on-service request) that appears next in a round-robin schedule, starting from the pointer position. If none on-service request exists, the output arbiter chooses the off-service request that appears next in a round-robin schedule, starting from its pointer position.

**Step 3: Accept.** If input arbiter  $a_i$  receives any grants, it accepts the on-service grant that appears next in a round-robin schedule, starting from the pointer position. If none on-service grant exists, the arbiter chooses the off-service grant that appears next in round-robin schedule, starting from its pointer position. The input and output pointers are updated to one position beyond the matched one. In addition,  $CF_{i,j}$  updates its value according to the following: if the input arbiter  $a_i$  accepts a grant from output arbiter  $a_j$ , and:

- i)  $CF_{i,j}(t) > 1$  then  $CF_{i,j}(t+1) = CF_{i,j}(t) - 1$  and  $VOQ(i, j)$  is set as on-service.

<sup>2</sup>We call this captured frame as it is the equivalent of having a snapshot of the occupancy of a VOQ at time  $t_c$ , where the occupancy determines the frame size.

- ii)  $CF_{i,j}(t) = 1$  then  $CF_{i,j}(t+1)$  is assigned the occupancy of  $VOQ(i, j)$ , and  $VOQ(i, j)$  is set as off-service.

This scheme can perform multiple iterations. However, that is out of the scope of this letter as we are interested in obtaining a matching result in a single iteration. The term unlimited, used in the name of the scheme, comes from letting the frame size have the size of the queue occupancy, without limitation.

### IV. PERFORMANCE EVALUATION

We consider  $i$ SLIP (with one iteration, or 1SLIP) for comparison purposes, the parallel iterative matching (PIM) scheme [3] and our framed version of it (uFPIM) to show the improvement of a matching scheme using the frame concept, and uFORM. uFPIM is a three-phase matching scheme, as uFORM, that uses random selection instead of round-robin. The performance evaluations are produced through computer simulation. The traffic models considered have destinations with uniform and nonuniform distributions, with Bernoulli arrivals. The simulation does not consider the segmentation and re-assembly delays for variable-size packets. Simulation results are obtained with a 95% confidence interval, not greater than 5% for the average cell delay.

#### A. Uniform Traffic

Figure 1 shows simulation results of four  $32 \times 32$  IQ switches, each one with a different matching scheme: 1SLIP, PIM, uFPIM, and uFORM under uniform traffic with Bernoulli arrivals. uFPIM and uFORM use the captured-frame concept. This figure shows that uFORM, as  $i$ SLIP, delivers 100% throughput under uniform traffic. PIM is known to offer about 60% throughput. However, when using the captured-frame concept in uFPIM, the performance is close to 100% throughput. The reason for the improvement shown by uFPIM is that once a match is achieved, the match between a VOQ and the output port is kept during the frame duration. Therefore, the number of contending ports is reduced with each match, during the frame service time.

Under uniform traffic, the desynchronization effect, shown by  $i$ SLIP, is also present in uFORM. This effect and the frame service policy allow uFORM to offer high throughput and low average cell delay under uniform traffic.

Under uniform traffic, the average frame size is small, as the uniform distribution of traffic among VOQs results in small average queue occupancies. Note that uFORM does not suffer from VOQ starvation, even in the case when VOQ occupancy is large, because the captured frame has a finite size, and the arrival of new cells does not affect the value of the frame counter arbitrarily.

#### B. Nonuniform Traffic

We simulated these four schemes under several nonuniform traffic models: unbalanced [13], Chang's [10], asymmetric [14], and power-of-two (PO2) [11].

The unbalanced traffic model uses a probability,  $w$ , as the fraction of input load directed to a single predetermined

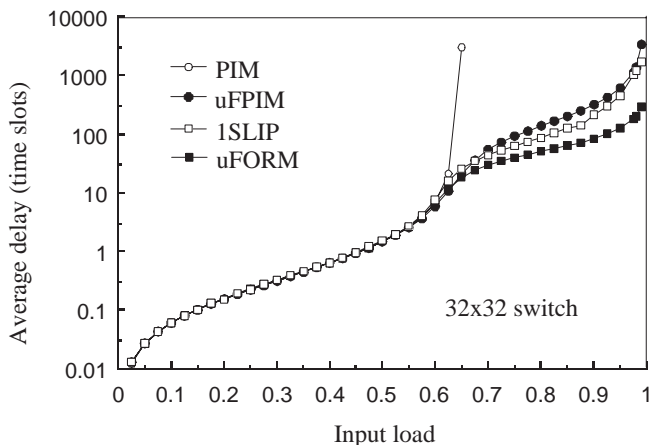


Fig. 1. Average delay under Bernoulli uniform traffic

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  $w = 1$ , it is completely directional, from input  $i$  to output  $j$ , where  $i = j$ . Figure 2 shows the throughput performance of 1SLIP, PIM, uFPIM, and uFORM under unbalanced traffic. This figure shows that uFORM provides over 99% throughput under the complete range of  $w$  and that uFPIM reaches just 99% throughput. The

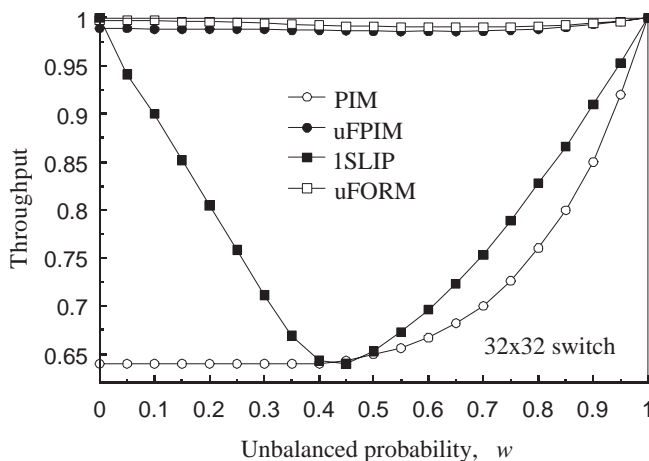


Fig. 2. Throughput performance under unbalanced traffic

high throughput of uFORM under this traffic model is the product of considering the VOQ occupancy. uFORM ensures service to queues with high load by capturing a large frame size for each, and to the queues with low load by using round-robin selection.

Chang's traffic model can be defined as  $\rho_{i,j} = 0$  for  $i = j$  and  $\rho_{i,j} = \frac{1}{N-1}$ , otherwise [10], where  $\rho_{i,j}$  is the input load. Under Chang's traffic, the obtained throughput is 64% by PIM, 97% by 1SLIP, and 99% by uFPIM and uFORM. Under asymmetric traffic, the throughput obtained is 70% by PIM, 71% by 1SLIP, and above 99% by uFPIM and uFORM.

We considered  $16 \times 16$  switches for simulation under PO2 traffic. Under PO2 traffic, the obtained throughput is 68% by PIM, 73% by 1SLIP, and 93% under uFPIM and uFORM.

Although uFPIM and uFORM have lower throughput under PO2 than in the other traffic models, these schemes show performance improvement.

## V. CONCLUSIONS

We introduced the captured frame size concept to determine cell eligibility for matching in input queued packet switches. We also proposed a matching scheme called uFORM. This scheme uses round-robin selection and the concept of captured frame size, where the frame size depends on VOQ occupancy at complete-service time. As the switch size increases, uFORM shows 99% throughput under several nonuniform traffic models, while retaining the high performance of round-robin based schemes under uniform traffic. This matching scheme does not need to compare the status of different VOQs as it is based on simple round-robin. The hardware complexity of uFORM is low, because the hardware additions are a CF counter and a service flag to each VOQ.

This letter also showed the performance improvement with the frame eligibility concept in uFPIM, which is a version of the PIM scheme using this concept. uFPIM shows an improved performance for both uniform and nonuniform traffic patterns.

## REFERENCES

- [1] M. Karol, M. Hluchyj, "Queuing in High-performance Packet-switching," *IEEE J. Select. Area Commun.*, vol. 6, pp. 1587-1597, December 1988.
- [2] 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.
- [3] T.E. Anderson, S.S. Owicki, J.B. Saxe, and C.P. Tacker, "High-speed Switch Scheduling for Local Area Networks," *ACM Trans. on Computer Systems*, vol. 11, no. 4, pp. 319-352, November 1993.
- [4] N. McKeown, "Scheduling algorithms for input-queued cell switches," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California at Berkeley, Berkeley, CA, 1995.
- [5] H.J. Chao, J-S. Park, "Centralized Contention Resolution Schemes for a large-capacity Optical ATM Switch," *IEEE ATM Workshop 1998*, pp. 11-16, May 1998.
- [6] Y. Li, S. Panwar, H.J. Chao, "The Dual Round-robin Matching Switch with Exhaustive Service," *IEEE HPSR 2002*, pp. 58-63, May 2002.
- [7] Y. Jiang and M. Hamdi, "A fully Desynchronized Round-robin Matching Scheduler for a VOQ Packet Switch Architecture," *IEEE HPSR 2001*, pp. 407-411, May 2001.
- [8] Y. Jiang and M. Hamdi, "A 2-stage Matching Scheduler for a VOQ Packet Switch Architecture," *IEEE ICC 2002*, vol. 4, pp. 2105-2110, May 2002.
- [9] D. N. Serpanos and P. I. Antoniadis, "FIRM: A Class of Distributed Scheduling Algorithm for High-speed ATM Switch with Multiple Input Queues," *INFOCOM 2000*, vol. 2, pp. 548-555, March 2000.
- [10] C-S. Chang, D-S. Lee, and Y-S. Jou, "Load Balanced Birkhoff-von Newman Switches," *IEEE HPSR 2001*, pp. 276-280, May 2001.
- [11] A. Bianco, M. Franceschinis, S. Ghisolfi, A.M. Hill, E. Leonardi, F. Neri, R. Webb, "Frame-based Matching Algorithms for Input-queued Switches," *IEEE HPSR 2002*, pp. 69-76, May 2002.
- [12] S. Li and N. Ansari, "Input-queuing Switching with QOS Guarantees," *IEEE INFOCOM 1999*, vol.3, pp. 1152-1159, March 1999.
- [13] R. Rojas-Cessa, E. Oki, Z. Jing, and H. J. Chao, "CIXB-1: Combined Input-One-cell-crosspoint Buffered Switch," *IEEE HPSR 2001*, pp. 324-329, May 2001.
- [14] R. Schoenen, G. Post and G. Sander, "Weighted Arbitration Algorithms with Priorities for Input-Queued Switches with 100% Throughput," *Broadband Switching Symposium '99*, 1999. <http://www.schoenen-service.de/assets/papers/Schoenen99bssw.pdf>