Exhaustive Output Arbitration of Input Buffered Switch with Buffered Crossbar

Bin Yeong Yoon, Man-Soo Han, Heyung Sub Lee,
Bongtae Kim, and Whan-Woo Kim

ABSTRACT— We propose a new arbitration method for an input buffered switch with a buffered crossbar. In the proposed method, an exhaustive polling method is used to decrease the synchronization. Using an approximate analysis, we explain how the proposed method improves the switch performance. Also, using computer simulations, we show the proposed method outperforms the previous methods under burst traffic.

Keywords—Input buffered switch, buffered crossbar, scheduling.

I. Introduction

A crossbar switch is very popular for high-speed switching because of its simplicity. Several types of crossbar switches with a virtual output queue (VOQ) [1]-[4] have been studied based on paralleled and iterative request-grant-accept cycles between input and output ports, or so called maximal size matching. To improve the performance of the crossbar switch with a VOQ, a buffered crossbar switch [5], [6] was introduced where a buffer exists in each crosspoint. Thanks to the crosspoint buffer, the switching performance was dramatically improved compared to the non-buffered crossbar switch. However, the starting point of the input arbitrations in these conventional methods [5], [6] is updated based on the result of selection, which causes synchronization to degrade the switching performance. In order to deal with the problem, an input arbitration method [7] was suggested in which two starting points of input arbitration are never the same at any moment. But the method in [7] uses a conventional round robin for output arbitration that can still yield the synchronization.

In this paper, we will address the synchronization phenomenon in the existing input arbitration technique and approximate the synchronization probability. To decrease the synchronization probability, we propose an exhaustive output arbitration method that is better than the existing methods in terms of mean delay and cell delay variance under heavy traffic.

This paper is organized as follows. In section II of this paper, we describe an input buffered switch with a buffered crossbar and propose a scheduling algorithm called exhaustive output arbitration to decrease the synchronization phenomenon. In section III, using computer simulations under uniform and on-off traffic, we validate that the proposed method is better than the conventional arbitrations in regard to the mean delay and cell delay variance.

II. Exhaustive Output Arbitration

Figure 1 shows an input buffered switch with N input modules and an $N \times N$ buffered crossbar, where N is the number of input and output ports. Each input module contains N VOQs. A VOQ at input module i that stores cells destined for output j is denoted as q_{ij}. A crosspoint buffer of the buffered crossbar that connects input module i to output j is denoted as b_{ij}. The larger the size of b_{ij}, the more improved the performance becomes. For implementation feasibility, however, we assume the size of b_{ij} is very small yet greater than 1. A flow control mechanism tells the input module whether or not each b_{ij} is full. VOQ q_{ij} is said to be eligible when q_{ij} is not empty and b_{ij} is not full. In the input arbitration, only an eligible

Manuscript received May 03, 2004; revised June 29, 2004.

This work is supported in part by the Ministry of Information and Communication of Korea under the title of "Optical Subscriber & Access Network Technology."

Bin Yeong Yoon (phone: +82 42 860 6859, email: byyun@etri.re.kr), Heyung Sub Lee (email: leehs@etri.re.kr), and Bongtae Kim (email: bkim@etri.re.kr) are with Broadband Convergence Network Research Division, ETRI, Daejeon, Korea.

Man-Soo Han (email: mshan@mokpo.ac.kr) is with the Division of Information Engineering, Mokpo National University, Chonnam, Korea.

Whan-Woo Kim (email: wwkim@cnu.ac.kr) is with the Division of Electrical and Computer Engineering, Chungnam National University, Daejeon, Korea.

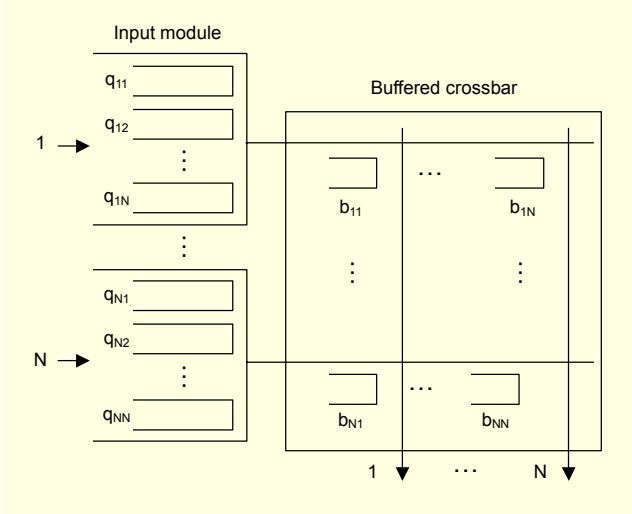

Fig. 1. Input buffered switch with buffered crossbars.

q_{ij} can send a cell to b_{ij}. Also, in the output arbitration, only a nonempty b_{ij} can send a cell to the output.

In [7], a synchronization phenomenon was studied where more than one input module selects cells destined for the same output. Since those cells compete for the same output and only one cell can depart from an output, the synchronization phenomenon degrades the switching performance. In the input arbitration, the destinations of the selected cells should be different from each other for a high performance. To alleviate the synchronization effect, M. S. Han and others [7] suggested a new input arbitration method where the round-robin starting points of input modules are different from each other in every case.

Here, we describe the more detailed input arbitration method of [7]. Let a_i be the starting point of input module i. Then the initial value of a_i is given by,

$$a_i = i, \ \ i = 1,2,\cdots,N. \tag{1}$$

At the end of each time slot, every starting point a_i is increased (modulo N) by 1 independently of the arbitration result. Input module i selects the first eligible VOQ from starting point a_i in a round-robin fashion.

Also, we explain the detailed conventional output arbitration method that is used in [7]. Let g_j be the starting point of output j. Output j selects the first nonempty buffer from starting point g_j in a round-robin fashion. If b_{ij} is selected, then $g_j = i+1$ at the next time slot. In this paper, we suggest a new output arbitration method based on the above output arbitration method to improve the switching performance.

Although the input arbitration method of [7] mitigates the synchronization phenomenon, it still can occur. For example, suppose $a_1 = 1$, $a_2 = 2$ and that q_{11} is not eligible but q_{12} and q_{22} are eligible. Then, input modules 1 and 2 choose the head of line cells of q_{12} and q_{22}, respectively. The head of line cells should compete for output 2. To analyze the synchronization effect of the input arbitration method, we study a probability where more than one input module selects cells destined for the same output. Suppose that a VOQ is eligible with probability e but is not with probability $d = 1-e$. Without a loss of generality, we assume $a_i = i$. Let P_{ij} be the probability that input module i selects a cell destined for output j.

Input module i selects a cell destined for output j when $q_{ii}, q_{i(i+1)}, \cdots, q_{i(j-2)}, q_{i(j-1)}$ are not eligible but q_{ij} is eligible, where $i \le j$. That is,

$$P_{ij} = d^{j-i}e, \ \ i \le j. \tag{2}$$

We consider four consecutive input modules. Without a loss of generality, suppose the consecutive input modules are input modules 1, 2, 3 and 4. Let S be the synchronization probability that more than one of the consecutive input modules will select cells destined for the same output. Since the arbitrations of input modules are performed independently from each other, S is given by

$$\begin{aligned} S = {} & P_{12}P_{22} + P_{13}P_{23} + P_{13}(1-P_{23})P_{33} + P_{23}(1-P_{13})P_{33} \\ & + P_{14}P_{24} + P_{14}(1-P_{24})P_{34} + P_{14}P_{44}(1-P_{23}P_{33} - P_{24}P_{34}) \\ & + P_{24}P_{34}(1-P_{14}P_{44}) + P_{24}P_{44}(1-P_{13}P_{33} - P_{14}P_{34}) \\ & + P_{34}P_{44}(1-P_{12}P_{22} - P_{13}P_{23} - P_{14}P_{24}) + \cdots. \end{aligned} \tag{3}$$

For simplicity, we consider only heavy traffic since the scheduling algorithm significantly affects the switch performance under that traffic.

Recall that VOQ q_{ij} is eligible when q_{ij} is nonempty and b_{ij} is not full. Hence, a cell can depart only from an eligible q_{ij}. It is well known that an arrival rate is equal to a departure rate in the steady state of a stable queuing system. That means, as the input traffic load increases, the departure rate of q_{ij} increases, i.e., the eligible probability of q_{ij} increases. Therefore, we assume that eligible ability e is not too small under heavy traffic.

Because of (2), and since e is not too small, we approximate (3) as follows:

$$S^* = e^2 d(3 + d + 3d^2 - d^3 + 7d^4 - 8d^5 + 9d^6 - 4d^7). \tag{4}$$

As we can see from Fig. 2, from a peak at around e=0.58, probability S^* decreases as e decreases. Under heavy traffic, most of the VOQs might be occupied by cells. On the other hand, since the size of the crosspoint buffers is small, the full probability of the crosspoint buffers might be large under heavy traffic. Hence, eligible probability e can't be large under heavy traffic. This implies that we need to decrease e in order

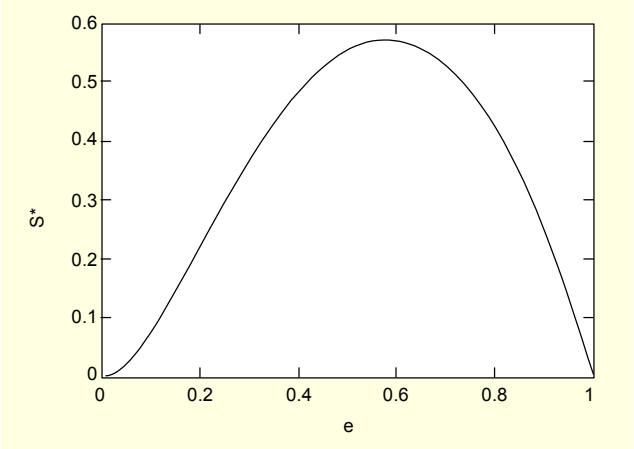

Fig. 2. Approximate synchronization probability.

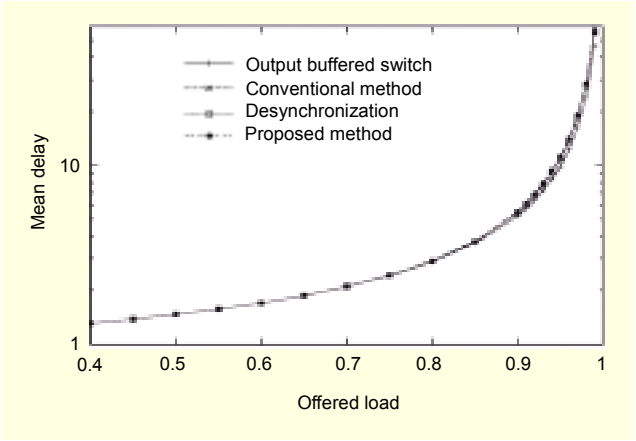

Fig. 3. Mean delay (time slot) under uniform traffic.

to decrease probability S^*. If the full probability of b_{ij} increases, e decreases and thereby S^*decreases.

To increase the full probability of b_{ij}, we can use an exhaustive service discipline. In the exhaustive service, output j serves b_{ij} as long as there are cells to be transmitted and can move on only when b_{ij} is empty. The exhaustive service increases the polling cycle time compared to the limited service that the conventional output arbitration method belongs to. The polling cycle time is the time to serve all nonempty buffers once in a polling system [9]. If the polling cycle time is increased, a buffer has to wait more time to be served. Thus, the full probability of a buffer is increased under the exhaustive service. Also, since burst traffic makes a buffer heavily loaded, the exhaustive service manner should be more effective for burst traffic.

We modify the conventional output arbitration method to satisfy the exhaustive service discipline. Output j selects the first nonempty buffer from starting point g_j in a round-robin manner. If b_{ij} is selected and has more than one cell, then $g_j = i$ at the next time slot. If b_{ij} is selected and has only one cell, then $g_j = i + 1$ at the next time slot.

III. Simulation

Under uniform and burst traffic, we compare the mean delay of the proposed method and those of the previous methods [5]-[8]. The uniform traffic is made up of independent, identically distributed Bernoulli arrivals with destinations uniformly distributed over all outputs. The burst traffic is the two-state on-off traffic as described in [7]. We use a 16×16 switch in which the size of b_{ij} is 2 for the simulation. The simulation time is 10^7 time slots for each plot point.

Figure 3 depicts the mean delay of various methods under the uniform traffic where the conventional method indicates the method shown in [5], [6], and [8], and the desynchronization is the method demonstrated in [7]. Figure 3 shows the performance

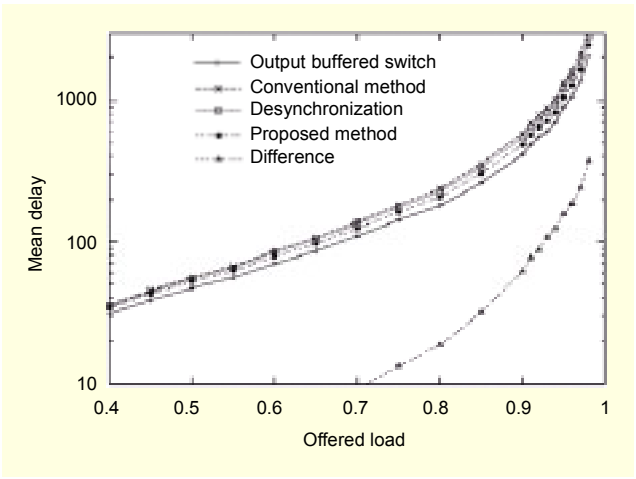

Fig. 4. Mean delay (time slot) under burst traffic.

of our method is almost the same as those of the previous methods under uniform traffic.

We now consider burst traffic with mean burst length $L = 50$. Figure 4 shows our method is better than the previous methods under burst traffic. In Fig. 4, 'difference' means the difference between the mean delays of the proposed and desynchronization methods. Notice that when the offered load is greater than 80%, the mean delay is improved by more than 5% when compared to the desynchronization method. Near the full offered load, the mean delay is improved by more than 10%.

In addition to the mean delay, another important parameter of switching systems is a cell delay variance. Although the proposed method outperforms the existing methods with regard to the mean delay, it increases the cell delay variance using the increase of the mean polling cycle time. Figure 5 illustrates the mean polling cycle time for each method under burst traffic, which shows that the proposed method is the worst of the methods.

Figure 6 shows the cell delay variance for each method

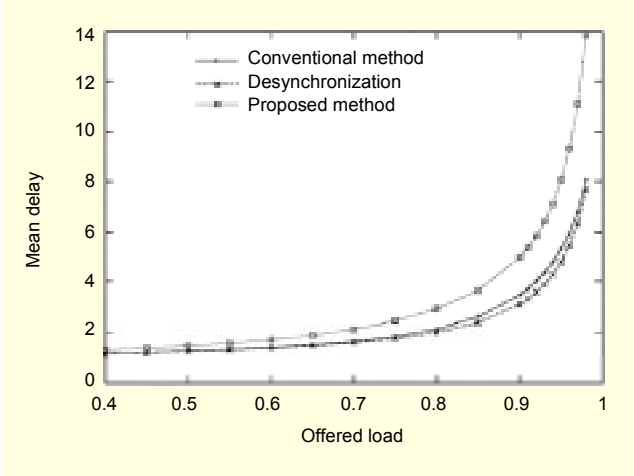

Fig. 5. Mean polling cycle time under burst traffic.

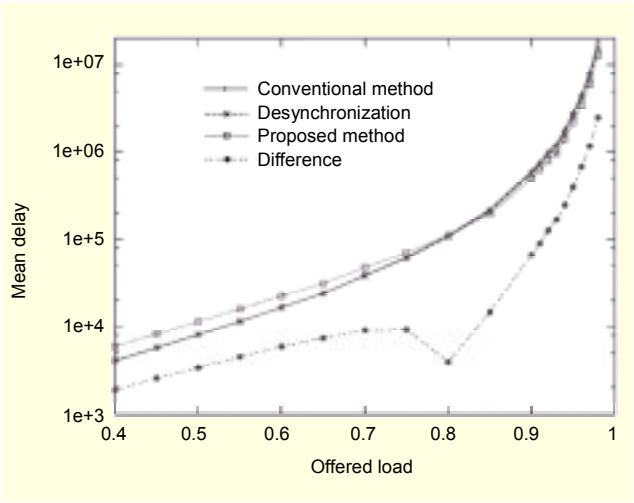

Fig. 6. Cell delay variance time under burst traffic.

where the 'difference' is the absolute value of the difference between the proposed and desynchronization methods in the cell delay variance. When the traffic load is less than 85%, the desynchronization methods are more efficient than the proposed method for the cell delay variance owing to the short mean polling cycle. However, this shows that the proposed method is better than the existing methods under heavy traffic. This means that there are two major factors that affect the cell delay variance: the synchronization effect and polling time. The synchronization effect deteriorates the cell delay variance more severely than polling time under heavy traffic.

IV. Conclusion

Synchronization in the input buffered crossbar switch with the crosspoint buffer increases the mean delay under heavy traffic. In this paper, we obtained the approximate synchronization probability that more than one input module simultaneously transmits cells to the same output. We showed that the synchronization effect is decreased when the crosspoint buffers are heavily occupied under heavy traffic. Based on this, we proposed the exhaustive output arbitration to increase the occupation probability of the crosspoint buffers.

The proposed exhaustive output arbitration selects the first nonempty buffer from a starting point. Then, the starting point can move on only when the selected buffer is empty, which increases the number of cells in the crossbar buffers. Using an approximate analysis, we showed that the proposed method improves the mean delay.

References

[1] N. McKeown, "The Islip Scheduling Algorithm for Input Queued Switches," *IEEE/ACM Trans. Network*, vol. 7, no. 2, Apr. 1999, pp. 188-201.

[2] T.E. Anderson, S.S. Owicki, J.B.Saxe, and C.P. Thacker, "High-Speed Switch Scheduling for Local-Area Network," *ACM Trans. Computer Syst.*, vol. 11, no. 4, Nov. 1993, pp. 319-352.

[3] Yihan Li, Shivendra Panwar, and H. Jonathan Chao, "The Dual Round-Robin Matching Switch with Exhaustive Service," *2002 Workshop High Performance Switching and Routing (HPSR 2002)*, May 2002, pp. 58-63.

[4] H.S. Lee, U.G Joo, H.H. Lee, and W.W. Kim, "Optimal Time Slot Assignment Algorithm for Combined Unicast and Multicast Packets," *ETRI J.*, vol. 24, no, 2, Apr. 2002, pp. 172-175.

[5 R. Rojas-Cessa, E. Oki, Zhigang Jing, and H.J. Chao, "CIXB-1: Combined Input One-Cell-Crosspoint Buffered Switch," *IEEE Workshop on High Performance Switching and Routing*, 2001, pp. 324-329.

[6] K. Yoshigoe and K.J. Christensen, "A Parallel-Polled Virtual Output Queued Switch with a Buffered Crossbar," *IEEE Workshop High Performance Switching and Routing*, 2001, pp. 271-275.

[7] M.S. Han, D.Y. Kwak, and B. Kim, "Desynchronized Input Buffered Switch with Buffered Crossbar," *IEICE Trans. Commun.*, vol. E86-B, no. 7, July 2003, pp. 2216-2219.

[8] R. Rojas-Cessa and E. Oki, "Round-Robin Selection with Adaptable-Size Frame in a Combined Input-Crosspoint Buffered Switch," *IEEE Commun. Lett.*, vol. 7, no. 11, Nov. 2003, pp.555-557.

[9] H. Takagi, "Queuing Analysis of Polling Models," *ACM Computing Surveys*, vol. 20, no. 1, 1988, pp. 5-28.