PERFORMANCE ANALYSIS OF CELL AND PACKET SWITCH WITH FINITE INPUT QUEUES AND ILPF SCHEDULING ALGORITHM

Igor Radusinovic[1], IEEE Student Member, Zoran Petrovic[2], IEEE Member
[1] Department of EE, University of Montenegro, Cetinjski put bb, 81000 Podgorica, Serbia&Montengro, igor.r@cg.yu
[2] Department of EE, University of Belgrade, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia&Montengro

***Abstract*- In this paper, packet loss probability performance in cell-based and packet-based switches with finite Input Queues (IQ) and iLPF scheduling algorithm are analysed. This algorithm has been proposed as a practical scheduling algorithm to achieve 100% throughput in cell-based switches. The scheduling algorithm is implemented in three cases: cell-based (CB-iLPF) switching, packet-based (PB-iLPF) switching and "waiting" packet-based (WPB-iLPF) switching. We investigate possible performance penalties dealing with variable-size data units and packet-based scheduling algorithms. We made performance comparison of considered IQ switches under input bursty traffics scenario. It is shown that WPB-iLPF switch performances are very sensitive to "waiting" period duration and that for input load smaller than 0.75, cell-based crossbar switch introduces packet loss probability greater than introduced by the packet-based switches.**

Keywords- cell-based switching, packet-based switching, input-queued switch, virtual output queueing, average packet delay, average input buffer size.

I. INTRODUCTION

Input Queueing (IQ) switches have become a dominant solution for the design of modern high capacity packet switches with many ports and fast line rates [1-2], as for example Cisco 12000, Lucent Cajun and Nortel Versalar TSR45000. The design of these high-performance switches is based on input queuing with a crossbar switch fabric, virtual output queues (VOQ) and fixed size cells. A crossbar switching fabric is very popular solution because under current technology, the switch matrix can support up to 10 Gbps line rate [3]. The use of virtual output queueing (in each input interface card, input buffers are organized into a set of queues, each queue storing data directed toward a specific output interface) avoids performance degradation due to the head-of-line (HOL) blocking [4]. At each time slot, a scheduler finds a matching between the inputs and outputs, and configures the crossbar according to this matching. A lot of algorithms have been proposed for finding this matching [4-7], but the need for switches with more ports and faster line rates makes these algorithms difficult to scale. Using fixed-size cells (cells from ATM terminology) in the switch makes the implementation of the scheduling algorithm much easier compared to the using variable-length packets. In other words, whenever a variable-length packet (for example IP datagram) arrives at the switch, it is divided into cells. After the switching is done, the cells are re-assembled in the form of the original variable-length packet.

There are three main problems with this approach:

1) This procedure induces significant implementation overhead.
2) A generation of incomplete cells (when size of packet is not multiple of cell size).
3) Even with fast switch matrix and memories, it may not always be possible to run cell-based scheduling algorithm fast enough.

Packet-based switching [1], [8], is one of the solutions that are proposed in literature in order to solve this disadvantages. Contrary to the cell-based switching, the packet-based switching is implemented in switch in which system directly works on packets without breaking them into cells. The two important design criteria for switching architectures have been considered: (i) throughput and (ii) average delay. Regarding on throughput, stability of matching algorithms is analyzed. It is well known that cell-based Maximal Weight Matching (MWM) algorithms are stable for any admissible traffic. According to [1], packet-based Maximal Weight Matching (PB-MWM) algorithm (canonical modification of cell-based MWM) achieve 100% throughput for any admissible Bernoulli i.i.d. traffic with packet lengths being bounded. In study [8] there is a prove that PB-MWM is stable even for any form of re-generative admissible traffic, with the time of regeneration being finite in mean. But, the same authors have found that for PB-MWM algorithms there exists a counter example (belonging to the general admissible traffic with Strong Law of Large Number Properties) for which PB-MWM is not stable. It has been shown that "waiting" modification of PB-MWM is stable for any admissible traffic with bounded packet lengths. The average time delay induced with different cell-based and packet based (their canonical extension) have been analyzed in [1], too. The results that have been presented there show that relative delays depend on the traffic characteristics.

In our previous paper [9], we considered well-known iterative Long Port First (iLPF) scheduling algorithm [10] (which belongs to MWM family) but in packet-based switches for the first time. Great effort had been done in closed-form analytical analysis of virtual output queueing systems for nonrealistic uniform input traffic with uniform and non-uniform cell destination. Therefore, in realistic case only way for obtaining right picture about switch performances are simulation study with ON-OFF input traffic models as an approximation of real traffic. The performances of the switches were investigated by simulated operation under ON-OFF traffics for different output port destination distributions. In this paper, we present further enhancement of the performance consideration in [9]. We consider packet loss probability in packet-based switches with finite input queues and iLPF scheduling algorithm. We use a simulation study with ON-OFF input traffic model as an approximation of real traffic. The packet loss probability is investigated by simulated operation under ON-OFF traffics with Zipf output port destination distribution [11]. For the first time in literature, we investigate possible performance penalties that use of variable-size data units and PB-MWM can bring in spite of packet loss probability.

The paper is organized as follows. Section II describes switching procedures for CB-iLPF, PB-iLPF and WPB-iLPF switches. Also, there is a overview of traffic model and performance definitions. Section III contains simulation results WPB-iLPF switch. Section IV presents performance comparison of switches under consideration. Finally, Section V concludes the paper.

II. SWITCH MODELS

A. CB-iLPF switch

Figure 1. shows the logical structure of an IQ switch. We assume that the switch has the same number of input and output ports denoted by N.

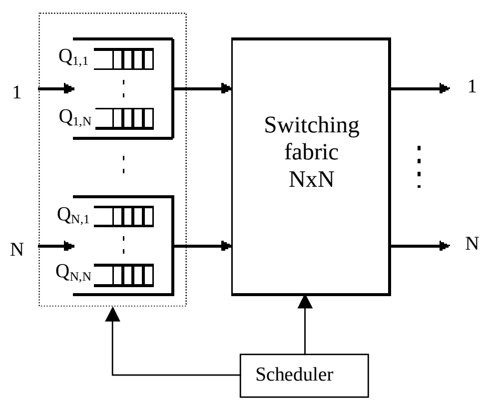

Fig. 1. Structure for an IQ switch with VOQs

We assume that the time is slotted and at each time slot, at most one cell can arrive to each input port. Cells arriving at input *i* and destined for output *j* are stored at input of a finite FIFO buffer called virtual output queue that is denoted by VOQ_{ij}. This procedure avoids the head-of-line problem. At each time slot, a scheduler finds a matching between the inputs and outputs, and configures the memory-less crossbar according to this matching. iLPF, (Fig. 2.) is well known MWM matching algorithm that has been developed for cell based switch. iLPF has a running-time complexity of $O(N^{2.5})$ [9] (or $O(N^3)$ [12]) and therefore it is very interesting solution for practical implementation. In fact, iLPF algorithm is modified maximum size algorithm. Each iLPF request weight is defined as the sum of the input and output occupancies. This sum is known as port occupancy. iLPF favors queues with high port occupancies. According to [9], iLPF can achieve 100% throughput for all admissible traffics with independent arrivals. With slide modifications of remarks from [1] it is easy to prove that this algorithm for cell-based switch (CB-iLPF) is stable for any admissible traffic satisfying the strong law of large numbers. More details can be found in [9].

B. PB-iLPF switch

PB-iLPF switch architecture is a modification of previous switch [1]. We assume that variable-length packets arrive to switch. The switching fabric works on cells. The received packets are segmented (without loss of generality) in an integer number of cells. Similarly like previous, in each time slot only one cell can be sent to each output. PB-iLPF switch is based on scheduling algorithm such that once it starts transmitting the first cell of a packet to an output port, it continuous the transmission until the whole packet is completely received at a corresponding output port. This avoids the problem of segmentation of packets and reassembly of cells in the system. However, in any cell-based switches different cells of the same packet observe different delay values. The cell-based scheduling is not aware of the existence of packets, and therefore there is a chance that packet scheduling (which is aware of the existence of packets) can be better in the sense of the delay observed by user. The packet-based algorithm makes distinction between:

1. Busy ports: the set of input-output ports that have been matched in the previous time slot and are still in the middle of a sending a packet.
2. Free ports: the set of input-output ports that either have no packet to send, or just finished sending a packet.

PB-iLPF keeps the matching already used by busy ports and finds a new matching for free ports using the CB-iLPF. We assume that all the ports are free at start. According to [1], any PB-MWM algorithm (this means PB-iLPF too) can achieve 100% throughput for regenerative admissible input traffic. But, there are the admissible traffic patterns for which PB-MWM algorithm is unstable. More details can be found in [1].

C. WPB-iLPF switch

To overcome problem with instability of any PB-MWM (this means PB-iLPF too) for some admissible traffic patterns a "waiting" procedure is proposed in [8]. During "waiting" period a packet gets served waits until all ports become free and schedule according, for example, to PB-iLPF. Switch based on this procedure we denoted WPB-iLPF switch. According to [8], this PB-iLPF is stable for any admissible traffic for arbitrarily small "waiting" periods. The proof was based on fluid model [13].

The performances of the switches are investigated by simulated operation under ON-OFF traffic. It means that our simulation models do not explicitly describe the arrival of IP datagrams at the packet-switch inputs. Instead of that we model the arrival of cell bursts at the inputs of the internal cell-switch. These cell bursts are assumed to originate from the segmentation of a packet. The simulated switching size is 16x16 and the random access buffers are considered with finite sizes. The traffic at each input is modeled as Interrupted Bernoulli Process (IBP) [14]. We consider Zipf traffic scenario as follows. Real-life traffic destinations are not uniformly distributed; traffic tends to be focused on preferred or popular destinations. Therefore in this scenario, we used the Zipf distribution to model burst destination distribution. The same ON-OFF model describes all input traffics where both busy and idle periods are geometrically distributed. The probability that an arriving burst is heading towards destination i is given by:

$$Zipf\,(i) = \frac{i^{-k}}{\sum_{j=1}^{N} j^{-k}}$$

where i is the burst destination, k is the Zipf order, and N is the number of switch ports. As obvious k=0 corresponds to uniform distribution. We used k=1 because many natural and human phenomena obey the Zipf law with k close to 1 [10]. We assigned each port with different cyclic shift of the distribution so that aggregate offered load from all input ports to an output port did not exceed 1. Packet lengths are uniformly distributed with the lengths between 1 and 192 cells.

We assume that packet will be lost if its first cell finds that corresponding VOQ has not enough places for all cells from fragmented packet. We define packet loss probability as the difference between incoming and outgoing (and queued) packets over the number of incoming packets during a representative time**.**

III. PERFORMANCE SIMULATION

In this section, the packet loss probability performances of WPB-iLPF switch are performed for different “waiting” periods (10%, 1% and 0.1% of time). Apart from that, the packet loss probability performance of IQ switch obtained by simulations, with various considered cell-based or packet-based scheduling algorithms are presented.

Fig. 3. shows the simulated packet loss probability results vs. input load p for the considered switches and different buffer lengths. All considered various scheduling algorithms produce slightly larger packet than CB-iLPF switch. It could be seen that WPB-iLPF-10% generates larger packet loss probabilities than other switches for any input load. Also, WPB-iLPF-1% and WPB-iLPF-0.1% have almost same performance no matter on input load. It means that decreasing of waiting periods don’t bring any packet loss probability improvement. Their performances are slightly bellowing PB-iLPF performances. But, it is important to notice that all considered IQ switches are very sensitive on buffer lengths.

IV. CONCLUSIONS

In this paper, we consider well-known iterative Long Port First (iLPF) scheduling algorithm [10] but in packet-based switches with finite buffers. Simulation results and performance comparisons of IQ switch using various proposed cell-based and packet-based switches scheduling algorithms are presented.

It is shown that all packet-based switches have quite a close packet loss probabilities performance for small “waiting” periods (1% and 0.1%). It means that the duration of implemented “waiting” procedure, from the point of packet loss probabilities is very relevant. It is proved that solution with “waiting” period equals to only 10% has performance significantly bellow ordinary all IQ switches that we considered in this paper. Therefore, determination of right value of “waiting” period from [9] point of view and our results is of extraordinary importance. It must be some kind of compromise between throughput stability, average packet delay and packet loss probability under traffic presented in [9].

Apart from that, it is concluded that all considered combinations of various scheduling algorithms produce bigger packet loss probabilities than CB-iLPF switch. We obtained that PB-iLPF and WPB-iLPF-1% have similar packet loss probabilities no matter on input load for buffer sizes bellow 700 cells.

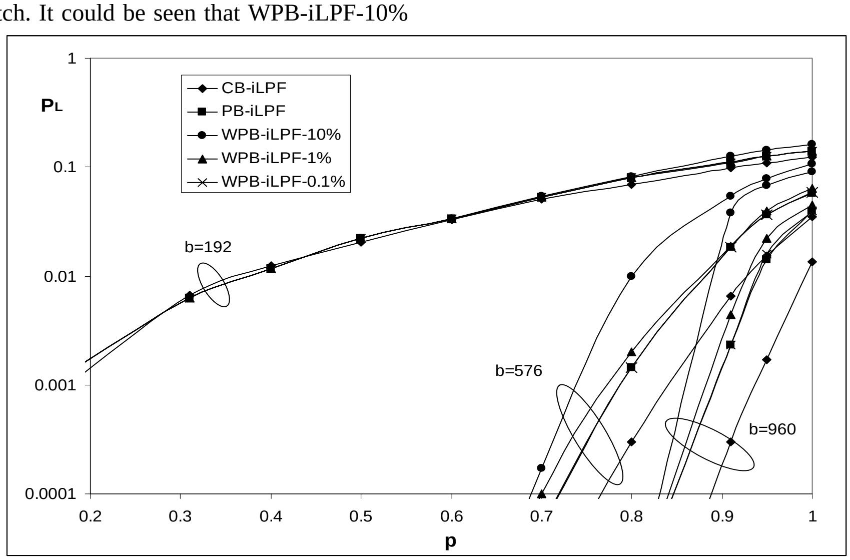

Fig. 3. Packet loss probabilities over input load for different buffer lengths

Taking into account other solutions for problems with cell-based scheduling approach for variable-length packet switching as well as IP datagrams arrival simulation, further enhancement of the presented performance consideration is already under our attention. Also, our new model and simulation results for determining instability regions and packet loss probabilities that various packet-based scheduling algorithms and different packet dropping procedures produce will be presented very soon.

References

[1] M. Ajmone Marsan, A. Bianco, P. Giaccone, E. Leonardi, F.Neri, "Packet Scheduling in Input-queued Cell-Based Switches," Proc. *IEEE INFOCOM*, vol. 2, 2001.

[2] E. Leonardi, M. Mellia, F.Neri, M. Ajmone Marsan, "Bounds on Avearge Dealys and Queue Size Averages and Variances in Input-Queued Cell-Base Switches," Proc. *IEEE INFOCOM*, vol. 2, 2001, pp. 1095--1103.

[3] Vitesse, Switch fabric products [Online], available at http://www.vitesse.com/, 2003.

[4] N. McKeown, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," *Proceedings of IEEE INFOCOM '96*, pp.296-302, 1996.

[5] N. McKeown, "The iSLIP scheduling algorithm for input-queued," *IEEE Transactions on Networking*, Vol.7, n.2, Apr.1999, pp.188-201.

[6] M. Ajmone Marsan, A. Bianco, E. Leonardi, L. Milia, "RPA: a flexibile scheduling algorithm for input-buffered switches," *IEEE Transactions on Communications*, vol. 47,n.12, December 1999, pp. 1921--1933.

[7] H.Duan, J.W.Lockwood, S. Mo Kang, "Matrix Unit Cell Scheduler (MUCS) for input-buffered ATM switches," *IEEE Communications Letters*, vol. 2, n.1, January 1998, pp. 20-23.

[8] Y. Ganjali, A. Keshavarzian, D. Shah, "Input Queued Switches: Cell Switching vs. Packet Switching," *Proceedings of IEEE INFOCOM 2003*, 2003.

[9] I. Radusinovic, Z. Petrovic, M. Pejanovic, Z. Veljovic, "Performance analysis of cell-based and packet-based input-queued switch with iLPF scheduling algorithm", *accepted for International Conference on Networking 2004*, Pointe-à-Pitre, Guadeloupe, French Caribbean, 2004.

[10] A. Mekkittikul, and N. McKeown, "A Practical Schedulng Algorithm to Achieve 100% throughput in an input-queued switch," *Proceedings of IEEE INFOCOM '98*, 1998.

[11] I. Elhanany, and D. Sadot, "A contention-free Tbit/sec packet-switches with virtual output queueing" IEICE Trans. Comm., Vol.E83-B, no.2, Feb. 2000.

[12] Y. Ganjali, A. Keshavarzian, and D. Shah, "An Efficient Scheduling Algorithm for CIOQ Switches with Space-Division Multiplexing Expansion," *Proceedings of IEEE INFOCOM 2003*, 2003.

[13] J.G.Dai, and B.Prabhakar, "The Throughput of Data Switches with and without speedup," *Proceedings of IEEE INFOCOM 2000*, March 2000, pp. 556-564.

[14] I. Radusinovic, M. Pejanovic, and Z. Petrovic, "Impact of scheduling algorithms on performance of buffered crossbar switch fabrics," *Proceedings of IEEE ICC 2002*, June 2002, pp. 2038-2042.