Impact of Unequal Buffer Lengths on DCQ Switch QoS-Awareness

Nebojsa Maletic, Gordana Gardasevic
University of Banja Luka
Faculty of Electrical Engineering
Banja Luka, Bosnia and Herzegovina
{nebojsa.maletic, gordana.gardasevic}@etfbl.net

Igor Radusinovic, Milutin Radonjic
University of Montenegro
Faculty of Electrical Engineering
Podgorica, Montenegro
{igorr, mico}@ac.me

Abstract— **In our previous work, we proposed the Crosspoint Queued (CQ) crossbar switch with two buffers at each crosspoint of crossbar fabric, called the Dual Crosspoint Queued (DCQ), to support the traffic differentiation and QoS-awareness. One buffer is dedicated to high priority traffic and the other one to low priority traffic with equal buffer lengths. The performance analysis has shown that static priority round robin based scheduling algorithms can provide the QoS only for high priority traffic due to their greedy nature. In this paper, we extend our performance analysis to DCQ crossbar switch with unequal buffer lengths, thus supporting the QoS-awareness. Additionally, we assume that the share of high priority traffic within incoming traffic is lower than the low priority traffic.**

Keywords-DCQ-switch; TDRR algorithm; QoS; buffer length

I. INTRODUCTION

The new generation of telecommunication networks and services brings several novel paradigms such as "emergent technologies, all-connected objects and things, converged networks and applications", etc. As a consequence, a dramatic increase in data generation and transferring requests efficient mechanisms for an overall Quality of Service (QoS) support. Obviously, this is not a simple task since heterogeneous data sources and diverse applications have multiple QoS constraints and requirements.

In recent years, the problem of QoS assurance in high-speed data switching has been particularly emphasized. This has resulted in creation of various crossbar-based switching architectures. Some of the most common architectures are Input Queued (IQ) [1], Output Queued (OQ) [2], Combined Input Output Queued (CIOQ) [3], Combined Input Crosspoint Queued (CICQ) [4], and Crosspoint Queued (CQ) [5, 6].

The CQ architecture consists of buffers placed only at the crosspoints of the crossbar fabric. The advantage of this architecture is its implementation simplicity and avoidance of common problems such as complex scheduling, head-of-line (HOL) blocking, etc.

The QoS support in these systems is based on providing specific priority mechanisms [7, 8, 9]. The actual network traffic posses certain properties, such as self-similarity and long-range dependence (LRD) [10]. These properties are in close relation to queuing performances. Additionally, the packet burstiness is one of the parameters with key impact on overall switching performances.

The problem of providing more advanced QoS-awareness mechanism has motivated us to perform research on different scheduling mechanisms that can support required traffic differentiation [11-15]. The main idea is to search for efficient use of crosspoint buffers thus reducing the packet drops and large delays.

The rest of paper is organized as follows. Section II presents the description of DCQ switch architecture and the proposed Threshold Dual Round Robin (TDRR) scheduling algorithm. In Section III, we present results of extensive simulations for performance evaluation purposes. The performance evaluation is based on mean delay and cell loss probability. The paper is concluded with the final remarks in Section IV.

II. DCQ SWITCH AND SCHEDULING ALGORITHMS

A. DCQ Switch Architecture

The Dual Crosspoint Queued (DCQ) crossbar switch architecture was proposed in [11]. In Fig.1, the principal DCQ organization is depicted. Each crosspoint of a DCQ switch contains two buffers, in contrast to the CQ switch where one buffer is accommodated at each crosspoint. DCQ switch consists of N inputs, N outputs, and $2N^2$ crosspoint buffers. In order to support the differentiation of incoming traffic on real-time (e.g. live audio and video streaming) and non-real-time (e.g. file transfer), the buffer that accommodates the real-time

traffic is denoted as high-priority (HP) buffer, and another one that accommodates the non-real time traffic is denoted as low-priority (LP). The HP traffic cells are stored in A buffers, while the LP traffic cells are stored in B buffers.

Packets are segmented into cells of fixed size lengths. Cells arriving from input *i* and addressed to output *j* will be forwarded directly to one of the crosspoint buffers ($A_{i,j}$ or $B_{i,j}$), depending on the traffic type. The cell will be discarded if the destination buffer is full, otherwise, the cell will be queued. In each time-slot, the output scheduler selects one of the non-empty CQ buffers from the output line and forwards the HOL cells to an output line-card.

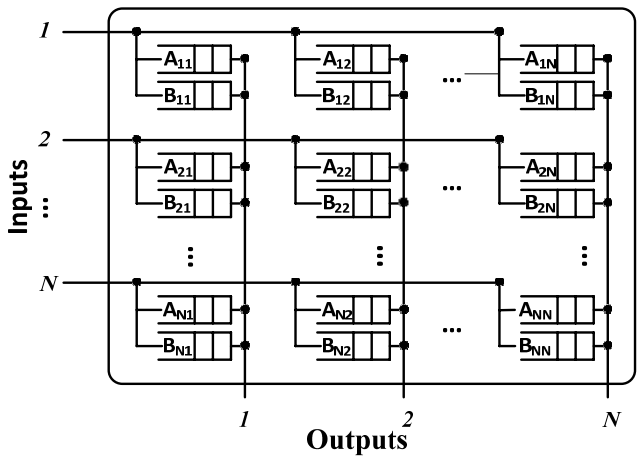

Fig. 1 Dual Crosspoint Queued (DCQ) switch architecture.

B. Description of Proposed Algorithm

In this paper, we have analyzed several variants of round robin based scheduling algorithms, namely round robin (RR), dual round robin (DRR), and priority frame dual round robin (PFRR) algorithm. Detailed descriptions of these algorithms that are easy for implementation can be found in [12].

In order to lower the greedy nature of DRR algorithm, here we introduce a variant of DRR named Threshold DRR (TDRR).

TDRR is a modification of DRR algorithm. The modification is made by introducing the lower and upper thresholds for HP buffers, denoted as α and β (α < β), respectively. TDRR switch operates in a manner similar to DRR's one. First, cells are departed from HP buffers at particular output line. When the average occupancy of HP buffers falls below the lower threshold α, the algorithm begins to empty LP buffers at the same output line, in RR manner. When the average occupancy of HP buffers reaches the upper threshold β, the departure of LP buffers stops, and algorithm starts to send cells to output from HP buffers in RR fashion. In this way, the algorithm supports different priority levels in order to assign adequate resources to HP and LP traffic.

III. Simulation Results And Performance Analysis

In our previous papers [11, 12], we have analyzed the performance of DCQ switch with equal HP and LP buffer lengths, under the LRD traffic. Here, we have extended our analysis to DCQ switch with unequal buffer sizes. We have chosen the same type of traffic, as in [11, 12], with Hurst parameter H=0.8, and normalized input traffic load of λ=0.8. Maximal burst flow size (the number of consecutive cells from an input to one output) is varied among B ∈ {100, 300, 500, 700, 1000} cells. The selection of outputs follows the uniform distribution, while the share of HP traffic within input traffic equals ρ=0.25. Simulations are performed for million time slots on a 32×32 DCQ switch, with the following combinations of HP and LP buffer lengths: (HP, LP) = {(8, 1016), (16, 1008), (32, 992), (64, 960), (128, 896)}. Therefore, the crosspoint "memory" size is equal to 1024 cells. For purpose of comparison, results will be compared to the case with equal buffer lengths (512, 512).

Due to lack of space, we will first present results for TDRR algorithm, while the comparison with other algorithms (RR, DRR, PFRR) will be given for fixed burst size. Although the extensive simulations have been performed for different values of TDRR's lower and upper threshold for HP buffers, results will be given for two sets of α and β values, namely (α, β) = {(1, 5), (5, 10)} ‰. These values allow us to keep the mean delay of HP traffic below the mean delay of LP traffic, thus enabling the traffic differentiation and supporting the QoS requirements.

For higher thresholds values, the mean delay for HP traffic becomes higher than mean delay for LP traffic. By lowering thresholds values, the mean delay of HP traffic approaches to the mean delay of DRR that was too low (around 1.2 time slots), which is also greedy behavior that we want to avoid.

The mean delay of TDRR for HP and LP traffic, for different burst flow sizes B and different buffer lengths is shown in Fig. 2. Here, we select (α, β) = (1, 5) ‰. Similarly, Fig. 3 presents results for (α, β) = (5, 10) ‰. One can notice that with the increase of *B*, the mean delay of LP traffic increases, while the mean delay of HP traffic does not change. This is true for both sets of threshold values. With the increase of threshold values, the mean delay of HP traffic increases too.

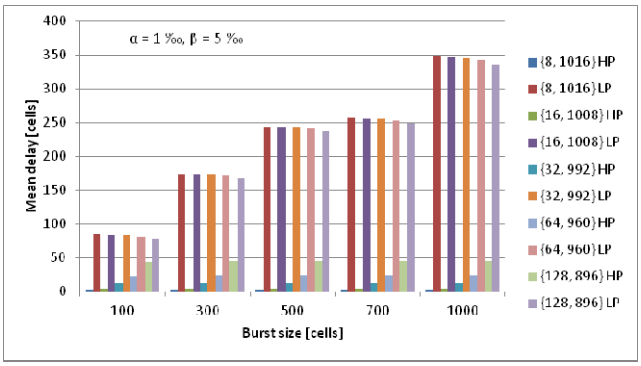

Fig. 2 Mean delay of TDRR for HP and LP traffic for different burst flow sizes B and different buffer lengths; (α, β) = (1, 5) ‰.

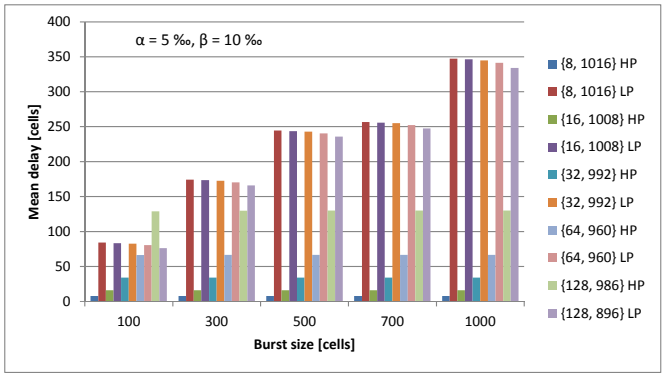

Fig. 3 Mean delay of TDRR for HP and LP traffic for different burst flow sizes B and different buffer lengths; (α, β) = (5, 10) ‰ .

However, the change of threshold values does not affect the mean delay of LP traffic, which is quite interesting. The mean delay of HP traffic can be increased by increasing the HP buffer length, which at the same time has decreasing effect on mean delay of LP traffic since the length of LP buffer decreases. However, the level of decrease is not so significant, while for HP traffic the increase of buffer lengths has much more influence on the mean delay, as shown in Fig. 4.

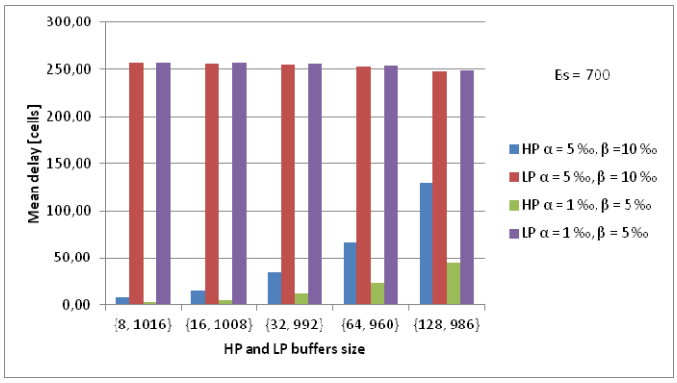

Fig. 4 Mean delay of TDRR for HP and LP traffic for different buffer lengths and different threshold lengths; B = 700 cells.

The probability of TDRR cell loss for HP and LP traffic for different burst flow sizes *B* and different buffer lengths, and for (α, β) = {(1, 5), (5, 10)} ‰ is shown in Figs. 5 and 6, respectively. By comparing these figures, it can be seen that the probability of cell loss for HP traffic is below 10^{-4} for all *B* values and all observed combinations of HP and LP buffer lengths. For most of HP and LP buffer lengths combination, it is below 10^{-5}. The probability of LP cells loss is zero, for burst flow size below 500 cells. However, with the increase of *B*, the probability of cell loss for LP traffic begins to increase. Clearly, for very high burst flow size, the LP buffers are quickly filled, and in order to keep low the probability of cell losses it is necessary to increase the length of LP buffers. The influence of different threshold values on probability of cell loss is depicted in Fig. 7, for B = 700 cells. For HP traffic, the difference is small, while for LP traffic there is no difference.

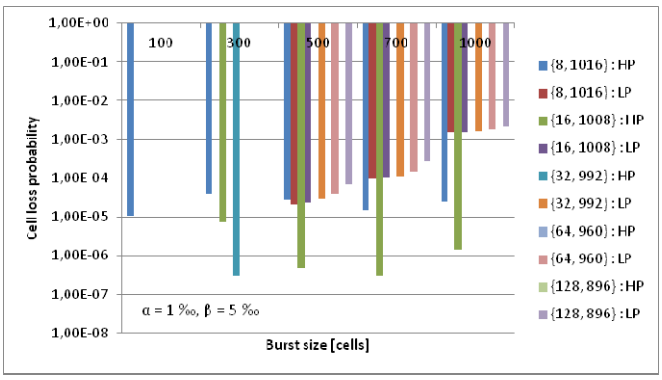

Fig. 5 Cell loss probability of TDRR for HP and LP traffic for different burst flow sizes B and different buffer lengths; (α, β) = (1, 5) ‰.

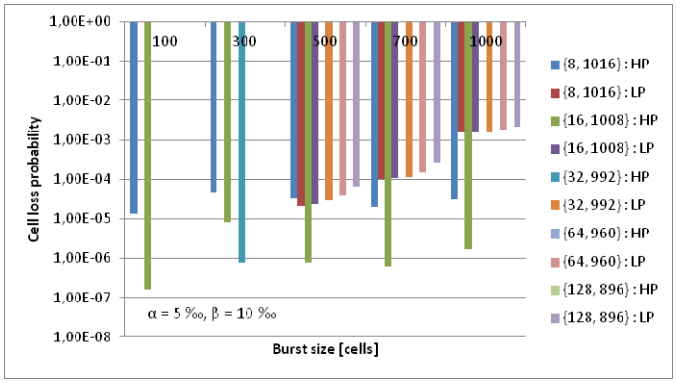

Fig. 6 Cell loss probability of TDRR for HP and LP traffic for different burst flow sizes B and different buffer lengths; (α, β) = (5, 10) ‰.

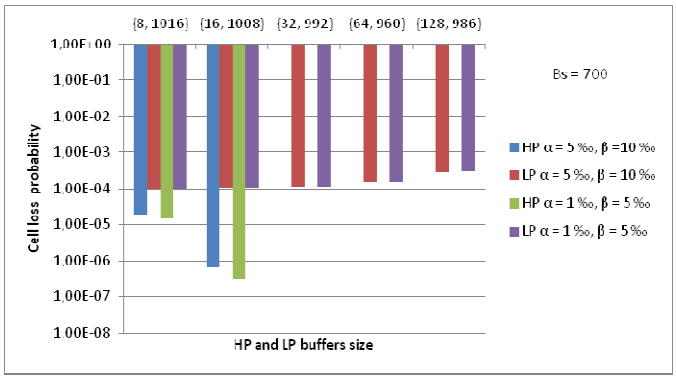

Fig. 7 Cell loss probability of TDRR for HP and LP traffic for different buffer lengths and different threshold lengths; B = 700 cells.

By introducing thresholds in DRR algorithms, we can alter delay of HP traffic, while there is no influence on delay of LP traffic, as well as on cell loss probability. Furthermore, the impact of threshold values on probability of HP cells loss can be considered as negligible.

In order to evaluate performance of TDDR algorithm, we perform the comparison with RR, DRR, and PFRR algorithms. The comparison of scheduling algorithms in terms of mean

delay for different combinations of HP and LP buffer lengths, and B = 700 cells is shown in Figs. 8 and 9.

The probability of cell loss is given in Figs. 10 and 11. The mean delay of HP traffic for PFRR and DRR algorithms is not affected by change of buffer lengths. Thus, their greedy nature is preserved. Similarly, the mean delay of LP traffic does not change too much with the change of LP buffer lengths, when these algorithms are applied.

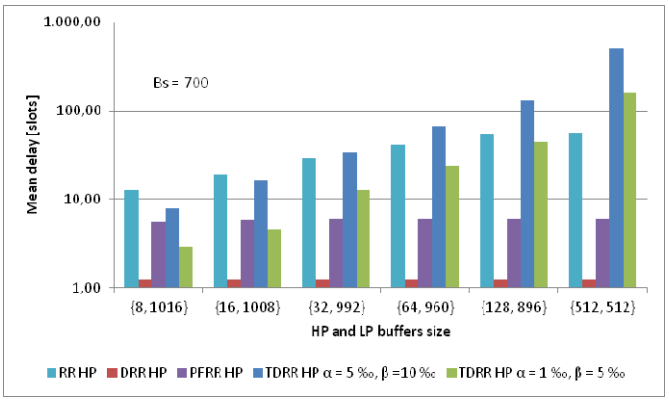

Fig. 8 Mean delay of HP traffic for various scheduling algorithms and different buffer lengths; B = 700 cells.

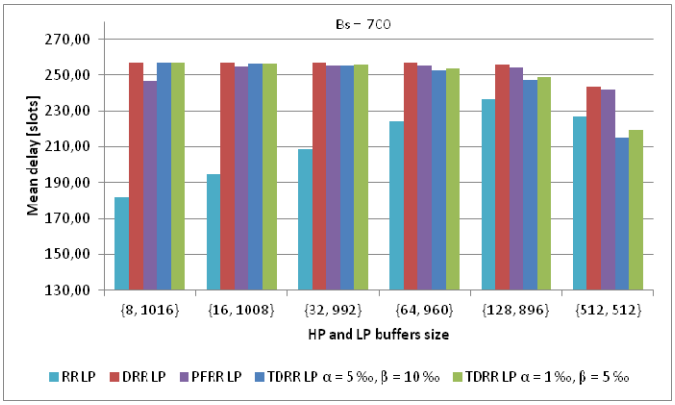

Fig. 9 Mean delay of LP traffic for various scheduling algorithms and different buffer lengths; B = 700 cells.

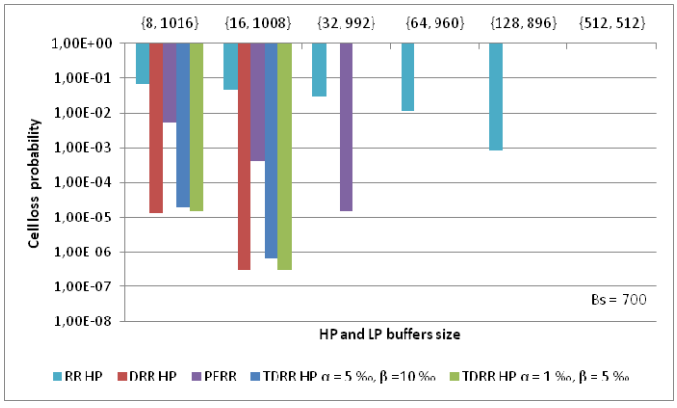

Fig. 10 Cell loss probability of HP traffic for various scheduling algorithms and different buffer lengths; B = 700 cells.

However, the proposed modified DRR algorithm, named TDRR, allows us to affect DRR's greedy nature in terms of mean delay by using simple thresholds and unequal HP and LP buffer lengths. The cell loss probability of HP traffic is not altered too much when compared to DRR, as seen in Fig. 10. However, there is no LP's mean delay or cell loss probability reduction for the incoming traffic with high load and high burst per flow. This traffic will keep LP buffers filled, thus leading to high losses.

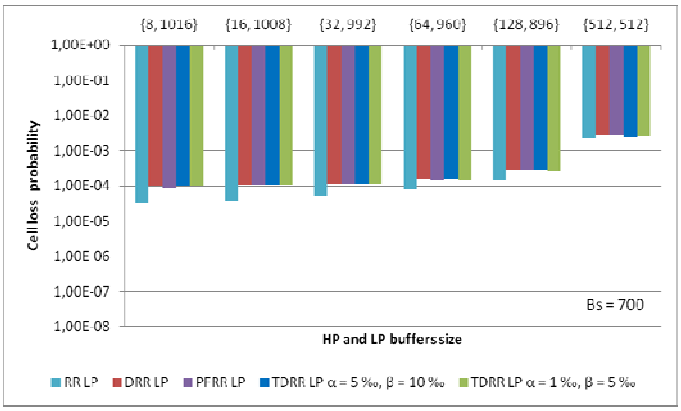

Fig. 11 Cell loss probability of LP traffic for various scheduling algorithms and different buffer lengths; B = 700 cells.

IV. Conclusions and Future Work

In this paper, we present the simulation results and performance analysis of DCQ crossbar switch with unequal buffer lengths dedicated to HP and LP traffic. We have also proposed a new Threshold Dual Round Robin algorithm (TDRR). We have compared TDRR DCQ switch performance with performance of other scheduling algorithms. Results obtained in case of moderate burst sizes are valuable. We showed that with usage of TDRR algorithm, it is possible to alter delay of HP traffic with positive influence on LP delay and cell loss probability and negligible increase of HP cell loss probability. Also, it is shown that unequal sized crosspoint buffers provide possibility for efficient balancing between the LP and HP performance in order to support the traffic differentiation and QoS assurance. For reducing the cell loss probability and mean delay for LP traffic, the dynamic priority scheduling should be applied. This will be the starting point for our further research.

Acknowledgment

This work is partially supported by the Ministry of Education and Science of Bosnia and Herzegovina and the Ministry of Science of Montenegro, under the bilateral Project "Architecture, design and performances of DCQ switch", and by the Ministry of Science of Montenegro under grant 01-451/2012 (FIRMONT).

REFERENCES

[1] Karol, M.J.; Hluchyj, M.G.; Morgan, S.P., "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Transactions on Communications,* vol.35, no.12, pp.1347-1356, Dec 1987.

[2] Elhanany, I.; Matthews, B., "On the performance of output queued cell switches with non-uniformly distributed bursty arrivals," *IEE Proceedings in Communications*, vol.153, no.2, pp.201-204, 1 April 2006.

[3] Shang-Tse Chuang; Goel, A.; McKeown, N.; Prabhakar, B., "Matching output queueing with a combined input output queued switch," in *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*. Proceedings. IEEE , vol.3, no., pp.1169-1178, vol.3, 21-25 Mar 1999.

[4] Katevenis, M.; Passas, G.; Simos, D.; Papaefstathiou, I.; Chrysos, N., "Variable packet size buffered crossbar (CICQ) switches," *IEEE International Conference in Communications*, vol.2, no., pp.1090-1096, 20-24 June 2004.

[5] Kanizo, Y.; Hay, D.; Keslassy, I,. "The Crosspoint-Queued Switch", *IEEE Infocom'09 Conference Proceedings*, pp.729-737, Brazil, 2009.

[6] Radonjic, M.; Radusinovic, I., "Impact of scheduling algorithms on performance of crosspoint-queued switch", *Annals Of Telecommunications*, vol.66, no.5-6, pp.363-376, 2011.

[7] Lee, T.-H.; Kuo, Y.-C., "Parallel matching algorithm for CIOQ switches with multiple traffic classes," *IEE Proceedings* in *Communications*, vol.150, no.5, pp.354-60-, 14 Oct. 2003.

[8] Hu, H.; Guo, Y.; Yi, P.; Chen, S., "Load-balanced differentiated services support switch," *in IET Communications,* vol.5, no.13, pp.1895-1906, Sept.5,2011.

[9] Balasubramanian, K.; Umamageswari, G.; Ganesh, M., "High-performance prioritization scheme for input-queued packet switches," *International Conference on Control, Automation, Communication and Energy Conservation, INCACEC 2009*., vol., no., pp.1-7, 4-6 June 2009.

[10] Leland, W. E.; Taqqu, M. S.; Willinger, W.; Wilson, D.V., "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, Vol. 2, No. 1, pp. 1-15, February 1994.

[11] Gardasevic G.; Divanovic S.; Radonjic M.; Radusinovic I., "A Qos-Aware Dual Crosspoint Queued Switch With Largest Weighted Occupancy First Scheduling Algorithm", *IEICE Transaction on Communications,* Vol.E98-B, No.01, January 2015, pp. 201-208.

[12] Maletic, N.; Divanovic, S.; Radonjic, M.; Radusinovic, I.; Gardasevic, G., "Performance evaluation of Dual Crosspoint Queued crossbar packet switch," *11th International Conference on in Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), 2013*, vol.01, no., pp.145-148, 16-19 Oct. 2013.

[13] Divanovic, S.; Radonjic, M.; Radusinovic, I.; Gardasevic, G., "Performance analysis of crosspoint queued crossbar switch with weighted round robin scheduling algorithm under unbalanced bursty traffic," *IEEE Symposium on Computers and Communications (ISCC*), 2013, vol., no., pp.000120-000124, 7-10 July 2013.

[14] Divanovic, S.; Radonjic, M.; Gardasevic, G.; Radusinovic, I., "Dynamic weighted round robin in crosspoint queued switch," *in Telecommunications Forum (TELFOR*), 2013 21st , vol., no., pp.109-112, 26-28 Nov. 2013.

[15] Radusinovic, I.; Divanovic, S.; Radonjic, M., "Analysis of WRR scheduling algorithm frame size impact on CQ switch performance," in *17th IEEE Mediterranean Electrotechnical Conference (MELECON*), 2014 , vol., no., pp.403-407, 13-16 April 2014.