# CQ Switch Analysis under Traffic Overload

Milutin Radonjić, *Member, IEEE*, Igor Radusinović, *Member, IEEE,*
Ivo Maljević, *Member, IEEE,* and Dušan Banović

*Abstract* — **An analysis of 2x2 crossbar packet switch with buffers at crosspoints and round robin scheduling algorithm is presented in this paper. The analysis is performed for a non-admissible traffic pattern, where output ports are overloaded. The case of full offered load is observed and output ports are loaded with packets that have different arrival probabilities. In addition to the parameters that are commonly observed in such an analysis (throughput and average packet delay), memory requirements for the implementation of the buffer, as well as fair representation when servicing the buffer - the so-called fairness are also analyzed. The results show that even for a switch with a small number of ports very large buffers should be implemented, if we want to achieve satisfactory performance under traffic overload.**

*Keywords* — **Average cell latency, Crossbar switch, Fairness, Scheduler, Throughput.**

## I. Introduction

With today's stage of development of computer communications, there is a constant need for performance improvement of the key components for data transmission, especially switches and routers. Crossbar architecture has been used for a long time in the construction of these devices, due to its simplicity and no internal blocking capability [1]. The literature survey also shows that many solutions are known based on the crossbar switching matrix. The most prevalent solution is with buffers at input ports, organized in so-called virtual output queues (VOQ) [2].

The VOQ switches are a very good solution as long as the line cards (containing buffers with packets awaiting forwarding) are close to the switching matrix. Namely, the scheduler which decides about the packages that will be forwarded at the next time slot must have accurate information about the buffers occupancy. This causes high level of control communication between the buffers and packet scheduler which is an integral part of the crossbar switch. If the buffers are located near the switch, it can be considered that this communication is instantaneous, in terms of packet transmission speed. However, in modern transport hubs it has become common for the buffers to be quite far from the switching fabric (Distributed Systems), which results in control communication being no longer instantaneous, which in turn affects the performance of the entire device.

Several approaches can be taken to solve this problem. The simplest one is to increase the duration of the time slot reserved for a transmission of a single packet through the switching matrix. This would solve the problem of time required for the control communication, but it would also degrade the transmission speed of switching device. Another approach, known from the literature, is based on the implementation of small buffers at the crosspoints of switching matrix, in addition to the existing buffers in the input line cards [3]. Such a solution alleviates the problem of the control communication duration, but does not completely eliminate it, because it is still necessary that the scheduler has the information about the input buffer occupancy.

The best solution would be the one that completely eliminates the control communication. This would mean that there are no buffers at the inputs, but only at the crosspoints of the switching matrix. This solution was not implementable in the past due to technological limitations. Namely, it was difficult to place the switching matrix with a scheduler and buffers of high capacity on the same chip. However, it has been recently shown that, using modern technology, it is now possible to implement large buffers at the crosspoints, and the larger problem is to ensure a sufficient number of pins to input/output data on the chip [4].

The switch with buffers only at crosspoints of switching matrix is called the CQ (Crosspoint Queued) switch (Fig. 1). Packets that arrive through the input *i* and are intended to the output *j* are placed in the buffer *Bij*. In each time slot the scheduler selects one of the occupied buffers belonging to the same output, and forwards its head-of-line packet to the output. Each output is considered independently. The choice of the buffer that will be served is made based on one of the pre-defined algorithms [5]. If the buffer is full upon arrival of the new packet, that packet is discarded.

The CQ switch performance analysis has been, in earlier research, performed under different traffic conditions and different configurations of the switch [6] - [8]. Simulations were done for switches with different numbers of ports and with different buffer sizes. Performance is observed in the case of uniform, unbalanced, IBP (Interrupted Bernoulli Process) and several variants of nonuniform traffic. However, in all these cases it was a so-called admissible traffic that was considered, which means that input/output is  not overloaded with traffic. In other words,

This work is partly supported by the Montenegrin Ministry of Science, under the Project "Research of switching circuit structure impact on NetFPGA based packet switching performance" (Grant No. 3358/2008).

Milutin Radonjić and Igor Radusinović are with the School of Electrical Engineering, University of Montenegro, Bul. Džordža Vašingtona bb, Podgorica, Montenegro; (phone: 382-20-245839; e-mail: {m.radonjic, igorr}@ieee.org).

Ivo Maljević is with the TELUS Mobility, 200 Consilium Place, Suite 1300 Scarborough, Ontario, Canada, M1H 3J3; (e-mail: ivom@ieee.org).

Dušan Banović is with Crnogorski telekom, Moskovska bb, Podgorica, Montenegro; (e-mail: ban@t-com.me).

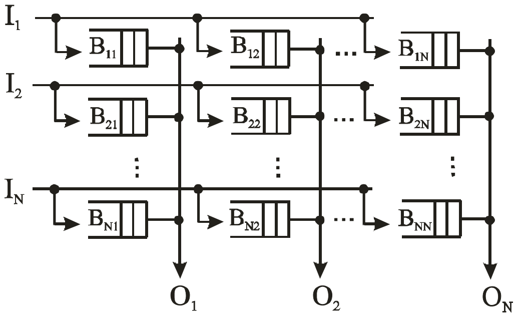

Fig. 1. CQ switch architecture.

statistically speaking, the load on any port is not greater than one. In this study we will observe the behavior of the CQ switch in non-admissible traffic conditions, where it is possible to overload some of the ports. The results show that it is required to implement very large-capacity buffers to achieve satisfactory performance. This is particularly true if some of the output ports are overloaded with high traffic.

This paper is structured as follows. In Section 2, after introductory remarks, the simulation model of the observed switch is explained. Simulation results and analysis are presented in Section 3. Concluding remarks are given in Section 4.

## II. Simulation Model

The results of the simulation of the 2x2 crossbar switch functionality with buffers only at the crosspoints are presented in this paper. It is assumed that at each time slot on both of the input ports arrives a single packet (input load $\rho = 1$). The probability that the packet which arrived on the first input is intended for the first output is denoted as $p_1$, which means that the probability that the packet arrived on the first input is intended for the second output equals 1-$p_1$. By the same token, the probability that the packet from the second input is intended for the first output is $p_2$, and that is intended for the second output is 1-$p_2$.

Simulations were run with 10 million time slots, for different values of parameters $p_1$ and $p_2$, in the range (0, 1], with the step of 0.1. The CQ switches with various buffer lengths ($L$) are observed: 1, 2, 3, 4, 8, 16, 32, 64, 128, 256 and 512. The buffer length implies the number of cells that can be accommodated. For the theoretical purpose, the unlimited buffer length is also considered.

Several scheduling algorithms are simulated. We will present results for Round Robin (RR) algorithm, as it is the simplest choice for implementation. With the Round Robin algorithm, the occupied buffers are serviced on the particular output line in the circular (round robin) order, handling all buffers without priority. After the departure of one cell from the buffer, the next occupied buffer is serviced in the following time slot. This is always performed in the same order as determined by the buffer position.

The commonly used parameters for evaluation of switch performance are the throughput and the average cell latency.

The switch throughput is defined as the ratio between the total number of cells entering it successfully and the maximum possible number of cell arrivals during the simulation. The latter is calculated as the number of ports multiplied by the number of simulated time slots.

The average cell latency is defined as an average delay of cells that are traversing the switch. It is calculated as the total time that cells spend inside the buffers (expressed in time slots) divided by the number of accepted cells during the simulation.

In addition to these parameters, it is of interest to investigate the impact of the scheduling algorithm on the fair representation of particular buffers (ports) during the transferring process. In the available literature this characteristic is known as fairness. One of the most popular measures for fairness estimation is *Jain's Fairness Index* (*JFI*) [9]. *JFI*, for a 2x2 switch, can be defined as:

$$JFI = \frac{\left(\frac{t_1}{p_1} + \frac{t_2}{p_2}\right)^2}{2\left(\left(\frac{t_1}{p_1}\right)^2 + \left(\frac{t_2}{p_2}\right)^2\right)}, \tag{1}$$

where $t_i$ ($i$=1,2) denotes the ratio of the number of accepted cells arrived on input $i$ and intended for the first output, and the number of observed time slots (throughput). Therefore, parameters within *JFI* are observed according to one of the outputs (in this case, it is the first output).

The *JFI* parameter value can be in the range of [0, 1], where *JFI*=1 means the total fairness between communication links, and *JFI*=0 means no fairness at all.

## III. Simulation Results

### A. Throughput analysis

The throughput of CQ switch as a function of probabilities $p_1$ and $p_2$, in the case of one-cell buffer length ($L$=1) is shown in Fig. 2. Throughput ranges from 0.5 (when $p_1$=$p_2$=0 and $p_1$=$p_2$=1) to 0.95 (when $p_1$=1 and $p_2$=0.1, as well as $p_1$=0.1 and $p_2$=1). Generally, the diagram is symmetrical to both diagonals of quadrant determined by $p_1$ and $p_2$. The most interesting part of the diagram is the one determined by probability values which satisfy following equation: $p_1$+$p_2$=1. This part of diagram represents the border of admissible traffic, where the drop in the throughput value as a consequence of small buffers can be clearly identified. The lowest throughput, in that

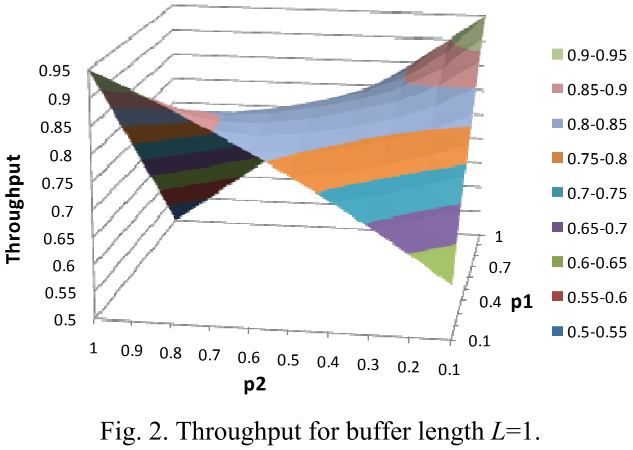

Fig. 2. Throughput for buffer length $L$=1.

case, is when the probabilities have very close values. The larger the difference between the probability values, the higher the throughput. Given the definition of these probabilities, such results were to be expected.

When the probabilities $p_1$ and $p_2$ move farther away from the linear relationship $p_1+p_2=1$, the throughput dramatically decreases. This is the consequence of the nature of such traffic, that is, because one of the output ports is overloaded.

In the case when longer buffers are used, the throughput increases when the conditions are close to the admissible traffic. However, when the traffic overloads one of the output ports, the increase of the length of the buffer does not result in a significant increase in throughput. As an illustration, Fig. 3 shows the throughput for the buffer length $L$=8. In the area around admissible traffic, switch has almost the same throughput for all values of the probabilitis $p_1$ and $p_2$, which is approximately 0.98. The shape of the diagram remains the same with further increase of the length of the buffer, and it changes only the maximum bandwidth that becomes closer to the one.

### B. *Average Cell Latency Analysis*

The diagram of the average cell latency as a function of probabilities $p_1$ and $p_2$, for one-cell buffer length ($L$=1), is shown in Fig. 4. Evidently, the average cell latency has very small values because of the very short buffers. Conversely, a different trend is noticed for longer buffers. When one of the output ports is overloaded, the average cell latency has a higher value, which corresponds to a low switch throughput. Within the area near the line $p_1+p_2=1$, a cells spend less time in buffers which corresponds to high throughput.

With the increase of the buffer size the average cell latency will also go up. The shape of the diagrams does not change significantly with the increase of the buffer length, which means that most of the earlier conclusions are valid for longer buffers as well. To illustrate the average cell latency diagrams for longer buffers, the diagram for the buffer length of $L$=64 is shown in Fig. 5.

### C. *Maximal buffer occupation during the simulation*

As a part of the analysis, we also observed a theoretical case where crosspoint buffers had no length limitation. In this case, it is of interest to determine what the maximum occupancy of the buffers is during the simulation. In other words, we try to determine the buffer length that enables the acceptance of all incoming cells. The results are shown in Fig. 6.

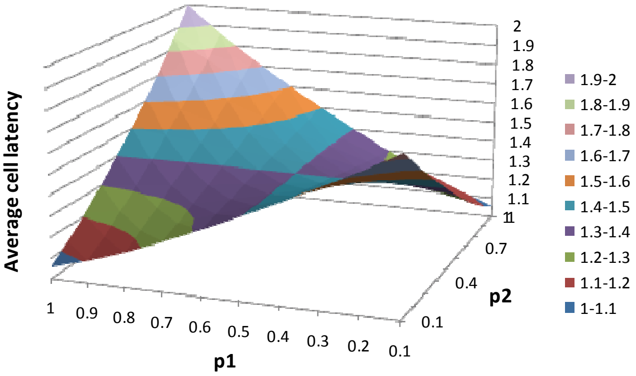

Fig. 4. Average cell latency for buffer length $L$=1.

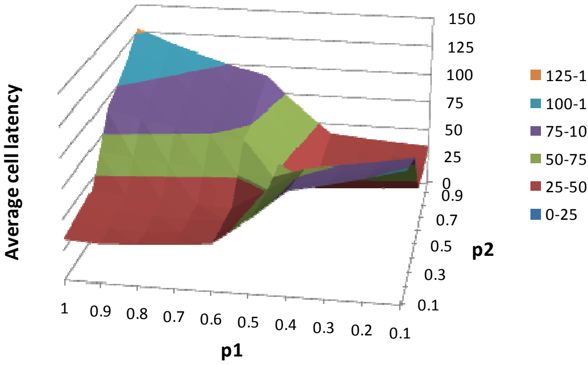

Fig. 5. Average cell latency for buffer length $L$=64.

The smallest memory requirements, as expected based on the throughput results, are around the admissible traffic zone. In the case when $p_1+p_2=1$, the required buffer length can be as high as 3000 cells. When moving away from the admissible traffic zone, the memory requirements dramatically increase so much that, for instance, in the case of $p_1+p_2=1.1$, memory buffers capable of accommodating around million cells are required. In the case one of the output ports is heavily loaded ($p_1+p_2$ very far from one) the memory requirement grows to around 5 million cells.

Clearly, buffers with such dimensions are not suitable for practical implementation. Nevertheless, the information about the maximum buffer occupancy can be

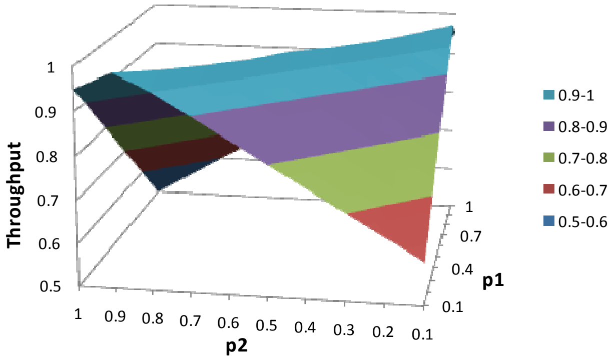

Fig. 3. Throughput for buffer length $L$=8.

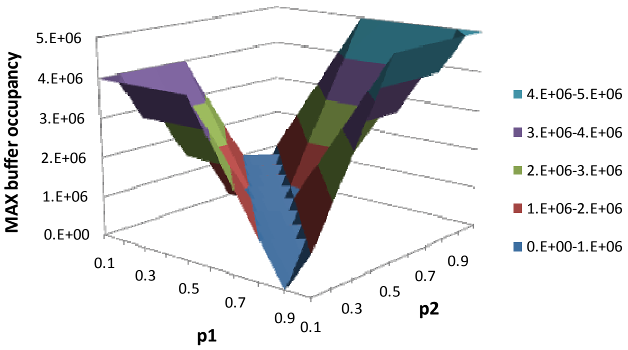

Fig. 6. Maximal buffer occupancies during the simulation.

useful for estimation of the buffer length that will be chosen for implementation, under certain traffic conditions.

### *D. Fairness analysis*

The *JFI* parameter values as a function of probabilities $p_1$ and $p_2$, for the CQ switch with the one-cell buffer length are shown in Fig. 7. The Round Robin algorithm exibits a completely equal service of ports (total fairness), where $p_1$=$p_2$. Such a result was to be expected since both buffers are equally loaded. Given that the RR algorithm alternately services the occupied buffers, they are both equally loaded and serviced. Consequently, the total fairness (calculated on the basis of throughput) is achieved.

For that reason, it is even more importrant to observe the fairness in the area where the output port is unequally loaded from the input ports. The lowest fairness is achieved when one of the probabilities equals one and the other equals 0.5. This corrsponds to the case where one buffer will be filled in every time slot, and the other (statistically speaking) once every two consecutive time slots. Because the RR algorithm alternately services buffers, it is clear that the buffer which is always filled will not be serviced fairly. Smaller values of $p_1$ and $p_2$, as well as their closer values, result in the increase of fairness.

Further analysis is performed for longer buffers, where it was shown that the *JFI* diagrams are very similar for a broad range of buffer lengths. The results differ only in the *JFI* index values for particular values of probabilities $p_1$ and $p_2$. The lowest *JFI* among all observed values of buffer lengths is observed when one probability equals 1 and the other equals 0.5. With the increase of buffers lengths, the values of minimum *JFI* decrease. For example, the minimal fairness index with buffer length of $L$=512 is around 0.9 (Fig. 8). Furthermore, longer buffers cause a wider range of probabilities which have total fairness (*JFI*=1). This is the consequence of the fact that *JFI* is based on the throughput, which increases with longer buffers. As can be noticed from Fig. 8, with buffer length of $L$=512, the probability range which ensures the total fairness satisfies $p_1+p_2 \leq 1$.

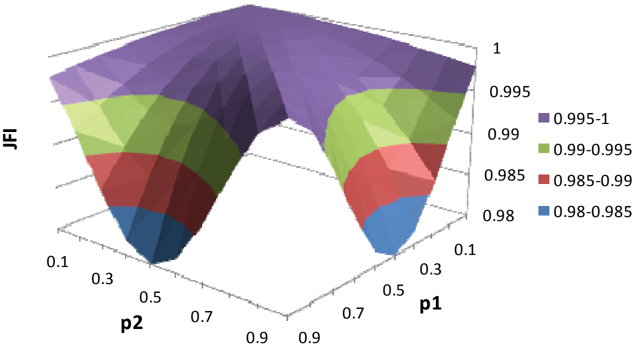

Fig. 7. JFI parameter for buffer length $L$=1.

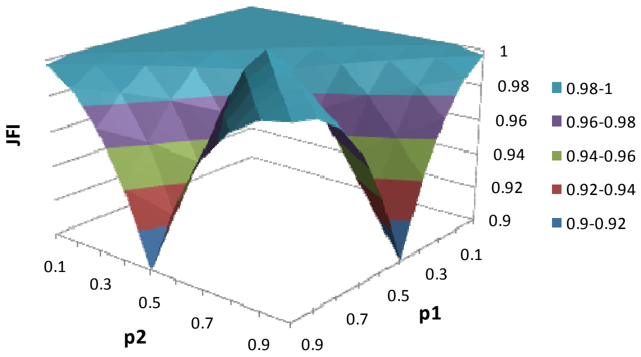

Fig. 8. JFI parameter for buffer length $L$=512.

## IV. Conclusion

The results presented in this paper show that very long buffers are required for the case of very high level of the incoming traffic, even for switches with a small number of ports. The highest throughput is obtained in the zone near the admissible traffic, and by moving away from this zone the throughput rapidly decreases. Additionally, the overload in one of the output ports causes very high average cell latency.

The Round Robin algorithm exhibits a very high fairness with regard to servicing buffers in case of the load balanced output ports. Longer buffers increase the area of the total fairness, as long as $p_1+p_2 \leq 1$ inequality is satisfied.

Further studies are under planning in the domain of performance analysis of some other scheduling algorithms, as well as larger switches.

## References

[1] A. Mekkittikul, N. McKeown, "A Practical Scheduling Algorithm to Achieve 100% Throughput in Input-queued Switches", in *Proc. of INFOCOM '98*, San Francisco, USA, 1998, pp. 792-799.

[2] N. H. Liu, K. L. Yeung, D. C. W. Pao, "Scheduling Algorithms for Input-queued Switches with Virtual Output Queueing", in *Proc. Of ICC'2001*, Helsinki, Finland, June 2001, pp. 2038 - 2042.

[3] R. Rojas-Cessa, E. Oki, Z. Jing, H. J. Chao, "CIXB-1: Combined Input-One-cell-crosspoint Buffered Switch", in *Proc. of IEEE HPSR '01*, Dallas, Texas, USA, 2001, pp. 324-329.

[4] Y. Kanizo, D. Hay, I. Keslassy, "The Crosspoint-queued Switch", in *Proc. of INFOCOM '09*, Rio de Janeiro, Brazil, 2009, pp. 729-737.

[5] I. Radusinović, M. Pejanović, Z. Petrović, "Impact of Scheduling Algorithms of Performances of Buffered Crossbar Switch Fabric", *IEEE ICC 2002,* New York, USA, 2002, pp. 2416-2420.

[6] M. Radonjic, I. Radusinovic, "Buffer Length Impact to Crosspoint Queued Crossbar Switch Performance", in *Proc. of the 15th IEEE MELECON*, Valleta, Malta, 2010, pp. 119-124.

[7] J. Cvorovic, I. Radusinovic, M. Radonjic, "Buffering in Crosspoint-queued Switch", in *Proc. of the 17th TELFOR 2009*, Belgrade, Serbia, 2009, pp.198-201.

[8] M. Radonjic, I. Radusinovic, "Impact of scheduling algorithms on performance of crosspoint queued switch", Annals of Telecommunications, Vol 66, No 5-6, pp.363-376.

[9] S. Bhatti, M. Bateman, D. Rehunathan, T. Henderson, G. Bigwood, D. Miras, "Revisiting inter-flow fairness", in *Proc. of BROADNETS 2008*, London, 2008, pp. 585–592.