Delay Performance of High-Speed Packet Switches with Low Speedup

Paolo Giaccone, Emilio Leonardi
E.E. Dept., Politecnico di Torino
e-mail: {giaccone,leonardi}@polito.it

B. Prabhakar, D. Shah
E.E. Dept., Stanford University
e-mail: {balaji,devavrat}@stanford.edu

Abstract— **The *speedup* of a switch is the factor by which the switch, and hence the memory used in the switch, runs faster compared to the line rate. In high-speed switches, line rates are already touching limits at which memory can operate. In this scenario, it is very important for a switch to run at as low a speedup as possible.**

In the past, it has been shown that 100% throughput can be achieved for any admissible traffic for an Input Queued (IQ) switch [1], [2] at speedup one. This gives finite average delays but does not guarantee control on packet delays. In [3], authors show that a Combined Input Output Queued (CIOQ) switch can emulate perfectly an Output Queued (OQ) switch at a speedup of 2 and, thus, control the packet delays. This motivates the study of possibility of obtaining delay control at speedup less than 2. To guarantee optimal control of delays for a general class of traffic, as shown in [3], speedup 2 is necessary. Hence, to obtain control of delays at lower speedup, we need to restrict the class of arrival traffics. In this paper, we study the speedup requirement for a class of admissible traffic, which we will denote as $(1, nF)$-regulated traffic, with parameters n and F. We obtain the necessary speedup for this class of traffic. Further, we present a general class of algorithms working at the necessary speedups and thus providing bounded delays.

I. Introduction

Recently, Input-Queued (IQ) and Combined-Input-Output-Queued (CIOQ) switches with Virtual Output Queueing (VOQ) have become an attractive architectural solution in very high speed routers [4], [5] as they scale well with the line rate.

At the same time, Output-Queued (OQ) switches are attractive as they achieve 100% throughput under any admissible traffic and give control over delays. But OQ switches require memory bandwidth (at the output ports) to scale as $O(rN)$, where r is the line rate and N is the number of ports. In other words, the internal switching speed has to run N times faster than the line rate, that is, speedup S is N. This constrains the speed at which OQ switches can run.

A pure IQ switch is able to achieve very high speeds, since the memory bandwidth scales as $O(r)$, being by construction its speedup equal to 1. The main drawback of this architecture is that it requires a scheduling algorithm which selects a non-conflicting set of packets to transfer across the switch. This scheduling algorithm should be simple, because it is implemented in hardware at very high speed. A class of Maximum Weight Matching (MWM) algorithms for IQ switches are known which provide 100% throughput for any admissible traffic [1], [2], [6]. In [7], [8] bounds on the average delay are obtained for MWM algorithm under admissible Bernoulli i.i.d. traffic pattern. But they do not guarantee delay bounds for each packet. Many practical scheduling algorithms [9], [10] have been proposed to approximate MWM performance. Their simplicity usually leads to some performance penalties, usually in the form of throughput degradation and/or larger delays.

In [2], [11] it is shown that at speedup 2, simple maximal matching kind of algorithms are stable (provide 100% throughput) under admissible arrival traffic. But again, there are no strict delay guarantees provided. In [3] it is shown that $S \geq 2$ is necessary and sufficient to emulate performance of OQ switches and, thus, to control the delays. Unfortunately the perfect emulation of OQ requires complicated stable-marriage style algorithms which are not feasible to implement at a very high-speed. In [12] it was shown that simpler scheduling algorithms can achieve the same performance of an OQ switch in terms of average delay.

Since speedup higher than 1 limits the speed at which a switch can operate, it is very desirable to operate at as low speedup as possible. This leads us to investigate a possible *tradeoff between speedup and delay*. However, if we want to obtain delay control for speedup $1 \leq S < 2$, we must restrict the arrival traffic. In this paper, we consider a general enough class of arrival traffic and study the necessary and sufficient speedup $1 \leq S < 2$ required to emulate OQ performance with guaranteed delay bounds.

II. Basic Model, Definitions and Notations

A. A CIOQ Switch

An $N \times N$ CIOQ switch has N inputs and N outputs with crossbar in the switch fabric. The queues at each input is logically divided into N Virtual Output Queues (VOQ) corresponding to N different outputs. There are queues at output too. When a CIOQ switch is working at speedup S (with $1 \leq S \leq N$), each input is able to transfer up to S packets per time slot, and each output is able to receive up to S packets per time slot. At speedup $S = 1$ a CIOQ switch is same as IQ switch, and does not require queues at the output side.

We assume that time is slotted. In a given time slot, at most one packet can arrive at each input. In every "scheduling cycle", the crossbar can transfer one packet from each input and one packet to each output. Effectively for a CIOQ switch operating at a speedup S, S scheduling cycles happen during 1 time slot. For example, if $S = 3/2$, then every 1 time slot 1.5 scheduling cycles happen. That is, in real switch, every 2 time slots, 3 scheduling cycles happen.

B. Work Conservation

Next we would like to consider the concept of work conservation for a switch. Consider the following definition, which was first proposed in [12] motivated from the classical queueing theory.

Definition 1. A switch is *work-conserving* if and only if, for any time slot, an output is always transferring one packet to the outgoing link whenever a packet is present in the system directed to the considered output.

0-7803-7632-3/02/$17.00 ©2002 IEEE

Note that this definition requires that the system should be "observed" at each time slot to check if it is work-conserving.

An OQ switch is by construction work-conserving whereas an IQ switch is not work-conserving. For example, consider a 3×3 IQ switch in which at time $t = 0$ no backlog exists and at time $t = 1$ two packets arrive: one at input 1 directed to output 3 and one at input 2 directed to output 3. An arrived packet is immediately transferred to the outputs and transmitted, while the other packet is stored at the input. At time $t = 2$ other two packets arrive: one packet at input 1 directed to output 2 and one packet at input 2 directed to output 1. Now at the inputs there are three packets directed to different outputs, but only two of them can be transferred to the outputs thus an output port remains idle even if there is a packet directed to it. As a conclusion an IQ switch can not be work conserving. Note that a work-conserving switch ensures the minimum average delays, (i.e. the same average delay than an OQ switch) since an output is never idling as long as a packet directed to it is in the switch.

The work-conserving property of OQ switch suggests the following equivalent work-conservation property which was first considered in [3]:

Definition 2. A switch, in particular CIOQ switch, is *work-conserving* iff, for any arrival sequence $\mathcal{A}$ the following holds for all the time: for each output j, the number of packets in the switch waiting for transmission to j equals the number of packets that would be stored in an OQ under the same $\mathcal{A}$.

From [3], speedup 2 is necessary to emulate OQ and hence to be strictly work-conserving for a CIOQ switch. The goal of this paper is to consider the switch operating at speedup $1 \leq S < 2$ while providing bounds on performance difference between CIOQ switch and an OQ switch. This leads to the notion of little less strict work-conserving property which we call as F-work-conservation. Basically, instead of requiring the system to be work conserving every time, we consider system with property of work-conservation holding at every F times.

Definition 3. A CIOQ switch is *F-work-conserving* iff, for any arrival sequence $\mathcal{A}$ the following holds for time $t = 0, F, 2F, \ldots, kF, \ldots$: for each output $j$the number of packets in the switch waiting for transmission directed to output j equals the number of packets that would be stored in an OQ under the same $\mathcal{A}$. We call the time interval $\{t \in \mathbb{Z}^+ : t \in [(k-1)F + 1, kF]\}$ as the *k^{th} observation window.*

The most important property about F-work conserving switches is about the control of the delays. We compare the delays experienced by packets in a CIOQ switch with an F-work-conserving policy and in an OQ switch under the same arrival sequence.

Theorem 1. *Fix any admissible arrival traffic sequence $\mathcal{A}$ at a switch of size N. Suppose an OQ switch and an F-work conserving CIOQ switch are given the same arrival traffic pattern $\mathcal{A}$. For any packet $P \in \mathcal{A}$, let T_{OQ}^P be the departure time from the OQ switch. Similarly, let $T_{\mathcal{D}}^P$ be the departure time of the same packet P under the F-work conserving CIOQ switch. Then for every P departing from OQ switch, there exists a unique packet $P' \in \mathcal{A}$ departing from CIOQ switch from the same output as P, such that,*

$$T_{\mathcal{D}}^{P'} - T_{OQ}^P \leq F - 1. \quad (1)$$

Hence, the average delay per packet experience by F-work conserving CIOQ switch is at most $F - 1$ more than the OQ switch for each feasible traffic pattern $\mathcal{A}$.

Proof. We apply exactly the same traffic sequence $\mathcal{A}$ to both: (a) an OQ switch, and (b) an F-work conserving CIOQ switch.

We would like to prove the statement by induction. At time $t = 0$, both systems start empty and hence statement is trivially true. Assume that the theorem statement is true for all packets departing from OQ till time kF. By F-work conservation property, the number of packets queued for any of the output in both OQ and CIOQ switch is the same at time kF. Consider $P_1, \ldots, P_m$ packets departed from output j in OQ switch between time $kF + 1, \ldots, (k+1)F$, where $m \leq F$, depending on arrival pattern $\mathcal{A}$. Since,
- at the end of time kF, both OQ and CIOQ had the same number of packets enqueued for output j,
- at the end of time $(k+1)F$, both OQ and CIOQ have the same number of packets enqueued for output j, and
- there are m packets $P_1, \ldots, P_m$ departing from output j in OQ switch between time $kF + 1, \ldots, (k+1)F$,
- there are m packets $P_1', \ldots, P_m'$ departing from output j of CIOQ by the end of time $(k+1)F$.

We can associate each of the P_i with unique P_i' and obtain,

$$T_{\mathcal{D}}^{P_i'} - T_{OQ}^{P_i} \leq F - 1$$

which means that the average departure time in CIOQ differs at most by $F - 1$ from OQ. Then the same property holds for the average delay, since the arrival sequence is the same for CIOQ and OQ. This completes the proof of Theorem 1. □

We would like to note that the Theorem 1 refers to a much stronger property than just a bounded average delays. For example, under admissible traffic an IQ switch running at speedup 1 and using MWM scheduling policy has a bounded average delay, and hence bounded average delay with respect to OQ switch too (by definition OQ has average delay ≥ 0). But it does not imply the property of Theorem 1.

C. Notations

Consider an $N \times N$ CIOQ switch. We observe the system at times $t_k = kF, \forall k \in \mathbb{Z}^+$, since we are interested in F-work conserving property. We define the following notations:
- B_{ij}^k is the number of packets enqueued at the input port i and destined to output j, sampled at the beginning of the observation window k, at time $t = kF, \forall\, k \in \mathbb{Z}^+$.
- $\hat{B}_j^k \triangleq \sum_i B_{ij}^k$ and $\bar{B}_i^k \triangleq \sum_j B_{ij}^k$.
- $A_{ij}(t)$ is the number of arrivals from input i to output j at time t, $\forall t \in \mathbb{Z}^+$; $A(t) = [A_{ij}(t)]$. A_{ij}^k is the cumulative number of arrivals from input i to output j occurring during the $(k-1)^{th}$ observation window: $A_{ij}^k = \sum_{t=(k-1)F}^{kF-1} A_{ij}(t)$. $A^k = [A_{ij}^k]$.

- $\hat{A}_j^k \triangleq \sum_i A_{ij}^k$ and $\bar{A}_i^k \triangleq \sum_j A_{ij}^k$.
- D_{ij}^k is the cumulative number of services from input i to output j, occurring during the k^{th} observation window. $D^k = [D_{ij}^k]$.
- $\hat{D}_j^k \triangleq \sum_i D_{ij}^k$ and $\bar{D}_i^k \triangleq \sum_j D_{ij}^k$.
- O_j^k is the number of packets enqueued at the output port j, sampled at the beginning of the k^{th} observation window.
- $Y_j^k = \sum_i B_{ij}^k + O_j^k$ is the total number of packet queued in the system and destined to output j.
- $\lceil x \rceil^+ = \max\{0, x\}$.

To model the system, we consider the switch evolving in a *gated-fashion* with period F, i.e. new arrivals are aggregated during each observation window and they are scheduled only at the beginning of the next observation window. It is like considering batch arrivals at the beginning of a new observation window, by batching all the arrivals during the previous observation window. The evolution of the state of the system is sampled at the beginning of a new observation window and can be modeled as follows:

$$B_{ij}^{k+1} = B_{ij}^k + A_{ij}^k - D_{ij}^k \qquad \forall i,j \qquad (2)$$

$$O_j^{k+1} = \lceil O_j^k + \sum_i D_{ij}^k - F \rceil^+ \qquad \forall j \qquad (3)$$

$$Y_j^{k+1} = \lceil Y_j^k + \hat{A}_j^k - F \rceil^+ \qquad \forall j \qquad (4)$$

Eq. (2) models the system evolving in a gated fashion. Indeed, the new backlogged packets are given by the old ones, plus the new arrivals and minus the departures, both occurring during the previous observation window. Note that, when $F = 1$, Eq. (2) degenerates into the evolution of a generic discrete-time queue. It is important to highlight that a system evolving in a gated fashion can increase the delay of a packet by at most F time slots, with respect to a slot-by-slot system. Eqs. (3) and (4) describe the transfer of all the scheduled packets directed to a generic output; in fact, during each observation window, at most F packets can be transferred to the output line cards.

Define the following norm:

Definition 4 (IO Norm). Given $X \in \mathbb{R}^{N^2}$:

$$\|X\|_{IO} \triangleq \max\{\max_j\{\sum_i X_{ij}\}, \max_i\{\sum_j X_{ij}\}\}$$

A policy $\mathcal{D}$ working with a speedup S is *feasible* if:

$$\|D^k\|_{IO} \leq SF \qquad \forall k, B_{ij}^k, A_{ij}^k \qquad (5)$$

Indeed, by Birkhoff von Neumann theorem, any set D^k can be scheduled [13] in a time window of $\|D^k\|_{IO}$ slots, since D^k can be decomposed in $\|D^k\|_{IO}$ switching configurations.

D. Traffic Class

In our context, we consider only controlled traffic, since it is the only one for which it is possible to guarantee delay bounds in an OQ switch architecture. We consider here only two kinds of controlled traffic: regulated and leaky bucket constrained traffic. Since at most one packet arrives per time slot, the following property holds when the arrivals are observed at the inputs:

$$\bar{A}_i \leq F \qquad (6)$$

D.1 Regulated traffic

The following definition is derived by the adversary queueing theory [14].

Definition 5. An arrival process $\mathcal{A}$ is (ρ, W)-*regulated* if:

$$\left\| \sum_{z=t}^{t+W-1} A(t) \right\|_{IO} \leq \rho W \qquad \forall t$$

i.e., at most ρW packets arrive during each interval of W time slots for each input-output couple. W is called "admissibility window".

We can say that a (ρ, W)-regulated traffic injects at most ρW packets during an admissibility window W, corresponding to a maximum average rate ρ for each input-output couple during the same window W. Furthermore, an arrival process (ρ, W)-regulated is also $(1, \rho W)$-regulated, but not viceversa. In other words, the family of all the possible arrival processes (ρ, W)-regulated is a subset of the bigger family of processes $(1, \rho W)$-regulated.

We focus on $(1, nF)$-regulated arrival processes for which it holds:

$$\left\| \sum_{z=k}^{k+n-1} A^z \right\|_{IO} \leq nF \qquad (7)$$

D.2 Leaky bucket constrained traffic

This second kind of source is the usual $[\rho, \sigma]$ leaky bucket constrained source ($[\rho, \sigma]$-LBC). We refer [15] for a detailed definition of this source.

III. Properties of F-work conserving policies

Property 1. *A policy $\mathcal{D}$ is F-work-conserving in an observation window of size F with speedup S if:*

$$\hat{B}_j^{k+1} \leq \lceil \hat{B}_j^k + \hat{A}_j^k + O_j^k - F \rceil^+ \qquad \forall k, j \qquad (8)$$

To understand the meaning of this property, start to consider the case $F = 1$. Eq. (8) means that if at least a packet is present at the input ports destined for output j, this (single) packet should be transferred to the output queue j, provided that no packet at the output queue j is present. For a generic F, Eq. (8) implies that, if at least $F - O_j^k$ packets are present at the input ports destined for output j, these packets should be transferred to the output queue j.

For F-work-conserving policies we state the following theorem:

Theorem 2. *Assume that policy $\mathcal{D}$ is F-work-conserving and the arrival process $\mathcal{A}$ is $(1, nF)$-regulated. If $Y_j^k > 0$ then:*

$$\exists n_0 : 0 \leq n_0 < n, \; Y_j^k Y_j^{k+n_0} = 0$$

i.e., there exists a k' close to k (that is, $k' - k < n$) such that $Y_j^{k'} = 0$.

We omit the proof for lack of space, the interested reader can find it in [16].

Note that Theorem 2 implies that the maximum delay experienced by packets of an $(1, nF)$-regulated arrival process in a CIOQ switch with an F-work-conserving policy is not greater than nF slots.

We now show one possible example of F-work-conserving policy:

Lemma 1. *The following policy $\mathcal{D}$:*

$$D_{ij}^k = (A_{ij}^k + B_{ij}^k) \min\left\{1, \frac{\theta F - \gamma O_j}{\hat{A}_j^k + \hat{B}_j^k}\right\} \qquad \forall i, j, k$$

is F-work-conserving for $\theta \geq 1$ and $0 \leq \gamma \leq 1$.

Proof. If $\hat{A}_j^k + \hat{B}_j^k \leq \theta F - \gamma O_j$ then $\hat{B}_j^{k+1} = 0$ and $\hat{D}_j^k = \hat{B}_j^k + \hat{A}_j^k$. Otherwise, if $\hat{A}_j^k + \hat{B}_j^k > \theta F - \gamma O_j$ then $\hat{B}_j^{k+1} = \hat{B}_j^k + \hat{A}_j^k - \theta F + \gamma O_j > 0$ and $\hat{D}_j^k = \theta F - \gamma O_j$. Hence, if $\theta \geq 1$ and $\gamma \in [0, 1]$:

$$\hat{B}_j^{k+1} \leq \lceil \hat{B}_j^k + \hat{A}_j^k - \theta F + \gamma O_j \rceil^+ \leq$$
$$\leq \lceil \hat{B}_j^k + \hat{A}_j^k - F + \gamma O_j \rceil^+ \leq \lceil \hat{B}_j^k + \hat{A}_j^k - F + O_j \rceil^+$$

and the policy $\mathcal{D}$ is F-work-conserving. □

Policy $\mathcal{D}$, to be feasible with the speedup S, satisfies the following relation, derived from Eq. 5, referred as *feasibility condition*: $SF \geq \|D^k(\theta, \gamma)\|_{IO}, \ \forall k$

Intuitively, policy $\mathcal{D}$, with $\gamma = 0$, is greedy, since it transfers completely all the backlogged packets if compatible with the available output bandwidth θF. Otherwise, the output bandwidth is distributed among all the inputs proportionally to the number of backlogged packets.

IV. On the minimum speedup under regulated traffic

The following three theorems are our main results. The first one is quite trivial and intuitive, but can be significant.

Theorem 3. *Consider a CIOQ switch. Under an arrival process $\mathcal{A}$ which is $(1, W)$-regulated, there exists a W-work conserving policy when $S \geq 1$.*

Proof. Fix the observation window size $F = W$. Consider the following policy:

$$D_{ij}^k = (A_{ij}^k + B_{ij}^k) \min\left\{1, \frac{F}{\hat{A}_j^k + \hat{B}_j^k}\right\}$$

We know, from Lemma 1, that it is F-work-conserving (in this the case, $\theta = 1$ and $\gamma = 0$). Now we will prove that it is feasible for $S \geq 1$. Thanks to Theorem 2, we can assume, for all k:

$$Y_j^k = 0 \Rightarrow B_{ij}^k = 0 \ \forall i \qquad \text{and} \qquad O_j^k = 0$$

By assumption, $\hat{A}_j^k \leq F$ and $\bar{A}_i^k \leq F$. Hence, the policy reduces to: $D_{ij}^k = A_{ij}^k$ and by imposing $\|D^k\|_{IO} \leq SF$, we obtain: $S \geq 1$. □

Theorem 4. *Consider a CIOQ switch. Under an arrival process $\mathcal{A}$ which is $(1, W)$-regulated, there exists a $W/2$-work conserving policy if and only if $S \geq 4/3$.*

Proof. Fix the observation window size $F = W/2$. We divide the proof in two steps, in the first we show that $S = 4/3$ is a sufficient speedup to deal with $(1, 2F)$-regulated traffic, in the second step we show that it is also a necessary condition. Note that in this case, $\mathcal{D}$ is also the optimal policy, minimizing the speedup needed.

Step 1. Fix $\theta_0 = 4/3$ and consider the following policy $\mathcal{D}$:

$$D_{ij}^k = (A_{ij}^k + B_{ij}^k) \min\left\{1, \frac{\theta_0 F}{\hat{A}_j^k + \hat{B}_j^k}\right\}$$

We know, from Lemma 1, that $\mathcal{D}$ is F-work-conserving (in this case, $\gamma = 0$ and $\theta = \theta_0$), hence it is a good representative for $\mathcal{D}$. We show now that $\mathcal{D}$ is feasible for $S \geq 4/3$. First we notice that, in general:

$$\hat{D}_j^k = \sum_i D_{ij}^k = \min\{\hat{A}_j^k + \hat{B}_j^k, \theta_0 F\} \leq \theta_0 F \leq SF$$

with $S \geq 4/3$. Thus, to decide the feasibility of $\mathcal{D}$, we have to compute the maximum possible value for $\bar{D}_i^k$. $\bar{D}_i^k$ can be split in two components, $\bar{D}_{i,A}^k$ which is the amount of services received by packets arrived during the k^{th} observation window at input i, and $\bar{D}_{i,B}^k$ is the amount of services received by backlogged packets from the previous observation window at input i: $\bar{D}_i^k = \bar{D}_{i,A}^k + \bar{D}_{i,B}^k$. It is $\bar{D}_{i,A}^k \leq F$ because of (6). We now find the maximum for $\bar{D}_{i,B}^k$. Note that if $\hat{D}_{j,B}^k > 0$ then $\hat{B}_j^k > 0$, being $\hat{D}_{j,B}^k$ the amount of service received by backlogged packets at output j. Then, $\hat{B}_j^{k-1} = 0$ and $D_{ij,B}^k = B_{ij}^k$ for Theorem 2.

$$\sum_j B_{ij}^k = \sum_j A_{ij}^{k-1} \left(1 - \min\left\{1, \frac{\theta_0 F}{A_j^{k-1}}\right\}\right) \leq$$
$$\leq \sum_j A_{ij}^{k-1} \left(1 - \min\left\{1, \frac{\theta_0 F}{2F}\right\}\right) \leq F(1 - \theta_0/2)$$

thanks to the fact that $A_j^{k-1} \leq 2F$. Thus, after maximizing $\bar{D}_{i,B}^k$, we can maximize $\bar{D}_i^k$ and imposing the feasibility conditions:

$$\bar{D}_i^k \leq F + F(1 - \theta_0/2) = \frac{4}{3}F \leq SF$$

which holds for $S \geq 4/3$.

In conclusion, with speedup $S \geq 4/3$ policy $\mathcal{D}$ is feasible.

Step 2. We want to show, by a counterexample, that the minimum speedup $4/3$ is also necessary to have an F-work-conserving policy. Consider a switch with 2 active inputs and 3 outputs. Assume $Y_j^k = 0$, hence $B_{ij}^k = 0$ for $1 \leq i \leq 2$ and $1 \leq j \leq 3$. Consider the following traffic pattern, $(1, 2F)$-regulated: $A_{11}^k = A_{21}^k = A_{12}^{k+1} = A_{23}^{k+1} = F$. At the end of the k^{th} observation window, to minimize the maximum backlog at both inputs, we set: $D_{11}^k = D_{21}^k = SF/2$.

After the arrival at time $k+1$, there are $(1 - S/2)F$ packets enqueued at the inputs and destined to output 1, whereas F are

Minimum speedup		Average delay	Maximum
sufficient	necessary	penalty w.r. OQ	delay
$S = 1$	$S = 1$	$3/2 \times \rho W$	$2 \times \rho W$
$S = 4/3$	$S = 4/3$	$3/4 \times \rho W$	$3/2 \times \rho W$
$S = 3/2$	-	$1/2 \times \rho W$	$4/3 \times \rho W$
$S = 2$	$S = 2$	0	ρW

TABLE I

TRADEOFF BETWEEN SPEEDUP, THE AVERAGE DELAY PENALTY WITH RESPECT TO AN OQ SWITCH AND MAXIMUM DELAY FOR A (ρ, W)-REGULATED TRAFFIC

Minimum speedup		Average delay	Maximum
sufficient	necessary	penalty w.r. OQ	delay
$S = 1$	$S = 1$	$3/2 \times \sigma/(1-\rho)$	$2 \times \sigma/(1-\rho)$
$S = 4/3$	$S = 4/3$	$3/4 \times \sigma/(1-\rho)$	$3/2 \times \sigma/(1-\rho)$
$S = 3/2$	-	$1/2 \times \sigma/(1-\rho)$	$4/3 \times \sigma/(1-\rho)$
$S = 2$	$S = 2$	0	$\sigma/(1-\rho)$

TABLE II

TRADEOFF BETWEEN SPEEDUP, THE AVERAGE DELAY PENALTY WITH RESPECT TO AN OQ SWITCH AND MAXIMUM DELAY FOR A $[\rho, \sigma]$-LBC TRAFFIC

destined to output 2 and 3. Hence, to have $\mathcal{D}$ work-conserving by setting $Y_j = 0$ and $B_{ij}^{k+2} = 0$: $D_{ij}^{k+1} = B_{ij}^{k+1} + A_{ij}^{k+1}$. Since D_{ij}^{k+1} must be feasible, we impose:

$$\frac{(2-S)F}{N} + \frac{2(2N-1)F}{3N} \leq SF \qquad \Rightarrow \qquad S \geq \frac{4}{3}$$

Hence, $S \geq 4/3$ is a necessary condition to have an F-work-conserving policy. □

Theorem 5. *Consider a CIOQ switch. Under an arrival process $\mathcal{A}$ which is $(1, W)$-regulated, there exists a $W/3$-work conserving policy, if $S \geq 3/2$.*

We omit the proof for lack of space, the interested reader can find it in [16].

V. MAIN RESULTS ABOUT DELAY PERFORMANCE

Under a $(1, nF)$-regulated arrival process, Theorems 3, 4 and 5 evaluate the compromise between speedup and average delay penalty with respect to an OQ switch, which is $3/2 \times F$. Indeed, the average delay penalty is sum of two contributions. The first is the average delay penalty equal to F due to the F-work-conserving property (see Theo. 1). The second is an additional average penalty equal to $F/2$ due to the switch working in a gated-fashion (see Eq. 2). On the contrary, the absolute delay is $nF + F$, thanks to the observation at the end of Theorem 2.

Now consider an arrival process (ρ, W)-regulated and an arrival process $[\rho, \sigma]$-LBC. Tables I and II show the average delay penalty with respect to OQ and the absolute delay, for regulated and LBC traffic. Note that, for $n > 3$, we did not compute the minimum speedup. Of course, with speedup $S = 2$, a CIOQ system can emulate perfectly an OQ and the average delay penalty is null.

VI. CONCLUSIONS

CIOQ switches that can control the packet delays at low speedups are very appealing. It is well known that, at speedup lower than 2, a CIOQ switch can not emulate OQ switch even with bounded delay penalty [3]. Hence, we considered the CIOQ switch operating under a restricted, but general enough, arrival traffic class. We defined a new notion of F-work conservation for CIOQ switches, which in turn implies the property of OQ emulation with average delay penalty bounded by F. Under regulated traffic, we were able to compute an upper bound of the delay penalty for $S = 1$, $S = 4/3$ and $S = 3/2$. We presented scheduling policy for $S = 4/3$ and $S = 3/2$. Thus, we showed that it is possible to emulate OQ switch under quite a general class of arrival traffic at lower speedup than 2 with bounded amount of average delay penalty.

REFERENCES

[1] McKeown N., Mekkittikul A., Anantharam V., Walrand J., "Achieving 100% throughput in an input-queued switch", *IEEE Transactions on Communications*, vol. 47, n. 8, Aug. 1999, pp. 1260-1267

[2] Dai J., Prabhakar B., "The throughput of data switches with and without speedup", *IEEE INFOCOM 2000*, vol. 2, Tel Aviv, Israel, Mar. 2000, pp. 556-564

[3] Chuang S.T., Goel A., McKeown N., Prabhakar B. "Matching output queueing with a combined input/output-queued switch", *IEEE Journal on Selected Areas in Communications*, vol. 17, n. 6, Jun. 1999, pp. 1030-39

[4] "Cisco 12000 Gigabit Switch Router", Product Overview, www.cisco.com, Apr. 2000

[5] Partridge C., et al., "A 50-Gb/s IP router", *IEEE Transactions on Networking*, vol. 6, n. 3, June 1998, pp. 237-248

[6] Tassiulas L., "Linear complexity algorithms for maximum throughput in radio networks and input queued switches", *IEEE INFOCOM'98*, vol. 2, New York, NY, 1998, pp. 533-539

[7] Leonardi E., Mellia M., Neri F., Ajmone Marsan M., "Bounds on Average Delays and Queue Size Averages and Variances in Input Queued Cell-Based Switches", *IEEE INFOCOM 2001*, Anchorage, AK, vol. 3, Apr. 2001, pp.1095-1103

[8] Shah D., Kopikare M., "Delay bounds for the approximate Maximum weight matching algorithm for input queued switches", *IEEE INFOCOM 2002*, New York, NY, USA, June 2002

[9] Ajmone Marsan M., Bianco A., Filippi E., Giaccone P., Leonardi E., Neri F., "On the behavior of input queuing switch architectures", *European Transactions on Telecommunications*, vol. 10, n. 2, Mar. 1999, pp. 111-124

[10] Giaccone P., Prabhakar B., Shah D., "Towards simple, high-performance schedulers for high-aggregate bandwidth switches ", *IEEE INFOCOM'02*, New York, NY, Jun. 2002

[11] Ajmone Marsan M., Leonardi E., Mellia M., Neri F., "On the stability of Input-Queued Switches with Speed-up", *IEEE/ACM Transactions on Networking*, vol. 9, n. 1, pp.104-118, Feb. 2001

[12] Krishna P., Patel N.S., Charny A., Simcoe R.J., " On the speedup required for work-conserving crossbar switches", *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 6, Jun. 1999, pp. 1057-1066

[13] Weller T., Hajek B., "Scheduling nonuniform traffic in a packet-switching system with small propagation delay", *IEEE/ACM Trans. on Networking*, vol. 5, n. 6, Dec. 1997, pp. 813-823

[14] Borodin A., Klejnberg J., Raghavan P., Sudan M., "Aversarial queueing theory", Sept. 1998

[15] Le Boudec J.Y., Thiran P., "Network calculus: a theory of deterministic queuing systems for the Internet", *Springer Publishing Company*, Jul. 2001

[16] P. Giaccone, E. Leonardi, B. Prabhakar, D. Shah, "Delay Prformance of High-Speed Packets Switches with Low Speed-up", internal report available on-line at: http://www.tlc-networks.polito.it/~emilio/db_papers/tec_rep/dspeedup.ps