Bounds on Average Delays and Queue Size Averages and Variances in Input-Queued Cell-Based Switches

E. Leonardi, M. Mellia, F. Neri, M. Ajmone Marsan

Abstract—In this paper we develop a general methodology, mainly based upon Lyapunov functions, to derive bounds on average delays, and on queue size averages and variances of complex systems of queues. We then apply this methodology to input-buffered, cell-based switch and router architectures. These architectures require a scheduling algorithm to select at each slot a subset of input-buffered cells which can be transferred towards output ports. Although the stability properties (i.e., the limit throughput) of input-buffered, cell-based switches was already studied for several classes of scheduling algorithms, no analytical results concerning cell delays or queue sizes are yet available in the technical literature. We concentrate on purely input-buffered switches that adopt a Maximum Weight Matching scheduling algorithm, that was proved to be the scheduling algorithm providing the best performance. The derived bounds proved to be rather tight, when compared to simulation results.

Keywords—Input Queued Switch; Performance Evaluation; Scheduling

I. INTRODUCTION

A number of high-performance IP routers (for example, the CISCO 12000 [1], the Lucent Versalar TSR45000 [3]) are built around fast cell-based switching fabrics: input IP datagrams are segmented into fixed-size data units, usually called cells, and transferred to output ports, where IP datagrams are re-assembled. The design of these high-performance routers generally does not adopt the classical output-queued (OQ) architecture (where cells are stored at the output of the switching fabric), preferring either input-queued (IQ) or combined input/output queued (CIOQ) structures. The reason is that, in OQ, both the switching fabric and the output (and possibly input) queues in line cards must operate at a speed equal to the sum of the rates of all input lines; since this speed grows linearly with the number of switch ports, the OQ approach is impractical for large, fast switches. Instead, in IQ schemes, all the components of the switch (input interfaces, switching fabric, output interfaces) can operate at a data rate which is compatible with the data rate of input and output lines, and which does not grow with the switch size. A traditional performance penalty of IQ architectures is due to head-of-the-line blocking in the case of a single queue per input interface [4], but can be largely reduced by Virtual Output Queuing (VOQ) (also called Destination Queuing) schemes [5], which organize input buffers in each line card into a set of queues where cells awaiting access to the switching fabric are stored according to their destination output cards.

A major issue in the design of IQ switches is that the access to the switching fabric must be controlled by some form of scheduling algorithm¹, which operates on a (possibly partial) knowledge of the state of input queues. This means that control information must be exchanged among line cards, either through an additional data path or through the switching fabric itself, and that intelligence and computational complexity must be devoted to the scheduling algorithm, either at a centralized scheduler, or at line cards, in a distributed manner.

The problem faced by the scheduling algorithm can be formalized as a maximum size or maximum weight matching on the bipartite graph in which nodes represent input and output ports, and edges represent cells to be switched. The optimal solution to this problem is known to be obtained with the Maximum Weight Matching (MWM) algorithm [8], but several lower complexity scheduling algorithms for IQ cell switches were also proposed and studied in the recent literature [5], [9], [10], [11], [12], [13], [14], [15], [16], [17]. MWM was also proved to sustain the same throughput of OQ switches [5].

To simplify the design and implementation of scheduling algorithms, often switches operate with a speed-up, i.e., the internal switching fabric, as well as the input and output memories, operate at a higher speed with respect to the data rate of input/output lines. In this case, buffering is required at outputs as well as inputs, and the term “combined input/output queueing” (CIOQ) is used. Obviously, when the speed-up is such that the internal switch bandwidth equals the sum of the data rates on input lines, input buffers become useless. In [18], [19], [20], [21], a speed-up equal to 2 in CIOQ switches, independent of the number of switch ports, was shown to be sufficient to exactly emulate an OQ architecture, at the expense of quite complex scheduling algorithms, whose implementation seems problematic. Previous papers [5], [22], [23], [24] proved that a wide class of simpler maximal size scheduling algorithms (comprising well-known algorithms, such as i-SLIP [10] and 2DRR [15]) provides the same throughput performance of OQ when a speed-up equal to 2 is available in the switch (although the behavior of an OQ architecture is not exactly emulated).

These results concerning the throughput of IQ and CIOQ switches provide a solid theoretical background, which is very useful to designers of high-speed switches and routers, that will be a major ingredient of future telecommunication infrastructures. However, throughput is only one of the key performance parameters used in the design and planning of communication networks; the other major performance metrics is delay. For a wide class of telecommunication services, in particular those with real-time requirements, delay is even more important than throughput. Hence, extending the theoretical analysis of IQ switches to delays is extremely important. Computing cell delays in cell-based IQ switches is not easy; indeed, to the best

¹The authors are with the Dipartimento di Elettronica, Politecnico di Torino, Italy. Email: {leonardi,mellia,neri,ajmone}@mail.elet.polito.it. This work was supported by the Italian Ministry for University and Scientific Research through the MQoS Project.

²The term “scheduling algorithm” for switching architectures is used in the literature for two different types of schedulers: switching matrix schedulers and flow-level schedulers [6], [7]. Switching matrix schedulers decide which input port is enabled to transmit in a non purely output-queued switch; they avoid

blocks and solve contentions within the switching fabric. Flow-level schedulers decide which cells must be served in accordance to QoS requirements. In this paper the term scheduling algorithm is only used to refer to the first class of algorithms.
of our knowledge, no analytical results are yet available for the delays faced by cells traversing an IQ switch.

In this paper we consider MWM scheduling algorithms, focusing on the case of uniform traffic, and we derive for the first time analytical bounds on average delays, and on average and variance of the size of the queues at switch inputs. After introducing some definitions and preliminary results, in Section II we provide the main results of our analysis methodology, which we then apply to the derivation of bounds on average delays and queue sizes of IQ switches in Section III, and on queue size variances in Section IV. In Section V we show that our analytical bounds match well with simulation results. Section VI provides some concluding remarks.

The availability of tight bounds on delays and queue sizes is quite important for switch and router designers: they allow the quantification of the buffer storage required for the provision of QoS guarantees, in terms of both loss probability, average delay, and delay jitter. The fact that our bounds are expressed in closed form, makes them extremely easy to compute and apply to realistic design problems.

II. ANALYSIS METHODOLOGY

In this section, we first briefly recall some previously published results that form the background necessary for our analysis methodology; then we derive the new results that lead to the delay and queue size bounds.

Given a system of $N$ discrete-time queues of infinite capacities, let $X_n$ be the row vector of queue lengths at time $n$; i.e., $X_n = (x_n^1, x_n^2, \ldots, x_n^N)$, where $x_n^i$ is the number of customers in queue $i$ at time $n$.

The queue length evolution is described by the expression

$$x_{n+1}^i = x_n^i + a_n^i - d_n^i,$$

where $a_n^i$ represents the number of customers arrived at queue $i$ in time interval $(n, n+1)$, and $d_n^i$ represents the number of customers departed from queue $i$ in time interval $(n, n+1)$. Let $A_n = (a_n^1, a_n^2, \ldots, a_n^N)$ be the vector of the numbers of arrivals at the different queues, and $D_n = (d_n^1, d_n^2, \ldots, d_n^N)$ be the vector of the numbers of departures from queues. With this notation, the evolution of the system of queues can be described as

$$X_{n+1} = X_n + A_n - D_n \quad (1)$$

We assume that vectors $A_n$ are independent and identically distributed. However, correlation between different components of the arrival vector $A_n$ is admissible. Extensions to other interesting traffic scenarios, such as batch arrivals or unbalanced traffic distributions, although not presented in this paper, are rather straightforward.

We indicate with $||Y||$ the norm of vector $Y = (y_1, y_2, \ldots, y_K)$. Three different norms will be introduced later in Definitions 3, 4, and 5.

Definition 1: A system of queues is said to be weakly stable if, for every $\epsilon > 0$, there exists $B > 0$ such that

$$\lim_{n \to \infty} \mathbb{P}\{|X_n| > B\} < \epsilon$$

Definition 2: A system of queues is said to be strongly stable if

$$\lim_{n \to \infty} \sup_{Y_n} E[||X_n||] < \infty$$

(where $E[X]$ is the average of random variable $X$).

We assume that the process describing the evolution of the system of queues is an irreducible Discrete Time Markov Chain (DTMC), whose state vector at time $n$ is $Y_n = (X_n, K_n)$. $Y_n \in \mathbb{N}^M$, $X_n \in \mathbb{N}^N$, $K_n \in \mathbb{N}^M$, and $M = N + N^r$. $Y_n$ is the combination of vector $X_n$ and a vector $K_n$ of $N^r$ integer parameters.

Let $H$ be the state space of the DTMC, obtained as a subset of the Cartesian product of the state space $H_\infty$ of $X_n$ and the state space $H_K$ of $K_n$.

If all states $Y_n$ are positive recurrent, the system of queues is weakly stable; however, the converse is generally not true, since queue lengths can remain finite even if the states of the DTMC are not positive recurrent due to instability in the sequence of parameter vectors $\{K_n\}$.

Note that most systems of discrete-time queues of practical interest can be described with models that fall in the DTMC class. The following general criterion for the (weak) stability of systems falling into this class is therefore useful.

Theorem 1: Given a system of queues whose evolution is described by a DTMC with state vector $Y_n \in \mathbb{N}^M$, if a lower bounded function $V(Y_n)$, called Lyapunov function, $V: \mathbb{N}^M \to \mathbb{R}$, can be found such that

$$E[V(Y_{n+1}) | Y_n] < \infty \quad \forall Y_n$$

and there exist $\epsilon \in \mathbb{R}^+$ and $B \in \mathbb{R}^+$ such that

$$E[V(Y_{n+1}) - V(Y_n) | Y_n] < -\epsilon \quad \forall |Y_n| > B \quad (2)$$

then all states of the DTMC are positive recurrent, and the system of queues is weakly stable.

Proof: This theorem is a straightforward extension of Foster’s Criterion; see [25], [26], [27].

If the state space $H$ of the DTMC is a subset of the Cartesian product of the denumerable state space $H_\infty$ and a finite state space $H_K$, the stability criterion can be slightly modified, since the stability of the system can be inferred only from the queue length state vector $X_n$.

Corollary 1: Given a system of queues whose evolution is described by a DTMC with state vector $Y_n \in \mathbb{N}^M$, and whose state space $H$ is a subset of the Cartesian product of a denumerable state space $H_\infty$ and a finite state space $H_K$, if a lower bounded function $V(X_n)$, called Lyapunov function, $V: \mathbb{N}^N \to \mathbb{R}$, can be found such that

$$E[V(X_{n+1}) | Y_n] < \infty \quad \forall Y_n$$

and there exist $\epsilon \in \mathbb{R}^+$ and $B \in \mathbb{R}^+$ such that

$$E[V(X_{n+1}) - V(X_n) | Y_n] < -\epsilon \quad \forall |X_n| > B$$

then all states of the DTMC are positive recurrent.

In this case the system is weakly stable iff all states of the DTMC are positive recurrent.

\[2\] In this paper $\mathbb{N}$ denotes the set of non negative integers, $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}^+$ denotes the set of non negative real numbers.
In the remainder of this paper we restrict our analysis to the class of systems of queues for which Corollary 1 applies. We also assume an aperiodic DTMC since our arrival processes normally confirm this assumption, hence having positive recurrent states is equivalent to having ergodicity of the DTMC. The generalization to periodic DTMCs is straightforward.

As an extension to the results above, we recall from [22] the following criterion for strong stability.

**Theorem 2:** Given a system of queues whose evolution is described by a DTMC with state vector $Y_n \in \mathbb{N}^M$, and whose state space $H$ is a subset of the Cartesian product of a denumerable state space $H_X$ and a finite state space $H_K$, if a lower bounded function $V(X_n)$, called Lyapunov function, $V : \mathbb{N}^N \to \mathbb{R}$, can be found such that

$$E[V(X_{n+1}) \mid Y_n] < \infty \quad \forall Y_n$$

and there exist $\epsilon \in \mathbb{R}^+$ and $B \in \mathbb{R}^+$ such that

$$E[V(X_{n+1}) - V(X_n) \mid Y_n] < -\epsilon \|X_n\| \quad \forall Y_n : \|X_n\| > B$$

then the system of queues is strongly stable.

**Proof:** Since the assumptions of Theorem 1 are satisfied, every state of the DTMC is positive recurrent. In addition, to prove that the system is strongly stable, we shall show that

$$\lim_{n \to \infty} E[\|X_n\|] < \infty$$

Let $H_B$ be the set of values taken by $Y_n$ for which $\|X_n\| \leq B$ (where (3) does not apply). It is easy to prove that $H_B$ is a compact set. Outside this compact set, Equation (3) holds, i.e.

$$E[V(X_{n+1}) - V(X_n) \mid Y_n] < -\epsilon \|X_n\|$$

Considering all $Y_n$’s that do not belong to $H_B$, taking the average of the previous expression, we have:

$$E[V(X_{n+1}) \mid Y_n \notin H_B] - E[V(X_n) \mid Y_n \notin H_B] < -\epsilon E[\|X_n\| \mid Y_n \notin H_B]$$

Instead, for $Y_n \in H_B$, being $H_B$ a compact set:

$$E[V(X_{n+1}) \mid Y_n \in H_B] \leq M < \infty$$

where $M$ is the maximum taken by $E[V(X_{n+1}) \mid Y_n]$ for $Y_n \in H_B$.

By combining the two previous expressions, we obtain:

$$E[V(X_{n+1})] = E[V(X_{n+1}) \mid Y_n \in H_B] P\{Y_n \in H_B\} + E[V(X_{n+1}) \mid Y_n \notin H_B] P\{Y_n \notin H_B\} <$$

$$< \epsilon M P\{Y_n \in H_B\} + E[V(X_n)] - \epsilon E[\|X_n\| \mid Y_n \notin H_B] <$$

where $M_0$ is a constant such that

$$M_0 > \{ -E[V(X_n) \mid Y_n \in H_B] + \epsilon E[\|X_n\| \mid Y_n \in H_B]\} P\{Y_n \in H_B\}$$

Note that $M_0$ is finite, being $H_B$ a compact set.

By summing over all $n$ from 0 to $N_0 - 1$, we obtain:

$$E[V(X_{N_0})] < N_0 M + E[V(X_0)] - \epsilon \sum_{n=0}^{N_0-1} E[\|X_n\|] + N_0 M_0$$

Thus, for any $N_0$, we can write

$$\frac{\epsilon}{N_0} \sum_{n=0}^{N_0-1} E[\|X_n\|] < M + \frac{1}{N_0} E[V(X_0) - V(X_{N_0})] + M_0$$

$$E[V(X_{N_0})]$$ is lower bounded by definition; assume $E[V(X_{N_0})] > K_0$. Hence

$$\frac{\epsilon}{N_0} \sum_{n=0}^{N_0-1} E[\|X_n\|] < M + \frac{1}{N_0} E[V(X_0)] - \frac{K_0}{N_0} + M_0$$

For $N_0 \to \infty$, being $E[V(X_0)]$ and $K_0$ finite, we can write

$$\frac{\epsilon}{N_0} \sum_{n=0}^{N_0-1} E[\|X_n\|] < M + M_0$$

Hence

$$\lim_{N_0 \to \infty} \frac{\epsilon}{N_0} \sum_{n=0}^{N_0-1} E[\|X_n\|]$$

is bounded. Since the DTMC $Y_n$ has positive recurrent states, there exists $\lim_{n \to \infty} E[\|X_n\|]$. Furthermore, it is easy to show that, if the sequence $E[\|X_n\|]$ is convergent, the sequence

$$\frac{1}{N} \sum_{i=0}^{N-1} E[\|X_i\|]$$

converges to the same limit, i.e.:

$$\lim_{N \to \infty} E[\|X_n\|] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} E[\|X_i\|]$$

But the right hand side was seen to be bounded; hence

$$\lim_{n \to \infty} E[\|X_n\|] < \infty$$

Note that (4) provides a first bound on the limit behavior of $E[\|X_n\|]$:}

$$\lim_{n \to \infty} E[\|X_n\|] \leq \frac{1}{\epsilon} (M + M_0)$$

Unfortunately, this bound often is very loose; thus, we will obtain a novel tighter bound by selecting a special class of Lyapunov functions.

**Corollary 2:** Given, as in Theorem 2, a system of queues whose evolution is described by a DTMC with state vector $Y_n \in \mathbb{N}^M$, and whose state space $H$ is a subset of the Cartesian product of a denumerable state space $H_X$ and a finite state space $H_K$, if there exists a symmetric copositive matrix $^T Z \in \mathbb{R}^{N \times N}$ defining the function $V(X_n) = X_n Z X_n^T$, such that

$$E[V(X_{n+1}) \mid Y_n] < \infty \quad \forall Y_n$$

and there exist two positive real numbers $\epsilon \in \mathbb{R}^+$ and $B \in \mathbb{R}^+$, such that

$$E[V(X_{n+1}) - V(X_n) \mid Y_n] < -\epsilon \|X_n\| \quad \forall Y_n : \|X_n\| > B$$

$^3$An $N \times N$ matrix $Q$ is copositive if $XQX^T \geq 0$ for all $X \in \mathbb{R}^{+N}$. 
then the system of queues is strongly stable. In addition, all the polynomial moments of the queue length distribution are finite.

Proof: This is a re-phrasing of the results presented in [28, Sect.IV]. Readers are again referred to such stable publications for a proof.

Since the DTMC describing the evolution of a stable system of queues has positive recurrent states, if we assume aperiodicity as above, the DTMC is ergodic; hence \( \lim_{n \to \infty} E[X_{n+1}] = \lim_{n \to \infty} E[X_n] \). Under the assumptions of Corollary 2, due to the fact that the moments of \( X_n \) are finite, we can further state that:

\[
\lim_{n \to \infty} E[V(X_{n+1}) - V(X_n)] = 0 \tag{7}
\]

The proof of Theorem 2 shows that \( \lim_{n \to \infty} E[||X_n||] < \infty \). If we substitute in (3) \( -\epsilon ||X_n|| \) with \( -\epsilon f(||X_n||) \), where \( f(\cdot) \) is a continuous function defined on \( \mathbb{R}^+ \), following the steps of the proof of Theorem 2, it is possible to show that \( \lim_{n \to \infty} E[f(||X_n||)] < \infty \). We can therefore state the following theorem.

**Theorem 3:** Given, as in Theorem 2, a system of queues whose evolution is described by a DTMC with state vector \( Y_n \in \mathbb{N}^M \), and whose state space \( H \) is a subset of the Cartesian product of a denumerable state space \( H_X \) and a finite state space \( H_K \), if a lower bounded function \( V(X_n), V : \mathbb{N}^N \to \mathbb{R} \), can be found, such that

\[
E[V(X_{n+1}) | Y_n] < \infty \quad \forall Y_n
\]

and there exist two positive real numbers \( \epsilon \in \mathbb{R}^+ \) and \( B \in \mathbb{R}^+ \), such that

\[
E[V(X_{n+1}) - V(X_n) | Y_n] < -\epsilon f(||X_n||) \quad \forall Y_n : ||X_n|| > B
\]

being \( f(x) \) a continuous function in \( \mathbb{R}^+ \), then

\[
\lim_{n \to \infty} E[f(||X_n||)] < \infty
\]

It is now possible to state the following important theorem that provides a stronger and more general bound than (5).

**Theorem 4:** Given a system of queues whose evolution is described by a DTMC with state vector \( Y_n \in \mathbb{N}^M \), whose state space \( H \) is a subset of the Cartesian product of a denumerable state space \( H_X \) and a finite state space \( H_K \), and for which all the polynomial moments of the queue length distribution are finite, if a lower bounded polynomial function \( V(X_n), V : \mathbb{N}^N \to \mathbb{R} \), can be found, such that

\[
E[V(X_{n+1}) | Y_n] < \infty \quad \forall Y_n
\]

and there exist two positive real numbers \( \epsilon \in \mathbb{R}^+ \) and \( B \in \mathbb{R}^+ \), such that

\[
E[V(X_{n+1}) - V(X_n) | Y_n] < -\epsilon f(||X_n||) \quad \forall Y_n : ||X_n|| > B
\]

being \( f(x) \) a continuous function in \( \mathbb{R}^+ \), then

\[
\lim_{n \to \infty} E[f(||X_n||)] \leq \lim_{n \to \infty} E\left[f(||X_n||) + \frac{V(X_{n+1}) - V(X_n)}{\epsilon} | Y_n \in H_B\right] P\{Y_n \in H_B\}
\]

\[
\lim_{n \to \infty} E[V(X_{n+1}) - V(X_n)] = 0 \tag{10}
\]

Following the steps of the proof of Theorem 2, and using (10):

\[
E[V(X_{n+1}) - V(X_n)] =
= E[V(X_{n+1}) - V(X_n) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[V(X_{n+1}) - V(X_n) | Y_n \notin H_B] P\{Y_n \notin H_B\} \leq
\leq E[V(X_{n+1}) - V(X_n) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[V(X_{n+1}) - V(X_n) | Y_n \notin H_B] P\{Y_n \notin H_B\} =
= E[V(X_{n+1}) - V(X_n) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[V(X_{n+1}) - V(X_n) | Y_n \notin H_B] P\{Y_n \notin H_B\} +
+ E[f(||X_n||) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[f(||X_n||) | Y_n \notin H_B] P\{Y_n \notin H_B\} =
= E[V(X_{n+1}) - V(X_n) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[f(||X_n||) | Y_n \in H_B] P\{Y_n \in H_B\} +
+ E[f(||X_n||) | Y_n \notin H_B] P\{Y_n \notin H_B\}
\]

Combining (10) and (11), the theorem is proved.

This result will be used in the next sections, in which polynomial forms for \( V(X_n) \) and for \( f(||X_n||) \) will be used to derive bounds on average delays as well as on queue size averages and variances for input-queued cell-based switches.

**III. BOUNDS ON AVERAGE DELAYS AND QUEUE SIZES**

We consider Input Queueing cell-based Switches with \( P \) input ports and \( P \) output ports, all at the same cell rate (and we call them \( P \times P \) IQS). The switching fabric is assumed to be non-blocking and memoryless, i.e., cells are only stored at switch inputs.

At each input, cells are stored according to a Virtual Output Queue (VOQ) policy: one separate queue is maintained for each output. Thus, the total number of queues in the switch is \( N = P^2 \).

In order to simplify our formulas, in the rest of this paper we adopt a vector notation rather than a matrix notation to indicate input queues: let \( q^{(k)} \), \( k = Pi + j \) be the queue at input \( i \) storing cells directed to output \( j \), with \( i, j = 0, 1, 2, \ldots, P - 1 \).

Before proceeding, we need to define three norm functions. Recall that the norm of vector \( Z \in \mathbb{R}^N \) is a function such that:

\[
||Z|| \geq 0 \quad \text{and} \quad ||Z|| = 0 \quad \text{if} \quad Z = 0.
\]

\[
||\alpha Z|| = ||\alpha|| ||Z|| \quad \text{for every scalar} \quad \alpha,
\]

\[
||Z_1 + Z_2|| \leq ||Z_1|| + ||Z_2||.
\]

**Definition 3:** Given a vector \( Z \in \mathbb{R}^N \), \( Z = \{z^{(k)}\} \), \( k = Pi + j \), \( i, j = 0, 1, \ldots, P - 1 \), the norm \( ||Z||_\alpha \) is defined as:

\[
||Z||_\alpha = \max_{j=0,\ldots, P-1} \left\{ \sum_{i=0}^{P-1} |z^{(Pi+j)}|, \sum_{i=0}^{P-1} |z^{(Pi+j)}| \right\}
\]

(12)
Definition 4: Given a vector $Z \in \mathbb{R}^N$, the norm $||Z||_1$ is defined as:

$$||Z||_1 = \sum_{k=0}^{N-1} |z^{(k)}|$$  \hspace{1cm} (13)

Definition 5: Given a vector $Z \in \mathbb{R}^N$, the norm $||Z||_2$ is defined as:

$$||Z||_2 = \sqrt{\sum_{k=0}^{N-1} (z^{(k)})^2}$$  \hspace{1cm} (14)

The normalized vector parallel to vector $Z$ is denoted by $\hat{Z}$, and is defined as:

$$\hat{Z} = \frac{Z}{||Z||_2}$$

Let $R \in \mathbb{R}^N$, $R = \{r(k), \ k = Pi + j, \ i, j = 0, 1, \ldots, P - 1\} = E[A]$ be the vector of the average cell arrival rates at queues $q^{(k)}$ in cells/slot.

Definition 6: The traffic pattern loading an IQS is admissible iff

$$||R||_0 = ||E[A]||_0 < 1$$

In plain words, a traffic pattern is admissible if the total average arrival rates in cells/slot are less than 1 for all input and output ports.

At each time slot, the switch scheduler selects cells to be transferred from input queues to output queues. The set of cells to be transferred during one time slot must satisfy two constraints: i) at most one cell can be extracted from the VOQ structure at each input, and ii) at most one cell can be transferred towards each output.

Definition 7: At each time slot, the scheduler of an IQS selects for transfer from queues $q^{(k)}$ a set of cells denoted by vector $D \in \mathbb{N}^N$, $D = \{d^{(k)}, \ k = Pi + j, \ i, j = 0, 1, \ldots, P - 1\}$ so that:

$$||D||_0 \leq 1$$

Set $D$ is said to be a set of non-contending cells, or a switching vector.

The selection of a switching vector was proved equivalent to the matching problem in a weighted bipartite graph, where nodes represent input/output ports, arcs represent the presence of cells to be transferred, and weights indicate the metrics to be used by the algorithm. As we already recalled, the optimal solution to this problem is obtained with the MWM algorithm [8].

Definition 8: An IQS adopts a MWM scheduler if the selection of the switching vector at each time slot is performed according to the MWM algorithm.

Let $W_n$ represent a weight vector at time $n$, and let $D_n$ be the switching vector at time $n$. The departure vector $D_n$ produced by a MWM scheduler is such that:

$$D_nW_n^T = \max_{D_n} (D_nW_n^T)$$

If we set $W_n = X_n$, where $X_n$ is the vector of input queues at time $n$, i.e., if we use queue lengths as the scheduler metrics, the switching vector produced by a MWM scheduler is such that:

$$D_nX_n^T = \max_{D_n} (D_nX_n^T)$$

In [5], it was proved that for each vector $Y_n \in \mathbb{R}^N$ such that $||Y_n||_0 \leq 1$, and for each vector $X_n \in \mathbb{N}^N$:

$$D_nX_n^T - Y_nX_n^T \geq 0$$  \hspace{1cm} (15)

From this property, using the Lyapunov function $V(X_n) = X_nX_n^T$, it was then proved that, for any IQS using an MWM scheduler with $W_n = X_n$:

$$E[X_{n+1}X_{n+1}^T - X_nX_n^T | X_n] \leq -\epsilon||X_n||$$  \hspace{1cm} (16)

for $||X_n||$ sufficiently large. Note that this property holds whichever norm is considered, as long as $N < \infty$, being the asymptotic behavior of $||X_n||$ the same for all norm functions.

As a consequence, due to Corollary 2 with matrix $Z$ equal to the identity matrix $I$, we can state the following theorem:

Theorem 5: Under any admissible i.i.d. input traffic pattern, an IQS using an MWM scheduler with weights equal to queue lengths is strongly stable, and all polynomial moments of the queue length distribution are finite (for a proof, the reader is invited to refer to [5]).

Given this result, we can apply (9) to derive a bound on the average cell delay in the switch.

Let us first prove the following

Proposition 1: For any $X_n \in \mathbb{R}^+$:

$$D_nX_n^T \geq \frac{1}{P}$$

Proof: From (15), we get:

$$\forall Y_n \in \mathbb{R}^+ \quad \text{such that} \quad ||Y_n||_0 \leq 1, \quad D_nX_n^T \geq Y_nX_n^T$$

Considering $Y_n = \frac{1}{P} \mathbb{I}$ (note that with this choice $||Y_n||_0 = 1$), where $\mathbb{I}$ is the all-ones vector, i.e., $\mathbb{I} = \{1, 1, \ldots, 1\}$, we obtain:

$$\frac{1}{P} \mathbb{I}X_n^T = \frac{1}{P} \sum_{i=0}^{N-1} \hat{x}^{(i)} = \frac{1}{P}$$

We now introduce the assumption of uniform traffic: all switch input/output ports are equally loaded. This assumption can be removed, but the resulting bounds would in general be looser. Let $\rho \leq 1$ be the average load of each switch input port in cells/slot. Considering that at each slot a cell arrives at each input with probability $\rho$, and no cell arrives with probability $1 - \rho$, and that each input port comprises $P$ VOQs, the vector of the average arrival rates at slot $n$ is such that:

$$E[A_n] = \frac{\rho}{P} \mathbb{I}$$

$$E[A_nA_n^T] = \rho^2$$  \hspace{1cm} (17)

Since $E[D_n] = E[A_n] = \frac{\rho}{P} \mathbb{I}$ due to system stability, similarly to (17), it is possible, under the assumption of uniform traffic, to state that:

$$E[D_nD_n^T] = \rho^2$$  \hspace{1cm} (18)

Moreover, being $A_n$ independent from $D_n$:

$$E[A_nD_n^T] = E[A_n]E[D_n^T] = \left(\frac{\rho}{P}\right)^2 \rho^2 = \rho^2$$  \hspace{1cm} (19)
In order to derive bounds on average delays and queue sizes, we use (9), with \( f(||X_n||) = ||X_n|| \) and \( V(X_n) = X_n^T \). Under these conditions, we must compute the terms in (9). For \( \epsilon \), which appears at the denominator, we compute \( \epsilon_{\text{max}} \), the maximum value of \( \epsilon \) for which equation (8) holds.

In order to compute \( \epsilon_{\text{max}} \), from (16) we can write:

\[
- \frac{E[X_{n+1}X_n^T - X_nX_{n+1}^T | X_n]}{||X_n||} \geq \epsilon
\]

\[\forall X_n : ||X_n|| > B \]  \hfill (20)

for a \( B > 0 \). The function at the left hand size of the inequality (20) admits a limit for \( ||X_n|| \rightarrow \infty \), which depends on the direction of \( X_n \). \( \epsilon_{\text{max}} \) is equal to the smallest value for this limit, since any smaller value of \( \epsilon_{\text{max}} \) verifies (20) for large values of \( ||X_n|| \).

Thus, recalling that \( X_{n+1} = X_n + A_n - D_n \), \( \epsilon_{\text{max}} \) can be obtained as:

\[
\epsilon_{\text{max}} = \lim_{||X_n|| \rightarrow \infty} \inf \frac{E[X_{n+1}X_n^T - X_nX_{n+1}^T | X_n]}{||X_n||} = \lim_{||X_n|| \rightarrow \infty} \inf \frac{2E[(D_n - A_n)X_n^T | X_n]}{||X_n||} = \lim_{||X_n|| \rightarrow \infty} \inf \frac{2(D_nX_n^T - E[A_n]X_n^T)}{||X_n||}
\]

Consider the vector \( Y_n = E[A_n] + (1 - \rho - \delta)D_n \). It is straightforward to prove that \( ||Y_n||_1 < 1 \) for each \( \delta > 0 \). Thus, from (15):

\[
\frac{D_nX_n^T - Y_nX_n^T}{||X_n||_1} = \frac{D_nX_n^T - E[A_n]X_n^T - (1 - \rho - \delta)D_nX_n^T}{||X_n||_1} \geq 0
\]

and

\[
\frac{D_nX_n^T - E[A_n]X_n^T}{||X_n||_1} \geq \frac{(1 - \rho - \delta)D_nX_n^T}{||X_n||_1}
\]

As a consequence, for each \( \delta > 0 \):

\[
\epsilon_{\text{max}} = \lim_{||X_n|| \rightarrow \infty} \sup \frac{2(D_nX_n^T - E[A_n]X_n^T)}{||X_n||_1} \geq \frac{2(1 - \rho - \delta)D_nX_n^T}{||X_n||_1} = 2(1 - \rho - \delta)D_nX_n^T
\]

Finally, since Proposition 1 holds, we obtain:

\[
\epsilon_{\text{max}} > \frac{2(1 - \rho - \delta)}{P}
\]

or

\[
\epsilon_{\text{max}} \geq \frac{2(1 - \rho)}{P}
\]

Let us now evaluate the term \( E[V(X_{n+1}) - V(X_n) | Y_n \in H_B] \) appearing in (9):

\[
E[X_{n+1}^TX_{n+1} - X_n^TX_n] = E[(X_n + A_n - D_n)(X_n + A_n - D_n)^T - X_nX_n^T] = E[2(A_n - D_n)X_n^T + E(A_n - D_n)(A_n - D_n)^T]
\]

Applying (21) and Proposition 1, we obtain:

\[
E[X_{n+1}^TX_{n+1} - X_n^TX_n] \leq \leq -2(1 - \rho)D_nX_n^T + E[(A_n - D_n)(A_n - D_n)^T] \leq -2(1 - \rho)\frac{\epsilon_{\text{max}}}{\epsilon} + E[(A_n - D_n)(A_n - D_n)^T]
\]

As a consequence, (9) with \( f(||X_n||) = ||X_n||_1 \) becomes, for large values of \( n \):

\[
E[||X_n||_1] \leq E \left[ ||X_n||_1 \left( 1 - \frac{2(1 - \rho)}{P} \epsilon_{\text{max}} \right) + \right.
\]

\[
+ \frac{(A_n - D_n)(A_n - D_n)^T}{\epsilon} \left[ Y_n \in H_B \right] P\{Y_n \in H_B\} = E \left[ ||X_n||_1 \left( 1 - \frac{\epsilon_{\text{max}}}{\epsilon} \right) + \right.
\]

\[
+ \frac{(A_n - D_n)(A_n - D_n)^T}{\epsilon} \left[ Y_n \in H_B \right] P\{Y_n \in H_B\} \leq E \left[ \frac{(A_n - D_n)(A_n - D_n)^T}{\epsilon} \right] \left[ Y_n \in H_B \right] P\{Y_n \in H_B\} \leq E \left[ \frac{(A_n - D_n)(A_n - D_n)^T}{\epsilon} \right] = E[A_n^T] + E[D_n^T] - 2E[A_nD_n^T]
\]

being \( 0 < \epsilon \leq \epsilon_{\text{max}} = 2(1 - \rho) \).

Thus, considering (17), (18), (19), and choosing \( \epsilon = \epsilon_{\text{max}} \) to minimize the bound value, we have:

\[
E[||X_n||_1] \leq \frac{\rho - \rho^2/P}{(1 - \rho)} P^2
\]

It is possible to repeat the same derivation considering \( ||X_n||_2 \), and recalling that \( ||X_n||_1/||X_n||_2 \leq P \); the result is:

\[
E[||X_n||_2] \leq \frac{\rho - \rho^2/P}{(1 - \rho)} P^{1/2}
\]

In the case of uniform traffic we can use the result in (22) to obtain a bound on the average length of individual input queues. Indeed, since \( ||X_n||_1 = \sum_{i=0}^{N-1} x^{(i)} \), \( E[||X_n||_1] = \sum_{i=0}^{N-1} E[x^{(i)}] = N E[x^{(i)}] \), so that:

\[
E[x^{(i)}] \leq \frac{\rho - \rho^2/P}{(1 - \rho)}
\]

The bound on the average cell delay is then obtained by applying Little’s result:

\[
E[T] \leq \frac{P - \rho}{(1 - \rho)}
\]

being \( E[a^{(i)}] = \rho/P \).

IV. BOUNDS ON QUEUE SIZE VARIANCES

In this section we derive bounds on the second moment of the queue lengths in a \( P \times P \) IQS implementing the MWM scheduling policy. The bounds are derived by repeating the same procedure used in the previous section, but using \( || \cdot ||_2 \) as the norm
operator, rather than $\| \cdot \|_1$, and by adopting the (polynomial) Lyapunov function:

$$V(X_n) = (X_n X_n^T)^{\frac{3}{2}}$$

We use (9) to obtain a bound on $E[f(||X_n||_2)]$, where $f(||X_n||_2) = (||X_n||_2)^2$ (also a polynomial function). From the bounds on the second moment of the queue lengths and the bounds on the average queue lengths derived in the previous section, it is then possible to compute bounds on the queue length variances.

We can now prove that the assumptions of Theorem 4 are verified.

**Proposition 2:** Under any admissible i.i.d. input traffic pattern, for an IQS using an MWM scheduler with weights equal to queue lengths there exist $\epsilon \in \mathbb{R}^+$ and $B \in \mathbb{R}^+$ such that

$$\frac{E[V(X_{n+1})] - V(X_n)}{E[||X_n||_2^2]} \leq \epsilon \|X_n\|_2 > B \tag{26}$$

**Proof:** From the definition of $V(X_n)$, we can write:

$$E[V(X_{n+1})] - V(X_n) = E\left[ ((X_n + A_n - D_n)(X_n + A_n - D_n)^T)^{\frac{3}{2}} - (X_n X_n^T)^{\frac{3}{2}} | X_n \right] =$$

$$= E \left[ \left( 1 + \frac{2(A_n - D_n) X_n^T}{||X_n||_2^2} + \frac{(||A_n - D_n||_2^2)^2}{2} \right)^{\frac{3}{2}} - \left( \frac{3}{2} \right) ||X_n||_2^3 \right]$$

Taking the limit for $||X_n||_2 \to \infty$, we can use a McLaurin expansion:

$$(1 + t)^{\frac{3}{2}} \leq 1 + \frac{3}{2} t$$

thus:

$$\lim_{||X_n||_2 \to \infty} \frac{E[V(X_{n+1})] - V(X_n)}{E[||X_n||_2^2]} \leq$$

$$\leq E \left[ \frac{3(A_n - D_n) X_n^T}{||X_n||_2^2} + \frac{3(||A_n - D_n||_2^2)^2}{2} ||X_n||_2 \right]$$

$$\tag{28} \leq -3 \frac{1 - \rho}{\sqrt{1 - \rho}} = -\epsilon_{\text{max}}$$

Using (15), similarly to the proof of Proposition 1, letting $Y_n = E[A_n] + (1 - \rho - \delta) D_n$, it is possible to derive that

$$E \left[ \frac{3(A_n - D_n) X_n^T}{||X_n||_2} \right] \leq -3 \frac{1 - \rho}{\sqrt{1 - \rho}} = -\epsilon_{\text{max}} \tag{29}$$

We can now use (9) to obtain the cell delay variance bound. Consider again an IQS under admissible uniform traffic, so that (17) hold. Then:

$$\lim_{n \to \infty} E[||X_n||_2^2] =$$

$$= \lim_{n \to \infty} E \left[ \frac{3}{2\epsilon_{\text{max}}} ||X_n||_2 (||A_n - D_n||_2)^2 \right] P\{X_n \in H_B\} \leq$$

$$\leq E \left[ \frac{3}{2\epsilon_{\text{max}}} ||X_n||_2 (||A_n - D_n||_2)^2 \right] \tag{30}$$

Using (17), (18) and (19), we get:

$$E(||A_n - D_n||_2^2) = 2(\rho P - \rho^2)$$

Finally, the uniform traffic assumption allows the derivation of a bound on the second moment of individual input queue lengths. Indeed, since $||X_n||_2^2 = \sum_{i=0}^{N-1} (x(i))^2$ we have $E(||X_n||_2^2) = \sum_{i=0}^{N-1} E[(x(i))^2] = NE[(x(0))^2]$, so that:

$$E[(x(0))^2] \leq \frac{(\rho - \rho^2)^2}{(1 - \rho)^2} \tag{32}$$

**V. SIMULATION RESULTS**

In this section we compare simulation results with our analytical bounds. We implemented a synchronous simulator of the dynamics of IQ switches, and run a set of simulations with different arrival processes. All simulation experiments were stopped when a confidence level equal to 0.95 was reached with a relative confidence interval width equal to 0.05 on the measured performance metrics.

The arrival process feeding each switch input port is Bernoulli, with average offered load $\rho$. The destinations of cells is uniformly chosen among all output ports.

Recalling that the bounds that were previously derived apply to a wide class of arrival processes, and in particular that $a_n$ and $a_n^{(i)}$ must be i.i.d., but a correlation between $a_n^{(i)}$ and $a_n^{(j)}$ is allowed, we ran different simulations of IQ switches with different arrival patterns.

Define the parameter $\nu$ as follows:

$$\nu = P\{a_n^{(j)} = 1 \mid a_n^{(i)} = 1\} \tag{33}$$

$\nu$ represents the degree of correlation between the arrival processes at two different VOQs directed to the same output. Note that, if the arrival processes at different VOQs are independent, then $\nu = \rho/P$. In simulations experiments, we set $\nu = \{1, 0.5, 0.25, 0.125, \rho/P\}$; note that when $\nu = 1$, a complete correlation between arrival processes at different VOQs.
directed to the same output exists, i.e., if a cell arrives at input \( i \), directed to output \( j \), then at every other input queue storing cells directed to the same output \( j \), a cell also arrives.

Figure 1 plots the average cell delay estimates obtained by simulation for the average delay \( E[T] \) at one VOQ, versus the input port load \( \rho \), in a 16 x 16 IQS, as well as the bound in (25). Numerical results show that the bound captures the behavior of \( E[T] \), and that the ratio between the bound value and the simulation estimates depends on the value of \( \nu \). The bound becomes tighter when the degree of correlation grows; for example, for \( \nu = 1 \) the ratio becomes close to 2.

In Figure 2, instead, we report the average delay simulation results and the bound versus the number of IQS ports \( P \) for a fixed offered load \( \rho = 0.9 \). In this case, under an uncorrelated arrival process, the average delay is almost independent from the switch size, but when correlation grows, \( E[T] \) exhibits a linear dependence from \( \rho \), as also captured by the bound.

Figures 3 and 4 plot simulation results and bounds for the average VOQ length \( E[\tilde{x}] \), versus the offered load \( \rho \) and versus the number of switch ports \( P \), respectively. Considerations similar to those presented for average delays also apply to average queue sizes, due to the deterministic relation between \( E[T] \) and \( E[\tilde{x}] \) given by Little’s theorem.

Figure 5 plots the simulation estimates for the second moment of individual input queue lengths \( E[x^{(1)}_i]^2 \), versus the offered load \( \rho \), and compares them with the bound in (32). Again, the curve of the bound follows that of the simulation estimates, but now it is not as close as for averages.

The curves in Figure 6 report simulation estimates and bounds on \( E[|x^{(1)}_i|^2] \), versus the switch size; in this case the bound exhibits a linear dependence on the switch size that is not observed in simulations results.

VI. CONCLUSIONS

In this paper we considered input-buffered, cell-based switch architectures, which are often used today in high performance IP routers. We assumed that the scheduling algorithm used to select at each time slot a subset of the cells buffered at input ports for transfer to output ports is based on maximum weight matching, with weights equal to the sizes of virtual input queues.

Using analytical techniques based upon the theory of Lyapunov functions, we derived original bounds to the average and variance of the size of input queues, and to average cell delays. These bounds complement the previously available theoretical results for input-buffered, cell-based switch architectures, that only referred to throughput values.

Simulation results under correlated Bernoulli arrivals showed
that analytical bounds are rather tight, specially if we consider that our bounds are valid for a large class of arrival processes. Within such class, it is probably possible to devise other (possibly more esoteric) forms of correlation, leading to performance even closer to the bounds.

Although we focussed on the case of uniform traffic, the proposed approach can be extended to more complex traffic patterns, but the tightness of the resulting bounds may be worse.

REFERENCES

[1] www.cisco.com