Packet scheduling in input - queued switches with a speedup of less than two

Claus Bauer, Dolby Laboratories, San Francisco, CA, cb@dolby.com

Abstract— **Input queued switches and the solution of input/output contention by scheduling algorithms have been widely investigated. Most research has focused on fixed-size packets. This contrasts with the variable size packets of IP networks. In this paper, we investigate how the class of maximal matching algorithms deployed in switches with a speedup of less than two can be modified to take into account the varying packet sizes. Using a novel model for the dynamics of maximal matching algorithms, we show that modified maximal matching algorithms guarantee stability of the switch and establish bounds on the average delay experienced by a packet.**

I. Introduction and Motivation

The main performance features of switches are throughput and packet delay. These features are determined by the switch architecture which includes the queue structure, the scheduling algorithm and the speedup. Because output-queued switches are becoming less practical due to the high speedup required in the switching core, input-queued (IQ) and combined input/output queued (CIOQ) switches have gained major importance. A typical CIOQ $N \times N$ switch is shown in figure 1. For each input i, there are N virtual output queues $VOQ_{i,j}$, $1 \leq j \leq N$. The packets arriving at input i and destined for output j are buffered in $VOQ_{i,j}$. The switching core itself is modeled as a crossbar requiring that not more than one packet can be sent simultaneously from the same input or to the same output.

Fig. 1. Architecture of an input queued switch

Two main classes of scheduling algorithms have been investigated in the literature. Maximum weight scheduling (MWM) algorithms have been proven to guarantee throughput and to provide bounds on average delay ([3],[6]). They do not require a speedup of the switching core, but are impractical due to their computational complexity of $O(N^3 \log N)$. A switch is said to work at a speedup of $S, S \geq 1$, if it works at a speed S times faster than the speed of the input links. The less complex - and with regard to a chosen metric less optimal - maximal matching algorithms have been widely researched in ([4],[7]). A maximal matching algorithm is defined for a set of weights $Q_{i,j}$, $1 \leq i,j \leq N$, where $Q_{i,j}$ is the weight assigned to $VOQ_{i,j}$, as follows:

1. Initially, all $VOQ_{i,j}$ are considered potential choices for a cell transfer.
2. The VOQ with the largest weight, say $VOQ_{a,b}$, is chosen for a cell transfer and ties are broken randomly.
3. All $VOQ_{i,j}$ with either $i = a$ or $j = b$ are removed.
4. If all $VOQ_{i,j}$ are removed, the algorithm terminates. Else go to step 2.

It has been shown that maximal matching algorithms provide stability for a speedup of $S = 2$ ([2]) or even less ([1]).

Most existing research has focused on cells of fixed size. This contrasts with the fact that the size of packets in IP networks varies. In order to forward packets with varying size, a cell-based scheduling algorithm requires an input module that segments packets into fixed size cells. Cells are then forwarded independently. This leads to bandwidth inefficiencies as described in [8]. Moreover, cells belonging to the same packet do not necessarily arrive contiguously at the output. Thus, at the output, cells belonging to various packets must be stored and re-assembled to a packet when all cells belonging to a packet have been received. In order to overcome these disadvantages of cell based switching, cell-based switch architectures that take the actual packet size into account have been investigated in [5] and [8]. In both papers, cell-based scheduling MWM algorithms are modified for packet scheduling. In [8], two procedures to derive a packet based scheduling algorithm from a cell based scheduling algorithm $ALGO$ are described. With both approaches, when an input - output connection is selected for the transfer of the first cell of a packet by the $ALGO$ algorithm, then the connection is maintained at least until all cells of the packet are forwarded. With the first approach, the matching is maintained until no port is transmitting an unfinished packet. Then, a new matching is calculated. This algorithm is called $PB_1 - ALGO$. With the second approach, at each timeslot, all input and output ports are divided in busy ports, which are busy sending unfinished packets, and free ports, which either have no packets to send or just finished sending a packet. At each timeslot the considered algorithm is only applied to the free ports. This algorithm is denoted as $PB_2 - ALGO$. It is shown in [5] that $PB_1 - MWM$ and $PB_2 - MWM$ provide stability.

It is desirable to understand if the less complex maximal matching algorithms MM can be modified into packet scheduling algorithms $PB_1 - MM$ and $PB_2 - MM$ such that they provide stability. This paper investigates this question for switches with a speedup of $S > R$ defined as follows. Let $\lambda_{i,j}$ define the arrival rate expressed in cells for $VOQ_{i,j}$ and set $\lambda_{\min} = \min_{\substack{i,j \\ \lambda_{i,j} > 0}} \lambda_{i,j}$. In the sequel, we will assume that traffic is admissible, i.e., $\sum_{j=1}^{N} \lambda_{i,j} \leq 1$, $\sum_{i=1}^{N} \lambda_{i,j} \leq 1$, $\forall\, i,j,\ 1 \leq i,j \leq N$. We set

$$E_{i,j} = \{(a,b),\, 1 \leq a,b \leq N : |a = i \bigvee b = j.\}$$

$$R_{i,j} = \sum_{(a,b)\in E_{i,j}} \lambda_{i,j}, \quad R = \max_{1\leq i,j\leq N} R_{i,j}. \tag{1}$$

We see $R < 2$. It is shown that $PB_1 - MM$ and $PB_2 - MM$ algorithms with a speedup of $S > R$ are stable and an upper bound for the average delay experienced by a packet is established. In contrast to earlier work that proved the stability of MM- algorithms with a fluid technique [2], this paper is based on a discrete model which allows to give bounds on the expected delay. A new model to describe the dynamics of maximal matching algorithms is used to establish results on stability and delay for the $PB_1 - MM - LQF$ algorithm, the $PB_1 - MM$ algorithm whose weights are proportional to the lengths of the VOQs. A more general characterization of the forwarding behavior of all inputs and outputs competing for specific resources allows us then to establish stability and delay bounds for all $PB_1 - MM$ and $PB_2 - MM$ algorithms.

The rest of the paper is organized as follows. Section II introduces the terminology to describe the dynamics of a switch and gives some preliminary results. In section III, an inequality that characterizes the behavior of maximal matching algorithms is derived. In the sections IV and V, results on stability and upper bounds on the average delay for the $PB_1 - MM - LQF$ algorithm are proven. Section VI describes an alternate technique to derive delay bounds for all $PB_1 - MM$ and $PB_2 - MM$ algorithms. We conclude in section VII.

II. Terminology and Model

Throughout this paper, the time t is described via a discrete, slotted time model. Cells are supposed to be of fixed size and packets are supposed to be of integer multiple cell size. An external timeslot is the time needed by a packet that is of cell size one to arrive completely at an incoming link. A larger packet will thus need several external timeslots to arrive completely. An external timeslot from time t to $t+1$ consists of the S internal timeslots $[t + (k-1)/S, t + k/S]$, $1 \leq k \leq S$, where an internal timeslot is defined as the time needed to forward one cell through the switching core. For the sake of simplicity, we assume that S is an integer. This assumption is only used in the proof of theorem 2. This proof can be easily generalized to any rational S. Packets arrive at the beginning of an external timeslot t and cells are transferred instantly at the end of an internal timeslot. We abbreviate the summation $\sum\limits_{1\leq i,j\leq N}$ as $\sum\limits_{i,j}$. $\underline{x}$ defines an $N \times N$ matrix $(x_{1,1}, ..., x_{1,N}, ..., x_{N,1}, ..., x_{N,N})^T$ and its norm $||x||_1$ is defined as $||x||_1 = \sum\limits_{i,j}^{N} x_{i,j}$. We define the arrival matrix $A(t)$, representing the arrivals of (parts of) packets of cell size one at each VOQ:

$$A_{i,j}(t) = \begin{cases} 1, & \text{if a (part of a) packet of cell size} \\ & \text{one arrives at } VOQ_{i,j} \text{ at time } t \\ 0, & \text{else.} \end{cases}$$

As at most one (part of a) packet of cell size one can arrive at an input during an external timeslot, there holds

$$\sum_{1\leq j\leq N} A_{i,j}(t) \leq 1, \tag{2}$$

$\forall i$, $1 \leq i \leq N$, $\forall t$. The service matrix $S^k(t)$, indicating which queues are chosen for cell transfer is defined as follows:

$$S^k_{i,j}(t) = \begin{cases} 1, & \text{if } VOQ_{i,j} \text{ is chosen for cell transfer at the} \\ & \text{end of the } k\text{-th internal time slot of} \\ & \text{the } t\text{-th external time slot,} \\ 0, & \text{else.} \end{cases}$$

Due to the crossbar structure of the switch, the following inequalities hold:

$$\sum_{1\leq i\leq N} S^k_{i,j}(t) \leq 1, \quad \sum_{1\leq j\leq N} S^k_{i,j}(t) \leq 1, \tag{3}$$

$\forall i, j$, $1 \leq i, j \leq N$, $\forall k$, $1 \leq k \leq S$, $\forall t$. We set $L(t) = (L_{1,1}(n), .., L_{N,N}(t))$ where $L_{i,j}(t)$ defines the occupancy of $VOQ_{i,j}$ (measured in integer cell sizes) at the beginning of t-th external timeslot. Accordingly, we define $L^k_{i,j}(t)$ $1 \leq k \leq S$ as the occupancy of $VOQ_{i,j}$ at the beginning of the k-th internal timeslot in the t-th external timeslot. Therefore, $L_{i,j}(t) = L^1_{i,j}(t)$, $\forall\, i, j$, $1 \leq i, j \leq N$. The development of the VOQ occupancy between consecutive internal timeslots is described by the following equations:

$$\begin{aligned} L^k_{i,j}(t) &= [L^{k-1}_{i,j}(t) - S^{k-1}_{i,j}(t)]^+, \ \forall k, 2 \leq k \leq S, \quad (4) \\ L^1_{i,j}(t) &= [L^S_{i,j}(t-1) - S^S_{i,j}(t-1)]^+ + A_{i,j}(t). \quad (5) \end{aligned}$$

Combining (4) and (5), the development of $L_{i,j}(t)$ over external timeslots is modelled as follows:

$$L_{i,j}(t+1) = [L_{i,j}(t) - \sum_{k=1}^{S} S^k_{i,j}(t)]^+ + A_{i,j}(t+1). \tag{6}$$

Now, we formalize the notion of the stability of a switch and give a sufficient criterion for stability which we will use later to establish some of our results.

Definition: Let $Y_n = (y_n(1), ..., y_n(M))$ be the row vector of a system of M queues at time n, where $y_n(i)$ is the length of the queue i at time n. A system of queues is said to be strongly stable if, for every $\epsilon > 0$, there exists $B > 0$ such that $\lim\limits_{n\to\infty} P\{||Y_n||_1 > B\} < \epsilon$, and $\lim\limits_{n\to\infty} \sup E||Y_n||_1 < \infty$.

Theorem 1 *Given a system of queues whose evolution is described by a discrete time Markov chain with state vector $X_n \in N^M$, if a lower bounded function $V(X_n)$, called Lyapunov function, $V : N^M \to R$ exists such that*

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

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

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

then the system of queues is strongly stable.

Proof: This is a special instance of theorem 1 in [3].□

For the subsequent analysis, the external timeslots t_n when all ports are free, will play a major role. In this context, we state the following lemma for further use:

Lemma 1: Consider an input-queued packet switch, that deploys an $PB_1 - MM$ or $PB_2 - MM$ algorithm whose input traffic is formed by variable length packets with i.i.d. random size. We assume that for each $VOQ_{i,j}$ the average packet size and the packet size variance are finite. We further assume that the transmission of packets from all VOQs selected by the $PB - MM$ algorithm start at the same rate. We define the sequence of instants t_n at which either the transmission of all packets at the head of selected queues ends at the same time, or all selected queues become empty. Then the sequence t_n is a non-defective renewal process, i.e. for any t_n, the evolution of the system following t_n is independent of the evolution of the system before t_n, and the sequence $z_n = t_{n+1} - t_n$ satisfies $E[z_n] < \infty$ and $E[z_n^2] < \infty$.

Proof: The proof follows the proof of lemmata 2 and 3 in [5]. □

In order to simplify the calculations, we assume in the sequel that the instants t_n always occur at the beginning of external timeslots. The general case of the instants t_n occuring at the beginning of any internal timeslot can be proved in the same way. We define the approximate next-state vector to describe the development of the VOQs between stopping times t_n :

$$\hat{L}_{i,j}(t_{n+1}) = L_{i,j}(t_n) + \sum_{h=0}^{z_n-1} A_{i,j}(t_n + h + 1)$$
$$-S_{i,j}(t_n + h), \;\; S_{i,j}(t) = \sum_{k=1}^{S} S_{i,j}^k(t). \qquad (7)$$

III. A characterization of maximal matching algorithms

In this section, we develop an inequalitiy that describes the dynamics of a maximal matching algorithm for fixed cell sizes. We will use this inequality to prove stability for the $PB_1 - MM - LQF$ algorithm in the next section. We define the weights of the algorithm as a function of the actual time as follows: $Q(t) = (Q_{i,1}(t), ..., Q_{N,N}(t))$ are the weights at the beginning of the t-th external timeslot. $Q_{i,j}^k(t)$, $1 \le k \le S$, is the weight of the $VOQ_{i,j}$ at the beginning of the k-th internal timeslot of the t-th external timeslot. Thus, $Q_{i,j}(t) = Q_{i,j}^1(t)$. Now, we prove the following theorem:

Theorem 2: *For any input buffered switch that applies a maximal matching algorithm and a speedup-up S, there holds at any time t*

$$\frac{1}{R} \sum_{k=1}^{S} \sum_{i,j} Q_{i,j}^k(t) \lambda_{i,j} \le \sum_{i,j} \sum_{k=1}^{S} Q_{i,j}^k(t) S_{i,j}^k(t).$$

Proof: In the first internal timeslot, in its first iteration, the algorithm selects the queue with the largest weight, say Q_{a_1,b_1}, for transfer. Thus from (1):

$$Q_{a_1,b_1}(t)R \ge Q_{a_1,b_1}(t)R_{a_1,b_1} \ge \sum_{(m,n)\in E_{a_1,b_1}} Q_{m,n}(t)\lambda_{m,n}.$$

All $VOQ_{i,j}$ with either $i = a_1$ or $j = b_1$ are removed. In the second iteration, the remaining queue with the largest weight, say Q_{a_2,b_2}, is chosen. Thus,

$$Q_{a_2,b_2}(t)R \ge Q_{a_2,b_2}(t)R_{a_2,b_2} \ge \sum_{\substack{(m,n)\in E_{a_2,b_2} \\ m\neq a_1,\, n\neq b_1}} Q_{m,n}(t)\lambda_{m,n}. \quad (8)$$

The matching algorithms stops after k, $k \le N$ iterations when only empty queues remain. For each of these k iterations, an inequality analogous to (8) holds. For the remaining empty queues, additional $(N - k)$ inequalities as in (8) hold where both sides are equal to zero. Summing over all N inequalities, we obtain

$$\sum_{m=1}^{N} Q_{a_m,b_m}(t) = \sum_{i,j} Q_{i,j}(t) S_{i,j}(t) \ge R^{-1} \sum_{i,j} Q_{i,j}(t)\lambda_{i,j}.$$

Applying this analysis to all S internal timeslots, we get

$$\sum_{k=1}^{S} \sum_{i,j} Q_{i,j}^k(t) S_{i,j}^k(t) \ge \frac{1}{R} \sum_{k=1}^{S} \sum_{i,j} Q_{i,j}^k(t)\lambda_{i,j}. \qquad \square$$

IV. Stability of $PB_1 - MM - LQF$

In this section, we prove stability for the $PB_1 - MM - LQF$ algorithm. We show the following theorem:

Theorem 3: *The $PB_1 - MM - LQF$ algorithm with a speedup $S > R$ is stable for all admissible i.i.d. arrival processes.*

Proof: The $PB_1 - MM - LQF$ algorithms behaves as a $MM - LQF$ algorithm for fixed cell sizes at the instants t_n and keeps its matching unchanged between these instants. This proof considers the development of the lengths of the virtual output queues between two consecutive instants t_n and t_{n+1}. In order to describe the scheduling behavior at the instant t_n, we will use theorem 2. We define the quadratic Lyapunov function as

$$V(L(t)) = \sum_{i,j} L_{i,j}^2(t). \qquad (9)$$

We first assume $L_{i,j}(t_n) \ge S z_n \; \forall i, j, \; 1 \le i, j \le N$. This implies $L_{i,j}(t_{n+1}) = \hat{L}_{i,j}(t_{n+1})$ and

$$S_{i,j}(t_n + h) \;\; = \;\; S_{i,j}(t_n), \forall h, \, 0 \le h \le z_n - 1. \quad (10)$$

We derive an upper bound for the expression $E[V(\hat{L}(t_{n+1})) - V(L(t_n))|L(t_n)]$. We see from (7) and (9):

$$\begin{aligned} & V(\hat{L}(t_{n+1})) - V(L(t_n)) \\ = \;\; & 2\sum_{i,j} \sum_{h=0}^{z_n-1} (A_{i,j}(t_n + h + 1) - S_{i,j}(t_n + h))L_{i,j}(t_n) \\ & + \sum_{i,j} \left(\sum_{h=0}^{z_n-1} A_{i,j}(t_n + h + 1) - S_{i,j}(t_n + h) \right)^2 . (11) \end{aligned}$$

Obviously, from (2), (3) and (7)

$$\left(\sum_{h=0}^{z_n-1} A_{i,j}(t_n+h+1) - S_{i,j}(t_n+h)\right)^2 \leq S^2 z_n^2. \tag{12}$$

Using (10), (12) and following an idea in [5], we apply Wald's equation [9], §2 - 13, to the sequence of stopping times t_n on the right hand side of (11):

$$\begin{aligned} & E\left[V(\hat{L}(t_{n+1})) - V(L(t_n))|L(t_n)\right] \\ \leq\ & N^2S^2E[z_n^2] + 2E\left[\sum_{h=0}^{z_n-1}\sum_{i,j}(A_{i,j}(t_{n+h+1}) \right. \\ & \left. -S_{i,j}(t_{n+h}))L_{i,j}(t_n)|L(t_n)\right] \\ =\ & N^2S^2E[z_n^2] + 2E\left[\sum_{i,j}\left[\sum_{h=0}^{z_n-1} A_{i,j}(t_{n+h+1}) \right.\right. \\ & \left.\left. \times L_{i,j}(t_n) - z(n)S_{i,j}(t_n)L_{i,j}(t_n)|L(t_n)\right]\right] \\ =\ & N^2S^2E[z_n^2] + 2E[z_n]E\left[\sum_{i,j}(A_{i,j}(t_n) \right. \\ & \left. -S_{i,j}(t_n))L_{i,j}(t_n)|L(t_n)\right]. \end{aligned} \tag{13}$$

We see from (4) that $L_{i,j}^{k+1}(t) \leq L_{i,j}^k(t)$, $L_{i,j}^S(t) + S - 1 \geq L_{i,j}(t)$, $\forall k$, $1 \leq k \leq S-1$. Using these inequalities together with (7) and theorem 2 with $Q_{i,j}^k(t) = L_{i,j}^k(t)$, we obtain:

$$\begin{aligned} & E\left[\sum_{i,j}(A_{i,j}(t_n) - S_{i,j}(t_n))L_{i,j}(t_n)|L(t_n)\right] \\ =\ & \sum_{i,j}\left[\lambda_{i,j}L_{i,j}(t_n) - \sum_{k=1}^{S} S_{i,j}^k(t_n)L_{i,j}(t_n)\right] \\ \leq\ & \sum_{i,j}\left[\lambda_{i,j}(S-1+L_{i,j}^S(t)) - R^{-1}\sum_{k=1}^{S}\lambda_{i,j}L_{i,j}^k(t_n)\right] \\ \leq\ & \left(1-\frac{S}{R}\right)\sum_{i,j}\lambda_{i,j}L_{i,j}^S(t_n) + (S-1)\sum_{i,j}\lambda_{i,j} \\ \leq\ & \left(1-\frac{S}{R}\right)\min_{\substack{i,j \\ \lambda_{i,j}>0}}\lambda_{i,j}||L^S(t_n)||_1 + (S-1)\sum_{i,j}\lambda_{i,j}. \end{aligned}$$

Inserting the last result in (13), we see

$$\begin{aligned} & E\left[V(\hat{L}(t_{n+1})) - V(L(t_n))|L(t_n)\right] \\ \leq\ & N^2S^2E[z_n^2] + 2E[z_n]\left(1-\frac{S}{R}\right)\min_{\substack{i,j \\ \lambda_{i,j}>0}}\lambda_{i,j} \\ & \times||L^S(t_n)||_1 + 2E[z_n](S-1)\sum_{i,j}\lambda_{i,j}. \end{aligned} \tag{14}$$

Therefore, using lemma 1 and noting that $||L^S(t_n)||_1 + N(S-1) \geq ||L(t_n)||_1$ by (3) and (4), we obtain

$$E\left[V(\hat{L}(t_{n+1})) - V(L(t_n))|L(t_n)\right] \leq -\epsilon||L(t_n)||_1 \tag{15}$$

$\forall L(t_n) : ||L(t_n)|| > B$. We now remove the assumption $L_{i,j}(t_n) \geq Sz_n, \forall\, i,j$. We see from (6) and (7) that if $L_{i,j}(t_n) < Sz_n$, then $L_{i,j}(t_{n+1}) \leq \hat{L}_{i,j}(t_{n+1}) + Sz_n$ or $L_{i,j}^2(t_{n+1}) \leq \hat{L}_{i,j}^2(t_{n+1}) + 2\hat{L}_{i,j}(t_{n+1})Sz_n + S^2z_n^2 \leq \hat{L}_{i,j}^2(t_{n+1}) + 3S^2z_n^2$, where we have used that $\hat{L}_{i,j}(t_n+1) \leq L_{i,j}(t_{n+1}) < Sz_n$. Thus, using lemma 1 and (15):

$$\begin{aligned} & E[V(L(t_{n+1}+1)) - V(L(t_n))|L(t_n)] \\ \leq\ & 3N^2S^2E[z_n^2] - \epsilon||L(t_n)||_1 \leq -\epsilon_1||L(t_n)||_1, \end{aligned} \tag{16}$$

$\forall L(t_n) : ||L(t_n)|| > B$. The theorem follows from theorem 1, (16), and the fact that stability at the stopping points t_n also implies stability between them as $E[z_n] < \infty$. □

V. Bounds on average delay for $PB_1 - MM - LQF$

In this section, we derive bounds on the average delay experienced by cells at the VOQs. We first prove the following bound on the average sum of the length of all VOQs.

Theorem 4: *Under the assumption of i.i.d. admissible traffic, for the $PB_1 - MM - LQF$ algorithm with a speedup $S > R$, the average sum of all the queue lengths $E[||L(t)||_1]$ is bounded as follows:*

$$\begin{aligned} E[L(t)] \ \leq\ & \frac{E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2) + 4E[z_n^2]\sum_{i,j}\lambda_{i,j}^2}{2E[z_n]\left(\frac{S}{R}-1\right)\min_{\substack{i,j \\ \lambda_{i,j}>0}}\lambda_{i,j}} \\ & +N(E[z_n]+1). \end{aligned}$$

Proof: We revisit the proof of theorem 3. Here, we only argue with the approximate state vector $\hat{L}(t_{n+1})$ and not with the exact state vector $L(t_{n+1})$. An analysis of the proof below shows that the exact state vector would incur an additional term of $N^2(S+1)E[z_n]$ to the delay bound. The approximate state vector will give a bound C. We will see that in general $C \geq N^2(S+1)E[z_n]$, and hence we will not consider the term $N^2(S+1)E[z_n]$ for the rest of this section. We note from (5)

$$||L^S(t_n)||_1 + N \geq ||L(t_n+1)||_1. \tag{17}$$

Using (17) and arguing as in (11), (13), (14), without using the estimate (12), we obtain:

$$\begin{aligned} & E\left[V(\hat{L}(t_{n+1})) - V(L(t_n))|L(t_n)\right] \\ \leq\ & E\left[\sum_{i,j}\left(\sum_{h=1}^{z_n-1} A_{i,j}(t_n+h+1) - S_{i,j}(t_n+h)\right)^2\right] \\ & +2E[z_n]\min_{\substack{i,j \\ \lambda_{i,j}>0}}\lambda_{i,j}\left(1-\frac{S}{R}\right)[E[||L(t_n+1)||_1]+N] \\ & +2E[z_n](S-1)\sum_{i,j}\lambda_{i,j}. \end{aligned} \tag{18}$$

Following an argument from [7], we see that by definition

$$E[A_{i,j}(t)] = E[A_{i,j}^2(t)] = \lambda_{i,j}, \tag{19}$$

for large enough t. As the $MM-LQF$ algorithm is stable, the switch state becomes a discrete time Markov chain with a stationary distribution. Thus, for large enough t,

$$E[S_{i,j}(t)] = E\left[\sum_{k=1}^{S} S_{i,j}^k(t)\right] = E\left[\sum_{k=1}^{S} (S_{i,j}^k(t))^2\right] = \lambda_{i,j}. \tag{20}$$

Then, by the Cauchy - Schwarz inequality:

$$E\left[\sum_{k=1}^{S} S_{i,j}^k(t)\right]^2 \leq E\left[\sum_{k=1}^{S} (S_{i,j}^k(t)^2\right] S = \lambda_{i,j} S. \tag{21}$$

Thus, from (19), (20), and (21):

$$\begin{aligned} & E\left[\sum_{i,j}\left(\sum_{h=1}^{z_n-1} A_{i,j}(t_n+h+1) - S_{i,j}(t_n+h)\right)^2\right] \\ = \; & E\left[\sum_{i,j}\sum_{h=1}^{z_n} A_{i,j}^2(t_n+h+1) + S_{i,j}^2(t_n+h)\right] \\ & +2\sum_{i,j}\sum_{\substack{h,u=1,\\ h<u}}^{z_n}\Bigg[A_{i,j}(t_n+h+1)A_{i,j}(t_n+u+1) \\ & +S_{i,j}(t_n+h)S_{i,j}(t_n+u)\Bigg] \\ & +2\sum_{i,j}\sum_{h=1}^{z_n} A_{i,j}(t_n+h+1)\sum_{h=1}^{z_n} S_{i,j}(t_n+h)\Bigg] \\ = \; & \sum_{i,j}\left(E[z_n](S+1)\lambda_{i,j} + (4E[z_n^2]-2E[z_n])\lambda_{i,j}^2\right) \end{aligned} \tag{22}$$

In summary, from (18) and (22) we obtain:

$$\begin{aligned} & E[V(\tilde{L}(t_{n+1}+1)) - V(L_n(t))|L(t_n)] \\ \leq \; & E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2) + 4E[z_n^2]\sum_{i,j}\lambda_{i,j}^2 \\ & +2E[z_n]\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}\left(1-\frac{S}{R}\right)[E[||L(t_n+1)||_1 + N]. \end{aligned}$$

We see from (6) and (7), $V(\hat{L}(t)) \geq V(L(t))$. From this and the last inequality , we see:

$$\begin{aligned} & E[V(\hat{L}(t_{n+1}))] \\ \leq \; & E[V(\hat{L}(t_{n+1})) - V(L(t_n)) + V(\hat{L}(t_n))] \\ \leq \; & E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2) + 4E[z_n^2]\sum_{i,j}\lambda_{i,j}^2 \\ & +2E[z_n]\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}\left(1-\frac{S}{R}\right)[E[||L(t_n+1)||_1] + N] \\ & +E[V(\hat{L}(t_n))]. \end{aligned}$$

Summing over $n = 0$ to $n = T-1$, we get

$$\begin{aligned} & E[V(\hat{L}(t_T))] \leq E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2) \\ + \; & 4TE[z_n^2]\sum_{i,j}\lambda_{i,j}^2 + 2E[z_n]\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}\left(1-\frac{S}{R}\right) \\ & \times\sum_{n=0}^{T-1}[E||L(t_n+1)||_1 + N] + E[V(\hat{L}(t_0))]. \end{aligned} \tag{23}$$

As $t_0 = 0$, assuming $E[V(\hat{L}(0))] = 0$ and noting that $V(\cdot) \geq 0$, we see from (23):

$$\begin{aligned} \frac{1}{T}\sum_{n=0}^{T-1} E[||L(t_n+1)||_1] \leq \; & \frac{E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2)}{2E[z_n]\left(\frac{S}{R}-1\right)\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}} \\ & + \frac{4E[z_n^2]\sum_{i,j}\lambda_{i,j}^2}{2E[z_n]\left(\frac{S}{R}-1\right)\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}} + N. \end{aligned} \tag{24}$$

As we assume an i.i.d. arrival process, the switch state is a discrete, irreducible aperiodic, i.e. ergodic Markov chain. Thus, the left hand side of (24) converges to the expected value of $||L(t_T+1)||_1$ in the equilibrium state. Thus

$$\lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1} E[||L(t_{n+1}+1)||_1] = \lim_{T\to\infty} E[||L(t_T+1)||_1]. \tag{25}$$

We obtain from (24) and (25):

$$\begin{aligned} E[||L(t_T+1)||_1] \leq \; & \frac{E[z_n]\sum_{i,j}((3S-1)\lambda_{i,j} - 2\lambda_{i,j}^2)}{2E[z_n]\left(\frac{S}{R}-1\right)\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}} \\ & + \frac{4E[z_n^2]\sum_{i,j}\lambda_{i,j}^2}{2E[z_n]\left(\frac{S}{R}-1\right)\min_{\substack{i,j\\ \lambda_{i,j}>0}}\lambda_{i,j}} + N. \end{aligned} \tag{26}$$

By the definition of the stopping points t_n, (2), and (4), for any $t \in [t_T+1, t_{T+1}+1[$: there is $E[||L(t)||] \leq E[||L(t_T+1)||] + NE[z_n]$. This and (26) imply the theorem.

For the specific case of uniform arrival traffic for all VOQs, i.e. $\lambda_{i,j} = \lambda\,\forall\, i,j,\, 1 \leq i,j \leq N$, there holds $E[L(t)]/N^2 = E[L_{i,j}(t)]$. Using Little's law, we derive from theorem 4 that for the expected delay $E[T]$ holds:

$$E[T] \leq \frac{E[z_n]((3S-1)-2\lambda) + 4E[z_n^2]\lambda}{2E[z_n]\left(\frac{S}{R}-1\right)\lambda} + \frac{E[z_n]+1}{N\lambda}.$$

VI. Stability and delay bounds for all $PB_1 - MM$ and $PB_2 - MM$ algorithms

In this section, we prove stability and establish delay bounds for all $PB_1 - MM$ and $PB_2 - MM$ algorithms with a speedup $S > R$. We introduce additional notation:

$$C_{i,j}(t) = \sum_{(a,b)\in E_{i,j}} L_{a,b}(t), \quad B_{i,j}(t) = \sum_{(a,b)\in E_{i,j}} A_{a,b}(t),$$
$$D_{i,j}(t) = \sum_{(a,b)\in E_{i,j}} S_{a,b}(t). \qquad (27)$$

The corresponding exact and approximate next state vector equations can be derived from (4), (5), (6), and (7).

Theorem 5: *Every $PB_1 - MM$ and every $PB_2 - MM$ algorithm with a speedup $S > R$ is stable for all admissible i.i.d. arrival processes.*

Proof: We consider a specific timeslot $[t, t+1[$. We first assume that $L_{i,j}(t) \geq S \,\forall i,j,\ 1 \leq i,j \leq N$. In each internal timeslot, a cell will be sent from any of k, $k \leq N$ busy inputs. Among the $N-k$ free inputs and outputs, an $PB_1 - MM - LQF$ or $PB_2 - MM - LQF$ algorithm will select $N-k$ new connections for cell transfer. Thus, a cell is sent from each input in each internal timeslot, i.e.,

$$C_{i,j}(t) \geq S. \qquad (28)$$

Because $||C(t)||_1 = (2N-1)||L(t)||_1$, we prove stability for $C(t)$. We consider the Lyapunov function defined in (9) with the argument $C(t)$. Arguing as in (11), we find

$$V(\hat{C}(t+1) - V(C(t)) = 2\sum_{i,j}\Big(B_{i,j}(t+1)$$
$$-D_{i,j}(t)\Big)C_{i,j}(t) + \sum_{i,j}\left(B_{i,j}(t+1) - D_{i,j}(t)\right)^2 \quad (29)$$

Obviously, from (2), (3) and (7)

$$\sum_{i,j}\left(B_{i,j}(t+1) - D_{i,j}(t)\right)^2 \quad \leq \quad (2N-1)^2N^2S^2. \quad (30)$$

We obtain from (1), (28), (29), and (30):

$$E\left[V(\hat{C}(t)) - V(C(t))|L(t)\right] \leq (2N-1)^2N^2S^2$$
$$+2\sum_{i,j}\left(R_{i,j}(t+1) - D_{i,j}(t)\right)C_{i,j}(t)$$
$$\leq (2N-1)^2N^2S^2 + 2(R-S)||C(t)||_1 < -\epsilon E||C(t)||_1,$$

$\forall C(t)$, $||C(t)|| > B$. Thus the theorem follows from theorem 1 if $L_{i,j}(t_n) \geq Sz_n$. Otherwise, we compare $V(\hat{C}(t_{n+1}))$ and $V(C(t_{n+1}))$ analogously to (16). □

Theorem 6: *Under the assumption of i.i.d. admissible traffic, for every $PB_1 - MM$ and every $PB_2 - MM$ algorithm with a speedup $S > R$, the expression $E[||C(t)||_1]$ is bounded as follows:*

$$E\left[||C(t)||_1\right] \leq \frac{(2N-1)(S+1)\sum_{i,j} R_{i,j} + 2\sum_{i,j} R_{i,j}^2}{2(S-R)}.$$

Proof: The proof follows the proof of theorem 4. Instead of using the bound (22), the bound (31) is applied. Using the Cauchy Schwarz inequality and (21), we see

$$E[B_{i,j}^2(t)] \leq (2N-1)\sum_{(k,l)\in E_{i,j}} E[A_{k,l}(t)]^2 = (2N-1)R_{i,j},$$
$$E[D_{i,j}^2(t)] \leq (2N-1)\sum_{(k,l)\in E_{i,j}} E[S_{k,l}(t)]^2 \leq (2N-1)SR_{i,j}.$$

Using these estimates and (27), one finds

$$E\left[\sum_{i,j}\left(B_{i,j}(t+1) - D_{i,j}(t)\right)^2\right]$$
$$\leq \quad (2N-1)(S+1)\sum_{i,j} R_{i,j} + 2\sum_{i,j} R_{i,j}^2. \qquad (31)$$

For the specific case of uniform arrival traffic for all VOQs, i.e. $\lambda_{i,j} = \lambda \,\forall i,j,\ 1 \leq i,j \leq N$, there is $E[C(t)]/(2N-1)N^2 = E[L_{i,j}(t)]$. Using Little's law, we see from theorem 5 that for the expected delay $E[T]$ holds:

$$E[T] \leq \frac{(2N-1)(S+1) + 2(2N-1)\lambda}{2(S-R)}.$$

VII. Conclusions

The inefficiencies of cell based scheduling algorithms in data networks with varying packet sizes motivate the study of packet scheduling algorithms. This paper considers the application of maximal matching algorithms with a speedup of less than than two to combined input/output queued switches that use packet scheduling. Using a new way to model the dynamics of maximal matching algorithms, the stability of the switches are proven and bounds on average delay are established. The proofs rely on the theory of the Lyapunov function.

References

[1] Benson, K., *Throughput of crossbar switches using maximal matching algorithms,* Proc. of IEEE ICC 2002, New York City.

[2] Dai, J.D.; Prabhakar,B., *The throughput of data switches with and without speedup,* Proc. of IEEE Infocom 2000, Tel Aviv.

[3] Leonardi, E., Mellia, M., Neri, F., Marsan, M.A., *Bounds on average delay and queue size averages and variances in input queued cell-based switches,* Proc. of IEEE Infocom 2001, Anchorage, Alaska.

[4] Leonardi, E., Mellia, M., Neri, F., Marsan, M.A., *Stability of maximal size matching scheduling in input queued cell switches,* Prof. of IEEE ICC 2000, New Orleans.

[5] Leonardi, E., Neri, F., Marsan, M.A., Bianco, A., Giaccone, P., *Packet scheduling in input-queued cell-based switches,* Transactions on Networking, 10(5): 666-678 (2002).

[6] N. McKeown, A.Mekkittikul, V. Anantharam, J. Walrand, *Achieving 100% throughput in an input-queued switch,* IEEE Trans. on Comm., vol. 47, no. 8, Aug. 1999, 1260 - 1272.

[7] Shah, D.; Kopikare, M., *Delay bounds for approximate maximum weight matching algorithms for input queued switches,* Proc. of IEEE Infocom 2002, New York City, June 2002.

[8] Shah, D,; *Input-queued switches: Cell switching versus packet switching,* IEEE Infocom 2003, San Francisco.

[9] Wolff, R.W., *Stochastic modeling and the theory of queues*, Prentice-Hall, NJ, 1989.