Bandwidth Allocation for Best Effort Traffic to Achieve 100% Throughput

Masoumeh Karimi, Zhuo Sun, and Deng Pan
Florida International University, Miami, FL
E-mails: {mkari001, zsun003, pand}@fiu.edu

***Abstract*—Generalized Processor Sharing (GPS) is a powerful model and there are many practical scheduling algorithms that can perfectly emulate it. GPS is widely used as an ideal fairness model to schedule packets for guaranteed performance traffic. However, there has not been a way for GPS to properly handle the best effort traffic. In this paper, we propose a bandwidth allocation scheme for GPS called Queue Length Proportional (QLP) in crossbar switches without speedup. QLP dynamically obtains a feasible bandwidth matrix to schedule best effort flows. In QLP, the amount of service that each flow receives is proportional to the length of its backlogged queue. We analytically prove that QLP is strongly stable and hence provides 100% throughput for any admissible traffic, no matter whether the traffic distribution is uniform or non-uniform. Moreover, we show that QLP is feasible, which means the allocated bandwidth does not exceed the available capacity. We also discuss how to track the queue length in GPS. Finally, we perform simulations to verify the theoretical results and to measure the performance of QLP.**

I. Introduction

Generalized Processor Sharing (GPS) has long been known as a simple and powerful fluid model for traffic scheduling [1]. All fair queueing algorithms ultimately emulate the GPS ideal model because it achieves perfect fairness for packet scheduling [2], [3], [4]. GPS is a theoretical fluid model and divides the available bandwidth into logically independent channels. Thus, traffic of each flow is smoothly transmitted through its own exclusive channel from the input port to the output port. GPS is widely used as an ideal fairness model to schedule packets for guaranteed performance traffic. However, there has not been a way for GPS to properly handle the best effort traffic. The objective of this paper is to address the appropriate scheduling of best effort traffic in order to be employed in crossbar switches.

Crossbar switches have received significant attention due to the non-blocking capability and large bandwidth utilization in comparison with bus based switches [6], [7]. The challenge of bandwidth allocation in a crossbar switch is how to efficiently share the available capacity of each input port and output port. Simple proportional bandwidth allocation for a shared link is not proper for crossbar switches [8], [9] because flows of a switch are subject to two bandwidth constraints: the available bandwidth at both the input port and output port of the flow. The scheme should be efficient to fully utilize the available bandwidth, and should be feasible in order to be applied in practice.

Our motivation for this work arises from the fact that in crossbar switches we need to dynamically obtain an admissible bandwidth matrix to properly handle best effort traffic. Guaranteed performance flows reserve resources for an allocated transmission rate [10]. However, best effort flows try to make the best use of the available transmission capacity but have no guarantee to the quality of service [7]. As can be seen, the bandwidth allocation scheme plays several important roles in guaranteeing the high performance of a switch [11]. First, the scheme helps to determine the traffic admission policy and buffer management strategy. Second, an efficient scheme makes it possible for a switch to achieve 100% throughput. Third, the scheme is used as the scheduling criterion by fair scheduling algorithms. There are many fair scheduling algorithms designed for bandwidth allocation in different crossbar switch architectures [2], [3], [12], [13], [14], [15]. However, most of them focused on providing quality of service for guaranteed performance traffic and there is relatively less work on how to better support best effort traffic.

In this paper, we present a bandwidth allocation scheme for GPS called Queue Length Proportional (QLP) to properly handle best effort traffic in crossbar switches without speedup. In QLP, the amount of service that each flow receives, or its dedicated bandwidth, is proportional to the length of its backlogged queue. QLP essentially favors the queues with the greatest occupancy and thus assists the crossbar switch to be more work-conserving. We conduct theoretical analysis to prove that QLP is strongly stable and therefore achieves 100% throughput for any admissible traffic, no matter whether the traffic distribution is uniform or non-uniform. We also show that QLP is feasible, which means the allocated bandwidth does not exceed the available capacity. Furthermore, we discuss how to track the queue length is GPS. Lastly, we conduct simulations to verify the analytical results and to measure the performance of QLP.

The organization of this paper is as follows. In Section II, we present the abstract switch model and the QLP bandwidth allocation scheme. In Section III, we theoretically analyze the stability and throughput of QLP and then provide some discussions. We show the simulation results in Section IV. Finally, we conclude the paper in Section V.

II. Queue Length Proportional (QLP) scheme

In this section, we present our bandwidth allocation scheme. First, we briefly explain the switch model. Then, we describe the QLP bandwidth allocation algorithm for best effort traffic.

A. The Abstract Switch Model

The considered switch architecture includes N input ports and N output ports, connected by a crossbar with no internal speedup. Let In_i denote the i^{th} input port and Out_j denote the j^{th} output port. The available bandwidth of each input port and output port and also the crossbar is R. Define the traffic from In_i destined to Out_j to be a flow F_{ij}. Use $R^a_{ij}(t)$ to represent the allocated bandwidth of F_{ij} at time t. Denote the queue of packets at In_i destined to Out_j as Q_{ij}.

B. Algorithm Description

In this subsection we present our Queue Length Proportional (QLP) bandwidth allocation scheme and investigate its feasibility property.

As mentioned earlier, simple proportional bandwidth allocation policy of GPS does not apply to switches [8], [9], [11]. For a GPS server works at a fixed bandwidth R, the rate of $\frac{\phi_j}{\sum_j \phi_j} R$ will guarantee the performance of each flow, where ϕ_j is the weight of each flow [1]. However in contrast with a single server, flows of a switch are subject to two bandwidth constraints: the available bandwidth at both the input and output port of the flow. Naive bandwidth allocation at the output port may make the flows violate the bandwidth constraints at their input ports, and vice versa.

QLP dynamically assigns the bandwidth to each best effort flow proportional to the backlogged queue length. We use queue length $Q_{ij}(t)$ as a dynamic weight of each flow F_{ij} and define $Q_{*j}(t) = \sum_i Q_{ij}(t)$ to be the number of bits queued at all input ports directed to a particular Out_j at time t, and $Q_{i*}(t) = \sum_j Q_{ij}(t)$ to be the number of bits queued at a particular In_i destined to different output ports at time t. We also denote the allocated bandwidth of a flow F_{ij} respecting to the constraint of each input port and output port as $R^{a.in}_{i*}(t)$ and $R^{a.out}_{*j}(t)$, accordingly. Recalling the GPS fluid model, traffic of each flow can smoothly stream from the input port to the output port through its own exclusive channel, without buffering in the middle, as illustrated in Figure 1. Thus, by considering the bandwidth constraints at both the input port and output port of each flow F_{ij} we have

$$R^{a.in}_{ij}(t) = \frac{Q_{ij}(t)}{Q_{i*}(t)} R = \frac{Q_{ij}(t)}{\sum_j Q_{ij}(t)} R \tag{1}$$

$$R^{a.out}_{ij}(t) = \frac{Q_{ij}(t)}{Q_{*j}(t)} R = \frac{Q_{ij}(t)}{\sum_i Q_{ij}(t)} R \tag{2}$$

In order to avoid bandwidth violation at the input port and the output port, we consider the overall allocated bandwidth $R^a_{ij}(t)$ of F_{ij} to be the minimum [8] between two calculated bandwidths as follows

$$R^a_{ij}(t) = min\{R^{a.in}_{ij}(t), R^{a.out}_{ij}(t)\} \tag{3}$$

$$= min\{\frac{Q_{ij}(t)}{Q_{i*}(t)} R, \frac{Q_{ij}(t)}{Q_{*j}(t)} R\} \tag{4}$$

$$= \frac{Q_{ij}(t)}{max\{Q_{i*}(t), Q_{*j}(t)\}} R \tag{5}$$

As can be seen, QLP bandwidth allocation scheme essentially favors the queues with the greatest occupancy. It treats all queues fairly and assists the crossbar switch to be more work-conserving [20]. One of the advantages of our scheme is its simplicity. It does not require any sorting or re-sorting process. Using the calculated information of the queue lengths, QLP just needs to find the maximum between the aggregated queue length of flows waiting in a particular input port and flows directed to a specific output port.

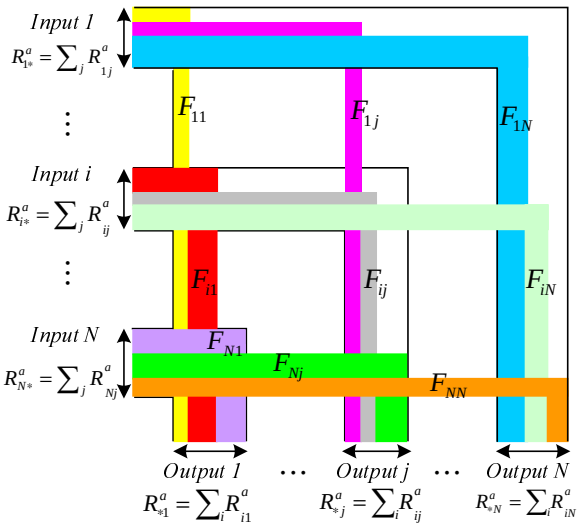

Fig. 1. GPS ideal fluid model used for scheduling in a crossbar switch.

Now we discuss the important properties of QLP, feasibility and stability. A bandwidth allocation scheme should efficiently utilize the available bandwidth while maintaining feasibility.

Definition 1: The allocated bandwidth $R^a_{ij}(t)$ of F_{ij} will be feasible if no over-subscription happens at any input port or output port. In other words, the aggregated assigned rate of flows does not exceed the available capacity, i.e.

$$R^a_{i*} \leq R \quad \text{and} \quad R^a_{*j} \leq R \tag{6}$$

For easy presentation, assume a normalized bandwidth; $R = 1$. Now, we show that our scheme is feasible. According to the QLP description we can write

$$R^a_{i*} = \sum_{j=1}^{N} R^a_{ij}(t) = \sum_{j=1}^{N} \frac{Q_{ij}(t)}{max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}}$$

$$R^a_{*j} = \sum_{i=1}^{N} R^a_{ij}(t) = \sum_{i=1}^{N} \frac{Q_{ij}(t)}{max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}}$$

Since $\sum_{j=1}^{N} Q_{ij}(t) \leq max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}$ and $\sum_{i=1}^{N} Q_{ij}(t) \leq max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}$, we have

$$R^a_{i*} \leq 1 \quad \text{and} \quad R^a_{*j} \leq 1 \tag{7}$$

which indicates QLP is feasible. We study the stability of our scheme in the next section.

III. Performance Analysis

A. Stability and Throughput

In this subection, we theoretically prove that QLP is strongly stable for any admissible traffic which implies that QLP achieves 100% throughput.

We first adopt the following definitions presented in [23]. The switch size is shown by $P = N \times N = N^2$.

Definition 2: U is the set of vector $Y = (Y_1, ..., Y_P)$ and $Y \in \mathbb{R}^{+P}$, such that

$$\sum_{i=1}^{N} Y_{(i+jN)} \leq 1 \quad j = 0, 1, ..., (N-1) \tag{8}$$

$$\sum_{j=0}^{N-1} Y_{(i+jN)} \leq 1 \quad i = 1, ..., N \tag{9}$$

Definition 3: $\|Y\|$ is the Euclidean norm of vector Y, i.e., $\|Y\| = \sqrt{y_1^2 + ... \ y_i^2 \ ... + y_P^2} = \sqrt{\sum_{i=1}^{P}(y_i^2)}$.

Definition 4: $\hat{Y}$ is the normalized vector parallel to Y, given that $Y \neq 0$, i.e., $\hat{Y} = \frac{Y}{\|Y\|} = \frac{\tilde{Y}}{\|\tilde{Y}\|}$.

Definition 5: $\tilde{Y}$ is the maximal vector parallel to Y, given that $Y \neq 0$ and $k \in \mathbb{R}$, i.e., $\tilde{Y} = max_k \ kY$.

Definition 6: Ψ_Y is the symmetric matrix associated with the projection operator along the direction of $\hat{Y}$, given that $Y \neq 0$, i.e., $\Psi_Y = \hat{Y}^T \hat{Y}$.

Property 1: Given that $Y \neq 0$, we have $Y\Psi_Y = Y$, which is a straightforward result of definition 6.

The following variables are used to represent the status of the crossbar switch. Their initial values are assumed to be zero at time $t = 0$.

$Q_{ij}(t)$ is the number of bits buffered in Q_{ij} at time t, belong to flow F_{ij}. $Q(t)$ is the vector of queue lengths at time t, i.e., $Q(t) = (Q_{11}, ..., Q_{ij}, ..., Q_{NN})$.

$A_{ij}(t)$ is the number of bits arriving at Q_{ij} up to time t. $A(t)$ is the vector of the number of arrivals at time t, i.e., $A(t) = (A_{11}, ..., A_{ij}, ..., A_{NN})$.

$D_{ij}(t)$ is the number of bits departing from Q_{ij} up to time t. $D(t)$ is the vector of the number of departures at time t, i.e., $D(t) = (D_{11}, ..., D_{ij}, ..., D_{NN})$.

Assume that the number of arriving bits $A_{ij}(t)$ satisfies the Strong Law of Large Numbers (SLLN), i.e.

$$\lim_{t \to \infty} \frac{A_{ij}(t)}{t} = \lambda_{ij} \tag{10}$$

where λ_{ij} is the arrival rate of Q_{ij}. Consider the incoming traffic is admissible, which means that no over-subscription at any input ports or output ports, i.e., for all $\lambda_{ij} \geq 0$ we have

$$\forall i, \sum_j \lambda_{ij} \leq 1, \quad \text{and} \quad \forall j, \sum_i \lambda_{ij} \leq 1 \tag{11}$$

Correspondingly, Λ is the vector of the average arrival rates, i.e., $\Lambda = (\lambda_{11}, ..., \lambda_{ij}, ..., \lambda_{NN})$, and due to the admissible traffic we have $E[A(t)] = \Lambda$.

Similar to that in [25][18], the evolution equation of the switch for the interval $[t, t+1]$ is described as follows

$$Q(t+1) = Q(t) + A(t) - D(t) \tag{12}$$

Before investigating the stability of our scheme, we first calculate the average departure rate as the following lemma.

Lemma 1: For a crossbar switch without speedup, the average departure rate in QLP bandwidth allocation scheme is proportional to the queue length, i.e.

$$E[D(t)] = \tilde{Q}(t) \tag{13}$$

Proof: According to the previous Section, the dedicated portion of the available bandwidth R for each queue Q_{ij} at time t, can be described as a weight factor $w_{ij}(t)$. It means that for a flow F_{ij} we can have

$$w_{ij}(t) = \begin{cases} 0, & \text{if } Q_{ij}(t) = 0 \\ \frac{Q_{ij}(t)}{max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}}, & \text{otherwise} \end{cases} \tag{14}$$

which is positive when there is an offered load $Q_{ij}(t) > 0$. Consequently, $R_{ij}^a(t)$, the allocated bandwidth of flow F_{ij} at time t will be $\frac{Q_{ij}(t)}{max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}} R$. Obviously, the departure rate of each flow will be equal to its allocated bandwidth as follows.

$$D_{ij}(t) = R_{ij}^a(t) = \frac{Q_{ij}(t)}{max\{\sum_j Q_{ij}(t), \sum_i Q_{ij}(t)\}} R = \tilde{Q}_{ij}(t) R$$

which is proportional to its queue length at time t. By considering the normalized available bandwidth $R = 1$, we have $D_{ij}(t) = \tilde{Q}_{ij}(t)$, and thus $E[D(t)] = \tilde{Q}(t)$. ■

In order to prove 100% throughput of our scheme, we introduce the following definition and lemma from [23][25].

Definition 7: A system of queues is strongly stable if

$$\lim_{t \to \infty} sup \ E[\|Q(t)\|] < \infty \tag{15}$$

which implies 100% throughput and bounded delay guarantee.

Lemma 2: Given a system of queues whose evolution is described by a DTMC with state vector $S(t) \in \mathbb{N}^M$, and whose state space H is a subset of the Cartesian product of a denumerable state space H_Q and a finite state space H_K, then, if a lower bounded function $V(Q(t))$, called Lyapunov function, $V : \mathbb{N}^P \to \mathbb{R}$ can be found such that $E[V(Q_{t+1})|S(t)] < \infty \ \ \forall S(t)$ and there exists $\epsilon \in \mathbb{R}^+$, $B \in \mathbb{R}^+$ such that $\forall S(t) : \ \|Q(t)\| > B$

$$E[V(Q(t+1)) - V(Q(t))|S(t)] < -\epsilon \|Q(t)\| \tag{16}$$

then the system of queues is strongly stable. It means that the queue length does not grow infinitely which implies 100% throughput and bounded average delay.

Proof: We need to show $\lim_{t \to \infty} sup \ E[\|Q(t)\|] < \infty$. For the detailed proof, see theorem 2 in [23]. ■

Now, we present the main theorem for the stability of QLP.

Theorem 1: For a crossbar switch without speedup, the QLP bandwidth allocation scheme is strongly stable for any admissible traffic, i.e., it achieves 100% throughput.

Proof: According to lemma 2 and equation (16), we can define a quadratic Lyapunov function $V(t) = Q(t)ZQ^T(t)$ as that in [24][25], if there exists a symmetric copositive matrix $Z \in \mathbb{R}^{P \times P}$. Similarly, we can consider

$$Z = I - \rho \Psi_\Lambda \tag{17}$$

where I is an identity matrix, $\rho \in \mathbb{R}$ such that $0 \leq \rho \leq 1$, $\Lambda = E[A(t)]$ is the vector of the average arrival rate, and by definition 6 we have $\Psi_\Lambda = \hat{\Lambda}^T \hat{\Lambda}$. It is easy to prove that Z is positive (semi)definite. We also assume that the state vector $S(t) = Q(t)$.

Now, we need to prove that for some $\epsilon \in \mathbb{R}^+$, $B \in \mathbb{R}^+$ (B is large enough), $\exists \rho \in \mathbb{R}$ such that for $\|Q(t)\| > B$, we have

$$E[Q(t+1)ZQ^T(t+1) - Q(t)ZQ^T(t)|Q(t)] < -\epsilon\|Q(t)\|$$

For notational convenience, we define the timing index as a subscript, e.g., Q_{t+1} is equivalent to $Q(t+1)$. Therefore

$$E[Q_{t+1}ZQ_{t+1}^T - Q_tZQ_t^T|Q_t] < -\epsilon\|Q_t\| \tag{18}$$

By substituting Q_{t+1} and Q_{t+1}^T from the evolution equation (12) into the left hand side of (18) we have

$$E[Q_{t+1}ZQ_{t+1}^T - Q_tZQ_t^T|Q_t] = \tag{19}$$
$$E[(Q_t + A_t - D_t)Z(Q_t^T + A_t^T - D_t^T) - Q_tZQ_t^T|Q_t] =$$
$$E[2(A_t - D_t)Q_t^T + (A_t - D_t)(A_t - D_t)^T|Q_t] \tag{20}$$

For simplicity, we find the limit of (20) when $\|Q_t\|$ tends to infinity, as follows.

$$\lim_{\|Q_t\|\to\infty} \frac{E[2(A_t - D_t)ZQ_t^T + (A_t - D_t)Z(A_t - D_t)^T|Q_t]}{\|Q_t\|}$$

As can be seen, the terms $(A_t - D_t)$ and $(A_t - D_t)^T$ are bounded since the number of arrivals and departures in time interval $[t, t+1]$ are bounded. Also, we know that Z is a positive (semi)definite matrix. It means that, the limit of $\frac{E[(A_t-D_t)(A_t-D_t)^T|Q_t]}{\|Q_t\|}$ when $\|Q_t\| \to \infty$ is 0. As a result, the remaining part of the equation will be

$$\lim_{\|Q_t\|\to\infty} \frac{E[2(A_t - D_t)ZQ_t^T\ |Q_t]}{\|Q_t\|} \tag{21}$$

By definition 4 and knowing that $\|Q_t^T\| = \|Q_t\|$, we obtain

$$\lim_{\|Q_t\|\to\infty} \frac{E[2(A_t - D_t)Z(\hat{Q}_t^T\|Q_t\|)\ |Q_t]}{\|Q_t\|} = \tag{22}$$
$$E[2(A_t - D_t)Z\hat{Q}_t^T\ |Q_t] \tag{23}$$

By using (17) in (23), we have

$$2E[(A_t - D_t)(I - \rho\Psi_\Lambda)\hat{Q}_t^T\ |Q_t] \tag{24}$$

By definition 6 and property 1, (24) can be written as follows

$$2(\Lambda\hat{Q}_t^T - E[D_t]\hat{Q}_t^T - \rho\Lambda\hat{Q}_t^T + \rho E[D_t]\Psi_\Lambda\hat{Q}_t^T)$$

By replacing the result of lemma 1 for $E[D_t]$, we obtain

$$2(\Lambda\hat{Q}_t^T(1-\rho) - \tilde{Q}_t\hat{Q}_t^T + \rho\tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T) \tag{25}$$

As can be seen, (25) is a function of ρ and $\hat{Q}_t$, i.e.,

$$f(\rho, \hat{Q}_t) = 2(\Lambda\hat{Q}_t^T(1-\rho) - \tilde{Q}_t\hat{Q}_t^T + \rho\tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T) \tag{26}$$

In order to proof the stability, we need to show that for the entire domain of $\hat{Q}_t$, there exists a ρ such that (26) is always less than a finite negative constant, i.e.,

$$\exists\rho\ ,\ \forall\hat{Q}_t\ :\ f(\rho, \hat{Q}_t) < -\epsilon$$

Since $\hat{Q}(t) = \frac{Q(t)}{\|Q(t)\|} = \frac{Q(t)}{\sqrt{\sum_{ij} Q_{ij}^2(t)}}$ is a normalized vector, for a given ρ, the domain of $f(\rho, \hat{Q}_t)$ is the surface of the unit sphere such that $\hat{Q}_t \in \mathbb{R}^{+P}$. On the other hand, for a given vector $\hat{Q}_t$, variation of $f(\rho, \hat{Q}_t)$ is linear versus the scalar ρ. Knowing that $0 \le \rho \le 1$, we analyze two cases as below, for all values of $\hat{Q}_t$.

Case 1. when $\rho = 1$: $f(1, \hat{Q}_t) = 2(-\tilde{Q}_t\hat{Q}_t^T + \tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T)$. If $\hat{Q}_t$ is in parallel with Λ, we will have $\hat{Q}_t = \hat{\Lambda}$, by definition 6 we can write $\Psi_\Lambda = \hat{Q}_t^T\hat{Q}_t$. Thus $f(1, \hat{Q}_t) = 2(-\tilde{Q}_t\hat{Q}_t^T + \tilde{Q}_t\hat{Q}_t^T\hat{Q}_t\hat{Q}_t^T) = 2\tilde{Q}_t(-\hat{Q}_t^T + \hat{Q}_t^T) = 0$.
If $\hat{Q}_t$ is not in parallel with Λ, we can find $\tilde{Q}_t\hat{Q}_t^T > \tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T$. It leads to have a negative value for $f(1, \hat{Q}_t)$. As a result, for all values of $\hat{Q}_t$ we obtain $f(\rho, \hat{Q}_t)|_{\rho=1} = f(1, \hat{Q}_t) \le 0$.

Case 2. when $0 \le \rho < 1$, or in other words: $-1 \le \rho - 1 < 0$. To investigate this case, we can write f as follows $f(\rho, \hat{Q}_t) = f(\rho, \hat{Q}_t)|_{\rho=1} + (\rho-1)\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t)$, where the partial derivative of f can be found as $\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t) = 2(-\Lambda\hat{Q}_t^T + \tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T)$. If $\hat{Q}_t$ is in parallel with Λ, i.e., $\hat{Q}_t = \hat{\Lambda}$ and thus $\Psi_\Lambda = \hat{Q}_t^T\hat{Q}_t$. We can write the partial derivative of f as $\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t) = 2(-\hat{Q}_t\hat{Q}_t^T + \tilde{Q}_t\hat{Q}_t^T\hat{Q}_t\hat{Q}_t^T)$, which is strictly positive, i.e., $\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t) > 0$. Since in case one we obtained $f(1, \hat{Q}_t) \le 0$, after comparison with $f(\rho, \hat{Q}_t) = f(1, \hat{Q}_t) + (\rho-1)\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t)$ for $0 \le \rho < 1$, we find that $f(\rho, \hat{Q}_t) < 0$.
If $\hat{Q}_t$ is not in parallel with Λ, we will have $\Lambda\hat{Q}_t^T < \tilde{Q}_t\Psi_\Lambda\hat{Q}_t^T$ and therefore, $\frac{\partial}{\partial\rho}f(\rho, \hat{Q}_t)$ will be strictly positive. Similarly, it can be shown that for $0 \le \rho < 1$ we have $f(\rho, \hat{Q}_t) < 0$.

Hence, QLP is always stable for a crossbar switch without speedup for any admissible traffic, no matter whether the traffic distribution is uniform or non-uniform. It means that the queue length at input buffers does not grow infinity and there exists a finite upper bound $B < \infty$ at which backlogged queues will settle. ■

B. Discussions

In this subsection, we discuss how to find out the queue length of each flow in GPS. For tracking the queue length of a flow F_{ij}, similar to equation (12) we can have an evolution equation for queue Q_{ij} during interval $[t_1, t_2]$ as follows

$$Q_{ij}(t_2) = Q_{ij}(t_1) + A_{ij}(t_2, t_1) - D_{ij}(t_2, t_1) \tag{27}$$

where $Q_{ij}(t_1)$ is the remaining backlogged queue of F_{ij} at time t_1. We know that $D_{ij}(t_2, t_1)$, the number of departed bits from a particular flow F_{ij} during interval $[t_1, t_2]$ in GPS, can be obtained as

$$D_{ij}(t_2, t_1) = \int_{t_1}^{t_2} R_{ij}^a(t)dt \tag{28}$$

Assume that the allocated bandwidth of F_{ij} is fixed to a constant R_{ij}^a during interval $[t_1, t_2]$, i.e., $R_{ij}^a(t) = R_{ij}^a$, equation (28) can be calculated as

$$D_{ij}(t_2, t_1) = \int_{t_1}^{t_2} R_{ij}^a dt = R_{ij}^a \times (t_2 - t_1) \tag{29}$$

Thus

$$Q_{ij}(t_2) = Q_{ij}(t_1) + A_{ij}(t_2, t_1) - (R_{ij}^a \times (t_2 - t_1)) \tag{30}$$

As can be seen, the length of each queue Q_{ij} can be found during any interval $[t_1, t_2]$.

IV. Simulation Results

In this section, we carry out simulations to verify the theoretical results in Section III, and to evaluate the performance of QLP. We consider a 16×16 crossbar switch. Each input and output has a bandwidth of $R = 1$ Gbps, and the crossbar has a speedup of one. We set the packet length to be distributed between 40 and 1,500 bytes [26]. For the destination of the packets, we consider both uniform traffic and non-uniform traffic pattern. For uniform traffic, the destination of a new incoming packet is uniformly distributed among all the output ports, i.e., $\lambda_{ij} = \eta R/N$, where η is the effective load and N is the switch size. The η takes one of the 10 possible values of [0.1,1] with a step of 0.1 For non-uniform traffic, we use the same model as that in [27]. The traffic arrival rate λ_{ij} is defined by i, j and an unbalanced probability w as follows.

$$\lambda_{ij}(t) = \begin{cases} R\left(w + \frac{1-w}{N}\right), & \text{if } i = j \\ R\frac{1-w}{N}, & \text{if } i \neq j \end{cases}$$

In this case, the η is fixed to 1 and w takes one of the 11 possible values of [0,1] with a step of 0.1. When $w = 0$, the traffic arrival is uniformly distributed among the outputs, i.e., $\lambda_{ij}(t) = R/N$. Otherwise, the incoming packets at In_i are more directed to Out_j rather than the other outputs, which is called the hotspot destination. A special case will happen when $w = 1$, i.e., $\lambda_{ii}(t) = R$. To constrain the burstiness of a flow F_{ij}, we consider a leaky bucket model $(\eta \times \lambda_{ij}, \sigma_{ij})$, where σ_{ij} is the burst size of F_{ij} [7]. We set σ_{ij} of every flow to a fixed value of 10,000 bytes, and the burst may arrive at any time during a simulation run.

To evaluate the performance of our scheme, We compare our simulation data with Localized Independent Packet Scheduling (LIPS), in which an input port or output port makes scheduling decisions solely based on the state information of its local crosspoint buffers [6]. We consider four different LIPS implementation versions with different arbitration rules as follows: FP (Fixed Priority) assigns a fixed priority order to all the virtual queues of the same input to the same output and always picks the candidate with the highest priority; RD (Random) makes the arbitration on a random basis; RR (Round Robin) alternatively chooses eligible candidates in a round robin manner to avoid starvation; OPF (Oldest Packet First) uses the packet arrival time as the arbitration criterion, i.e., the packet arriving earlier has a higher priority. Now, we investigate the results on the throughput and the average delay.

A. Throughput

To verify theorem 1, we present the simulation data to show that our scheme achieves 100% throughput.

Figure 2(a) displays the throughput under uniform traffic. It can be seen that, the throughput of different schemes grows consistently with the effective load, and finally reaches 100% when the effective load becomes 1, only FP slightly decreases when the load is the maximum. Figure 2(b) shows the throughput under non-uniform traffic. It clearly reflects the non-monotonic performance variations of other methods. As expected, QLP significantly yields higher throughput than the other schemes when the speedup is one. The other schemes

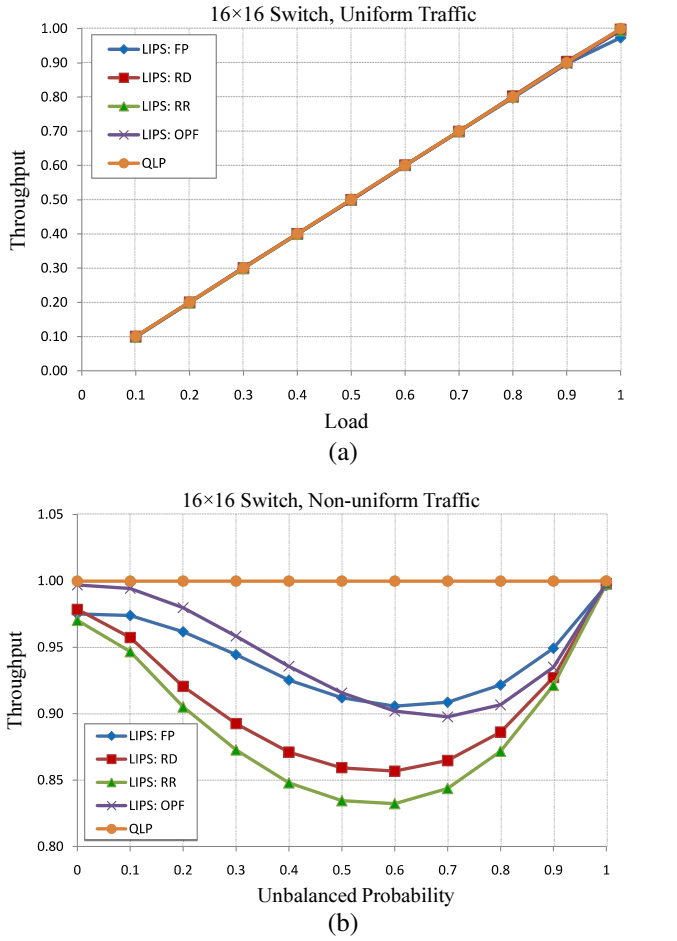

Fig. 2. Throughput of QLP. (a) With different loads (b) With different unbalanced probabilities.

achieve the lowest throughput when the unbalanced probability is around 0.5 and then, throughput is gradually improved until the unbalanced probability becomes 1. In fact at this point, all the packets of In_i go to Out_j and no scheduling is necessary. The results confirm that, QLP achieves 100% throughput and outperforms the other four methods when the speedup is one.

B. Average Delay

Next, we study the delay performance of QLP. We measure the total time that a packet stays in the switch. It is the interval from the time that the last bit of a packet arrives at its input to the time that the last bit of the packet is sent to the output. We plot the average delay of different schemes in logarithmic scale and the average delay is measured in seconds.

Figure 3(a) displays the average delay under uniform traffic. As can be seen, the delay grows gradually when the effective load increases, and jumps when the effective load becomes 1. Surprisingly, for the effective loads greater than 0.8, the QLP outperforms the other four schemes. Figure 3(b) depicts the average delay under non-uniform traffic. As expected, QLP shows an optimistic behavior in comparison with the other four methods and the delay difference is more than one decade. The average delay of other schemes increases with the unbalanced probability and reaches to the maximum when the unbalanced probability is around 0.5. Then, the average delay

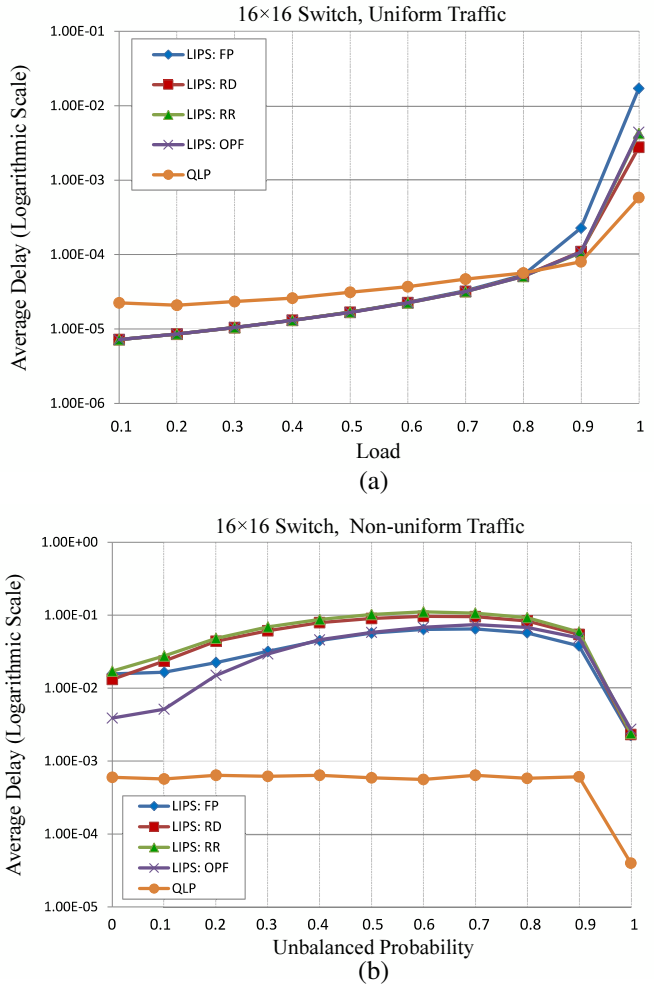

Fig. 3. Average Delay of QLP. (a) With different loads (b) With different unbalanced probabilities.

of all schemes drops when the unbalanced probability is equal to 1, because at this point all traffic of an input is destined to the same output and no switching is necessary. It is observed that, different unbalanced probabilities do not significantly affect the average delay of QLP, which demonstrates that our scheme works well under non-uniform traffic.

V. Conclusions

In this paper, we have presented the Queue Length Proportional (QLP) bandwidth allocation scheme for GPS to schedule best effort traffic in crossbar switches without speedup. In QLP, the amount of service that each flow receives is proportional to the length of its backlogged queue. QLP essentially favors the queues with the greatest occupancy and thus assists the crossbar switch to be more work-conserving. By theoretical analysis, we have proved that QLP is strongly stable and therefore provides 100% throughput for any admissible traffic, no matter whether the traffic distribution is uniform or non-uniform. We have also shown that QLP is feasible, which means the allocated bandwidth does not exceed the available capacity. Furthermore, we have discussed how to track the queue length in GPS. Finally, we have conducted simulations to verify the analytical results.

References

[1] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 344-357, Jun. 1993.

[2] A. Demers, S. Keshav and S. Shenker, "Analysis and simulation of a fair queueing algorithm," *ACM SIGCOMM '89*, vol. 19, no. 4, pp. 3-12, Austin, TX, Sept. 1989.

[3] H. Zhang, "WF2Q: worst-case fair weighted fair queueing," *IEEE INFOCOM '96*, pp. 120-128, San Francisco, CA, Mar. 1996.

[4] J. Xu and R. J. Lipton, "On Fundamental Tradeoffs between Delay Bounds and Computational Complexity in Packet Scheduling Algorithms," *IEEE/ACM Trans. on Netw.*, vol. 13, no. 1, pp. 15-28, Feb. 2005.

[5] S. He, S. Sun, W. Zhao, Y. Zheng, and W. Gao, "On Guaranteed Smooth Switching for Buffered Crossbar Switches," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 718-731, 2008.

[6] D. Pan and Y. Yang, "Localized Independent packet scheduling for buffered crossbar switches," *IEEE Trans. on Comp.*, vol. 58, Feb. 2009.

[7] J. Kurose and K. Ross, "Computer networking: a top-down approach," *Addison Wesley*, 4th edition, 2007.

[8] X. Zhang, S.R. Mohanty, and L.N. Bhuyan, "Adaptive Max-Min Fair Scheduling in Buffered Crossbar Switches Without Speedup," *26th IEEE International Conference on Computer Communications, INFOCOM'07*, pp. 454-462, Qualcomm Inc., San Diego, May 2007.

[9] M. R. Hosaagrahara and H. Sethu, "Max-Min Fair Scheduling in Input-Queued Switches," *IEEE Transactions on Parallel and Distributed Systems*, volume 19, number 4, pp. 462-475, Apr. 2008.

[10] Gerald R. Ash, "Traffic engineering and QoS optimization of integrated voice and data networks," *Morgan Kaufmann*, first edition, 2006.

[11] D. Pan and Y. Yang, "Max-min fair bandwidth allocation algorithms for packet switches," *IEEE Int. Par. and Dist. Processing Symp. (IPDPS)*, Long Beach, CA, Mar. 2007.

[12] S. Chuang, S. Iyer, and N. McKeown, 'Practical algorithms for performance guarantees in buffered crossbars," *IEEE INFOCOM '05*, Miami, FL, March 2005.

[13] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round robin," *IEEE/ACM Trans. Netw.*, vol. 4, no. 3, pp. 375-385, 1996.

[14] N. Ni and L. Bhuyan, "Fair scheduling for input buffered switches," *Cluster Computing*, vol. 6, no. 2, pp. 105-114, Hingham, MA, Apr. 2003.

[15] X. Zhang and L. Bhuyan, "Deficit round-robin scheduling for input-queued switches," *IEEE Journal on Selected Areas in Communications*, no. 4, pp. 584-594, May 2003.

[16] D. Pan and Y. Yang, "Credit based fair scheduling for packet switched networks," *IEEE INFOCOM*, pp. 843-854, Miami, FL, March 2005.

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

[18] N. McKeown, A. Mekkittikul, V. Anantharam and J. Walrand, "Achieving 100% throughput in an input queued switch," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1260-1267, 1999.

[19] D. Stephens and H. Zhang, "Implementing distributed packet fair queueing in a scalable switch architecture," *IEEE INFOCOM*, San Francisco, CA, March 1998.

[20] X. Zhang and L. Bhuyan, "An Efficient Scheduling Algorithm for Combined-Input-Crosspoint-Queued (CICQ) Switches," *IEEE GLOBECOM*, Dallas, TX, November 2004.

[21] L. Mhamdi and M. Hamdi, "Output queued switch emulation by a one-cell-internally buffered crossbar switch," *IEEE GLOBECOM*, San Francisco, CA, Dec. 2003.

[22] J. Turner, "Strong performance guarantees for asynchronous crossbar schedulers," *IEEE/ACM Transactions on Networking*, to appear, 2009.

[23] E. Leonardi, M. Mellia, F. Neri, and M. A. Marsan, "On the stability of input-queued switches with speed-up," *IEEE/ACM Trans. Netw.*, vol. 9, no. 1, pp. 104-118, 2001.

[24] P.R. Kumar and S. P. Meyn, "Stability of queueing networks and scheduling policies," *IEEE Transactions on Automat. Control*, vol. 40, pp. 251-260, Feb. 1995.

[25] M. A. Marsan, et al. "Packet Scheduling in Input-Queued Cell-Based Switches," *IEEE INFOCOM*, Alaska, USA, April 2001.

[26] G. Passas, M. Katevenis, "Packet Mode Scheduling in Buffered Crossbar (CICQ) Switches," *Proc. IEEE Workshop on High Performance Switching and Routing (HPSR 2006)*, pp. 105-112, Poznan, Poland, June 2006.

[27] Rojas-Cessa, E. Oki, Z. Jing and H. J. Chao, "CIXB-1: Combined input-once-cell-crosspoint buffered switch," *IEEE Workshop on High Performance Switching and Routing*, Dallas, TX, July 2001.