On the interaction between TCP-like sources and throughput-efficient scheduling policies

On the interaction between TCP-like sources and throughput-efficient scheduling policies

Paolo Giaccone, Emilio Leonardi, Fabio Neri
Dipartimento di Elettronica, Politecnico di Torino (Italy)

Abstract

We focus on the dynamic interaction in packet networks between regulated Additive-Increase Multiplicative-Decrease (AIMD) traffic sources and *max-scalar* scheduling policies (such as the popular Maximum Weight Matching – MWM) at switches. The latter were proved to be optimal in terms of throughput for stationary unregulated traffic sources.

We describe the average dynamics of both traffic sources and switch queues through a system of Delay Differential Equations (DDEs), whose properties are throughly analyzed. Our study allows to gain important insights both on the system efficiency and on the long-term bandwidth share among connections.

Our main finding is that AIMD sources and *max-scalar* switches co-exist well.

Keywords: Packet networks, optimal throughput scheduling, TCP/AIMD sources, input queued switches, wireless systems.

1. Introduction

In recent years a significant effort has been devoted by the networking research community to the definition of efficient scheduling policies that maximize the system throughput in several application contexts, such as wireless, satellite networks and high-capacity switching architectures [1, 4, 20, 23, 27, 28, 32, 36]. This problem dates back to the early '90, when Tassiulas and Ephremides, in their seminal work [34], have first shown that the maximization of throughput in a *network of interacting queues* (also called *constrained queueing systems*) can be achieved with a dynamic scheduling policy according to which the selection of packet transmissions, at servers, is driven by the instantaneous queues state.

It is worth noticing that the scheduling policy proposed in [34], i.e., the so-called *max-scalar policy*, and its later extensions [1, 20, 23, 27, 28, 36] (such as the popular Maximum Weight Matching – MWM), do not require any a pri-

ori knowledge of the long-term traffic behavior, thereby appearing amenable for implementation in contexts in which traffic is highly dynamic and unpredictable.

Optimality of the *max-scalar* policy and its extensions has been proved, however, only under assumptions of stationarity and admissibility for the traffic flowing through the system of queues. It is not clear how optimal policies behave in the case of either non stationary, or rate-adaptive traffic sources, which may induce temporary overloads of some system architectural elements. The suspect that the *max-scalar* scheduling policy and its extensions may be strongly unfair in the latter case has probably refrained a massive deployment of such policies in commercial systems.

Only recently the attention has been turned to the analysis of the interaction between optimal dynamic policies and regulated traffic sources. Results in this field can have a great practical significance in consideration of the fact that the majority of Internet traffic sources adopt the Transmission Control Protocol (TCP) and dynamically adapt their sending rate to the estimated traffic congestion level according to an Additive-Increase Multiplicative-Decrease (AIMD) scheme.

In [11, 29] the behavior of max-scalar policies under regulated sources has been analyzed. However, the rate adaptation algorithms considered in [11, 29] significantly differ from the AIMD source behavior, since they require the sources to gather detailed and updated information about the network status.

We consider the dynamical behavior of max-scalar policies under rate-controlled sources executing an idealized TCP-like algorithm, driven only by losses and delay information as observed by the sources. The average dynamics of both sources and queues are described through a fluid model, i.e., a system of Delay Differential Equations (DDEs), whose qualitative properties are throughly analyzed. Our study allows to gain important insights both on the system efficiency and on the long-term bandwidth sharing among traffic flows.

Our findings are rather surprising and intriguing; the adoption of *max-scalar* scheduling policies along with carefully designed Active Queue Management (AQM) schemes permits to efficiently exploit the bandwidth of complex systems such as either Input Queued (IQ) switches or wireless cells, without negatively affecting the fairness of TCP flows.

Recently [33] and [37] showed that possible extreme unfairness and rate oscillations may occur at routers implementing a *max-scalar* scheduling policy when the traffic is originated by TCP sources. However we emphasize that, differently from our work, both [33] and [37] assume a classical drop-tail packet discarding policy at the queues. We believe that conjugating *max-scalar* scheduling policy with a properly designed AQM packet dropping scheme is necessary to achieve a good degree of fairness.

2. Systems of interacting queues

We consider a system of Q discrete time, interacting queues. This provides an abstract model for several different communication scenarios such as the system of transmission queues either at a wireless access point or a satellite, or the Virtual Output Queueing (VOQ) system in an IQ switch.

In discrete time, the queue evolution is described by:

$$x_q(n+1) = [x_q(n) + a_q(n) - \mu_q(n)]^+ \qquad 1 \leq q \leq Q,$$

where $x_q(n)$ represents the queue length, $a_q(n)$ represents the number of arrivals at the queue, $\mu_q(n)$ represents the amount of service received by queue q during time $(n, n+1]$ and $[x]^+$ denotes $\max\{0, x\}$. These quantities can be expressed either in packets or in bytes. The set of achievable queue service rates is subject to a set of physical constraints, such as those expressing either capacity limits or the effects of possible interference between signals.

We formalize the previous concepts saying that the vector of service rates $\mu(n) = (\mu_1(n) \cdots \mu_Q(n))$ belongs to a convex set of achievable rates $\mathcal{S}$, i.e., $\mu(n) \in \mathcal{S}$ with $\mathcal{S} \subset \mathbb{R}_+^Q$, for every n.

We consider three possible application scenarios:

- **Work-conserving server:** in this first simple case, the bandwidth $\hat{\mu}$ of a work-conserving server is dynamically shared among Q queues. The sum of service rates allocated to queues is bounded by the transmission capacity $\hat{\mu}$ of the server: $\mathcal{S} = \{\mu : \mu_q \geq 0 \text{ and } \sum_q \mu_q \leq \hat{\mu}\}$.

  We emphasize that this simple toy case has limited real interest, but its analysis can provide important insights on the behavior of more complex systems.

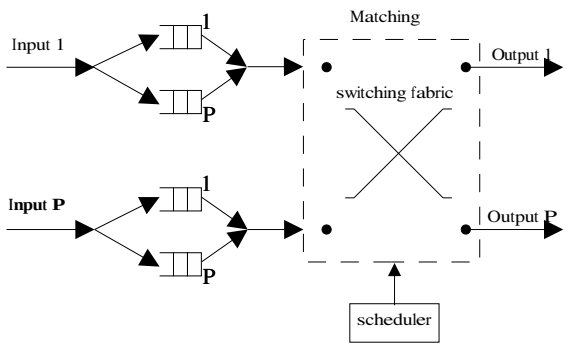

Figure 1: A $P \times P$ IQ switch architecture with VOQ

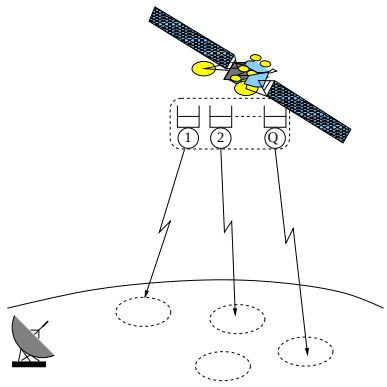

Figure 2: A wireless station

- **IQ switch:** Fig. 1 describes a $P \times P$ input queued switching architecture which represents the forwarding engine of a modern high-performance Internet router. An IQ switching architecture is a common forwarding engine of modern high-performance Internet routers. In a $P \times P$ IQ switch, each interface maintains P separate queues (called Virtual Output Queues - VOQ), one per output port ($Q = P^2$). The switching fabric operates in a synchronous fashion. At time slot n, a set $\pi(n)$ of non contending packets, called *matching*, is selected for transfer through the switching fabric. Matchings can comprise no more that one packet per input port and no more than one packet per output port. Denoting with $IQ(i)$ ($1 \leq i \leq P$) the set of VOQs at input i, and with $OQ(j)$ ($1 \leq i \leq P$) the set of queues directed to output j, in order to be feasible (i.e. a matching), the service vector $\mu(n)$ must satisfy the following capacity constraints:

$$\begin{cases} \mu_q(n) \in \{0,1\} & \forall q \\ \sum_{q \in IQ(i)} \mu_q(n) \leq 1 & 1 \leq i \leq P \\ \sum_{q \in OQ(j)} \mu_q(n) \leq 1 & 1 \leq j \leq P \end{cases}$$

  Finally, we denote with $\mathcal{S}$ the convex hull generated by the feasible service vectors, i.e., $\mathcal{S} = \{\mu : \mu_q \geq 0, \sum_{q \in IQ(i)} \mu_q \leq 1, \sum_{q \in OQ(j)} \mu_q \leq 1\}$

- **Wireless scenario:** it may represent a multi-beam satellite which transmits data to Q different ground locations (as depicted in Fig. 2) or a terrestrial wireless station, such as a node of a large-bandwidth WiMAX mesh network. Packets destined for each location are stored in separate queues. In this case the transmission rate vector $\mu(n)$ depends on the coding scheme and on the

power used for transmitting the signals. Using an access scheme which orthogonalizes transmitted signals, we can assume that the transmission rate μ_q depends only on the power P_q used for transmitting information from q, as in [28]. We further assume $\mu_q(P_q)$ to be a regular concave function. In addition, we assume that the total transmission power P_{tot} is bounded.

Hence, the set of possible transmission rates $\mathcal{S}$ is defined as the convex region defined by all the vectors $\mu = (\mu_1(P_1), \ldots, \mu_q(P_q), \ldots, \mu_Q(P_Q))$ where $\sum_q P_q \leq P_{tot}$.

The geometry of region $\mathcal{S}$ depends on the specification of functions $\mu_q(P_q)$, which in turn depend on the physical layer specification [28]. In this paper, just as matter of example, we assume $\mathcal{S}$ to be a circular region, defined by the following constraints:

$$\begin{cases} \mu_q(n) \geq 0 & \forall q \\ \sum_q \mu_q^2(n) \leq \hat{\mu} \end{cases} \tag{1}$$

where $\hat{\mu}$ is a positive constant. Note that this shape is an approximation of the achievable rate region in a multiple access channel MIMO system (cf. Fig. 8 in [13]) or in a fading environment (cf. Fig. 2.5 in [25]).

2.1. Max-Scalar policy definition

Under unregulated traffic sources, the problem of defining the optimal dynamic scheduling policy and the associated throughput region in complex systems of infinite-size interacting queues has attracted significant attention in the last decade from the research community since the pioneering work [34]. By assuming $a_q(n)$ to form a sequence of i.i.d. random variables, and applying the Lyapunov function methodology, it has been shown that a system of interacting queues achieves maximum throughput[1] if the max-scalar scheduling policy $\mathcal{P}_{MS}$ is applied. According to $\mathcal{P}_{MS}$, at each time slot n, the service vector is selected as follows:

$$\mu(n) = \arg\max_{\gamma \in \mathcal{S}} \sum_{q=1}^{Q} \gamma_q x_q(n) \tag{2}$$

The result in [34] has been generalized and adapted to different application contexts in recent years. As matter of example, we just briefly recall some of the

[1] A scheduling policy achieves maximum throughput if the system of queues is stable under any i.i.d. sequence $a_q(n)$ such that $(E[a_1(n)], E[a_2(n)], \cdots E[a_Q(n)]) \in \mathcal{S}$, i.e. the average arrival rates are within the set of admissible service rates.

related works. In the packet switching context, several studies aimed at the definition of the stability region in IQ switching architectures built around a buffer-less crossbar have appeared: papers [1, 20, 23, 32, 36] have proposed different extensions of $\mathcal{P}_{MS}$, which have been shown to achieve the maximum throughput; stability properties for simpler scheduling policies have been also studied in [10, 36]; in [2, 4, 20], finally, the problem of the definition of the stability region in networks of IQ switches has been considered. In the context of satellite and wireless networks, generalizations of $\mathcal{P}_{MS}$ have been proposed and shown to achieve the maximum throughput in [21, 27, 28, 35]. Finally, [9] has generalized the result in [34] under more general exogenous arrival processes applying a different analytical technique called fluid models. All previous works, however, have considered unregulated stationary traffic sources.

2.2. $\mathcal{P}_{MS}$ in three application scenarios

It is not difficult to particularize the policy $\mathcal{P}_{MS}$ for the previous three scenarios:

- **Work-conserving server**: $\mathcal{P}_{MS}$ selects simply the queue with the largest queue size i.e., this policy is usually referred as Longest Queue First (LQF) scheduling.
- **IQ switch**: $\mathcal{P}_{MS}$ selects the matching $\pi(n)$ to maximize $\sum_{q\in\pi(n)} x_q(n)$, thus degenerates in the popular maximum weight matching.[2]
- **Wireless scenario**: Assuming $\mathcal{S}$ to be defined according to (1), $\mathcal{P}_{MS}$ selects the maximum service vector $\mu(n) \in \mathcal{S}$ parallel to the queue size vector $X(n)$:
$$\mu_q(n) = \frac{x_q(n)}{\|X(n)\|_2}\sqrt{\hat{\mu}} \tag{3}$$
where $\|X(n)\|_2$ is the square norm of vector $X(n)$, i.e., $\|X(n)\|_2 = \sqrt{\sum_q x_q(n)^2}$.

2.3. $\mathcal{P}_{MS}$ under regulated sources: previous work

Only recently the attention has been turned to the analysis of the interactions between optimal dynamical policies and adaptive traffic sources. Papers [11, 29] have shown that $\mathcal{P}_{MS}$ behaves well in presence of regulated sources, and guarantees an acceptable degree of fairness to flows also in case of temporary overload.

[2]Note that, according to $\mathcal{P}_{MS}$, $\mu(n) = \arg\max_{\gamma\in\mathcal{S}} \sum_{q=1}^{Q} \gamma_q x_q(n)$ can always be selected to be an integer-valued vector; this is a consequence of the unimodularity of region $\mathcal{S}$.

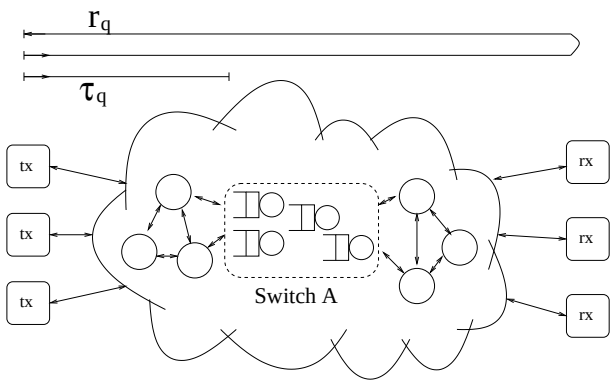

Figure 3: The system under study

However, these papers have focused on congestion control mechanisms significantly different from TCP, which require that traffic sources strictly interact with the network and gather detailed and up-dated information about the queues status.

Let $\rho_q(n)$ be the aggregate arrival rate at queue q, equal to the overall sending rate at the corresponding sources. In [11] sources were assumed to adjust $\rho_q(n)$ on the basis of the instantaneous queue size $x_q(n)$:

$$\rho_q(n) = \frac{\alpha_q K}{x_q(n) + \gamma_q}$$

where α_q, K and γ_q are suitable positive constants. Similarly, in [29] the rate of sources must be dynamically adapted on the basis of instantaneous queues lengths. According to one of the proposals in [29]:

$$\rho_q(n) = \min\left\{\left[\frac{V}{2x_q(n)} - 1\right]^+, \rho_{max}\right\}$$

where V is a control parameter and ρ_{max} is the maximum allowed source sending rate.

In both cases above, sources must be made aware of queue sizes at switches.

3. Our system

We consider a TCP/IP infrastructure comprising a set of hosts interconnected through a network of switches/routers as depicted in Fig. 3. In particular we focus on the network element identified in the figure as “switch A”. Switch A represents either an IQ switch or a wireless station implementing the max-scalar scheduling

policy, and acts as bottleneck for traffic flows traversing it (i.e., queuing and losses processes at switches/routers different from A, depicted as circles in Fig. 3, introduce negligible effects on TCP dynamics).

3.1. The system model

Aim of our work is to study the behavior of AIMD-based congestion control mechanisms in networks of interacting queues.

In our analysis, each queue q is fed with traffic originated in a set of M_q TCP sources. To study the interactions between sources and queues, we adopt a continuous time fluid approach [26] in which the average dynamics of both sources and queues are described by deterministic delay differential equations.

We assume that all the M_q TCP sources feeding queue q experience a constant round trip time r_q (see Fig. 3). The fluid evolution of the average window size $w_q(t)$ is driven by the classical, well-known AIMD fluid equation [26], derived by the saw tooth behavior of the window:

$$\frac{dw_q(t)}{dt} = \frac{1}{r_q} - \frac{w_q(t)}{2}\phi_q(t) \tag{4}$$

where $\phi_q(t)$ represents the rate of congestion indications experienced at time t by sources. The first term on the right-hand side represents the additive increase mechanism, while the second term represents the multiplicative decrease contribution.

We denote with $W(t) = (w_1(t), w_2(t), \cdots w_Q(t))$ the vector whose elements represent the average transmitter window sizes (modeling TCP congestion windows) at time t for sources feeding queue q.

The fluid evolution of queue lengths $x_q(t)$ is driven by the following equations:

$$\frac{dx_q(t)}{dt} = \Big[\frac{M_q}{r_q}w_q(t-\tau_q) + \lambda_q(t)\Big](1-d_q(t)) - \mu_q(t) \quad \text{if } x_q(t) > 0 \tag{5}$$

$$\frac{dx_q(t)}{dt} = \max\Big\{0, \Big[\frac{M_q}{r_q}w_q(t-\tau_q) + \lambda_q(t)\Big](1-d_q(t)) - \mu_q(t)\Big\} \text{ if } x_q(t) = 0 \tag{6}$$

The first term on the right of (5) represents the aggregate arrival rate at queue q; τ_q is the average propagation delay between sources and queue q; $M_q w_q/r_q$ is the overall average sending rate of the M_q sources; $\lambda_q(t)$ is the aggregate arrival rate of unregulated traffic; $d_q(t)$ is the dropping probability at buffer q; $\mu_q(t)$ is the service rate of queue q at time t. We denote with $X(t) = (x_1(t), x_2(t), \cdots x_Q(t))$ the vector whose elements represent queues lengths at time t. Furthermore, in the following denote with k_q the ratio M_q/r_q.

We suppose that each queue q implements a RED/ECN [12] AQM scheme, according to which packets are in general either dropped or marked. Both dropping probability $d_q(t)$ and marking probabilities $m_q(t)$ are driven by the buffer level. In particular, AQM schemes usually maintain a moving average $\hat{x}_q$ of the instantaneous queue size x_q, updated whenever a packet arrives according to the rule:

$$\hat{x}_q \leftarrow (1-z)\hat{x}_q + zx_q$$

The instantaneous mark/drop probability is computed as a function of $\hat{x}_q$ according to some relation $d_q(t) = f_d(\hat{x}_q(t))$ and $m_q(t) = f_m(\hat{x}_q(t))$ (for example, $m_q(t) = 0$ in the case of a pure dropping policy). For fluid modeling, we need a characterization of the temporal evolution of the moving average $\hat{x}_q(t)$ as a continuous function of time. This was originally done in [26], where the authors have shown that the evolution is represented by the differential equation:

$$\frac{d\hat{x}_q(t)}{dt} = \frac{\log(1-z)}{\delta(t)}\hat{x}_q(t) - \frac{\log(1-z)}{\delta(t)}x_q(t) \tag{7}$$

if $z < 1$ and $\hat{x}_q(t) = x_q(t)$ if $z = 1$; $\delta(t)$ is the average packet inter-arrival time, i.e., $\delta(t)^{-1} = k_q w_q(t-\tau_q) + \lambda_q(t)$.

We denote with $\hat{X}(t) = (\hat{x}_1(t), \hat{x}_2(t), \ldots, \hat{x}_Q(t))$ the vector whose elements represent the moving average of the instantaneous queue size.

The rate of congestion indications $\phi_q(t)$ experienced by sources at time t is given by:

$$\phi_q(t) = \frac{w_q(t-r_q)}{r_q}\Big[f(\hat{x}_q(t-r_q+\tau_q))\Big] \tag{8}$$

where $w_q(t-r_q)/r_q$ is the packet sending rate of sources at time $t-r_q$ and $f(y) = f_d(y) + f_m(y)$ is the sum of packet marking and dropping probabilities for buffer size equal to y.

Finally, according to the definition of $\mathcal{P}_{MS}$, queue service rates are determined in the fluid model as the solutions of:

$$\mu(t) = \arg\max_{\gamma\in\mathcal{S}} \sum_{q=1}^{Q} \gamma_q x_q(t) \tag{9}$$

where $\mu(t) = (\mu_1(t), \mu_2(t), \cdots \mu_Q(t))$ is the fluid service vector and $\mathcal{S}$ is the set of feasible service vectors.

Finally, absolutely continuous functional vector $(X(t), \hat{X}(t), W(t))$ satisfying (4), (5), (7) and (9) represents a solution of the above dynamic system.[3].

[3]Note that $\mu(t)$, is, by construction, continuous at every regular points of $(X(t), \hat{X}(t), W(t))$.

3.2. A critical discussion of the assumptions

In this sub-section, we critically discuss the assumptions and approximations of our model. First of all, we adopt a fluid approach to model both sources and queues dynamics. Indeed, several recent works [3, 7, 15, 16, 26] have clearly shown that the fluid approach is a viable alternative to detailed packet-level simulations for the analysis of large-bandwidth IP networks (i.e., supporting a large number of TCP flows). Moreover, fluid models were proved to be effective for the parameter design of AQM/ECN schemes in TCP/IP networks [18].

Fluid model equations can be formally derived from the jump process describing packet level dynamics through *fluid scaling* limits [6, 7, 9, 10]. This process permits to tightly relate the qualitative properties of fluid models to those of the original system [6, 9, 10, 31]. In particular, [9, 10] have shown that the throughput performance of routers implementing max-scalar scheduling policies can be derived from the analysis of the qualitative properties of fluid model solutions.

In this paper we skip a formal derivation of the fluid equations, that can follow exactly the same approach of [6, 7, 9, 10]: in Appendix A we report a brief overview on how to derive the fluid equations. We concentrate our investigation on the analysis of the fluid model properties.

Since our goal is to analytically study the interaction between TCP sources and max-scalar scheduling at nodes, enlightening structural properties of the system as a whole, we have tried to simplify as much as possible the description of every architectural element. This is the reason why we have modeled just the basic AIMD mechanism of TCP, ignoring slow-start, time-outs, etc. We notice, however, that this basic description of idealized AIMD sources is usually able to capture the dominant dynamics of the system, providing fairly accurate results in several scenarios [7, 15].

We restrict our analysis to long-lived connections, neglecting short-lived connections. This is essentially due to the fact that short-lived flows can be assimilated in the fluid model to unregulated flows (term $\lambda_q(t)$ in (5)) as shown in [8, 17], since the effect of AIMD congestion control feedback is not effective due to their limited durations.

We have neglected the effects of variable queuing delays on the round-trip time. This assumption does not affect the system equilibrium points derived in our main Theorem 1 in Sec. 4.1. It is, however, needed to simplify the stability analysis of equilibrium points. In Sec. 6 we validate the whole analytical model by comparing it with an accurate simulation model of a TCP/IP network, that takes into accounts

Thus even if (9) may potentially admit an infinite number of solutions, only one can grant continuity of $\mu(t)$ at regular points of $(X(t), \hat{X}(t), W(t))$

all the delay effects.

In (4) and (8) we have implicitly assumed that all the M_q TCP sources feeding queue q experience the same propagation delay. This assumption can be relaxed to the case in which sources feeding queue q may experience different propagation delays, but the dispersion in their values is not too large. In this case r_q and τ_q appearing in (4) and (8) can be reinterpreted in terms the average values. The model can be generalized when several classes of TCP source with significant different propagation delay coexist among the M_q sources feeding queue q. In this case an equation for each class of sources [15] has to be written. However in this paper we do not address this extension.

At last, note that in fluid models the packet by packet description of the system dynamics is completely neglected. The dynamics captured by the fluid model are those operating on the same timescales of the TCP control mechanism (around tens/hundreds milliseconds). This means that essentially (4) represents the average rates obtained by different flows according to the max-scalar policy over time-periods whose duration is comparable with the time-scale of the TCP dynamics.

4. Qualitative study of the model solutions

Now, we characterize the qualitative properties of the model solutions of the above system of differential equations.

First, we investigate on the existence of equilibrium points (i.e., stationary solutions) under the assumption of stationary traffic conditions, i.e. $\lambda_q(t) = \lambda_q \; \forall q$. We show that under mild assumptions a unique equilibrium point always exists. Then we turn our attention to the problem of the equilibrium point attractiveness (stability). We conjecture that, by carefully designing the AQM scheme equilibrium, global attractiveness can be obtained. In simple cases we are able to analytically prove the local attractiveness to the equilibrium point, while in more complex cases we report numerical results in support of our thesis. Unfortunately, the problem of establishing global attractiveness is very difficult, and has only received partial answers for the simple case of TCP flows feeding a FIFO queue [18].

We emphasize that a unique, globally attractive, equilibrium point unequivocally determines the long-term behavior (i.e., for $t \to \infty$) of system dynamics. As a consequence, by looking at the equilibrium point, we gain important insights on the system efficiency and long-term bandwidth share among connections, as shown in the next section.

4.1. System equilibrium point

The following statement fully characterizes the equilibrium points of our dynamical system.

Theorem 1. *Consider a network with the following assumptions: (i) for every q the arrival unregulated traffic rate is stationary, i.e., $\lambda_q(t) = \lambda_q$; (ii) $\mathcal{S}$ is a convex compact set in $\mathbb{R}_+^Q$ with non null interior; (iii) $f_m(y)$ and $f_d(y)$ are non decreasing continuous and differentiable functions; (iv) for some finite B_q, $f_d(B_q) = 1$; (v) $f(y) = f_m(y) + f_d(y)$ is strictly increasing for $0 \leq y \leq B_q$, with $f(0) = 0$. In this case the system of differential equations (4), (5), (7) and (9) admits a unique stationary solution $(X^*, \hat{X}^*, W^*)$ satisfying the following conditions:*

$$\mu^* = \arg\max_{\mu \in \mathcal{S}} G(\mu) \tag{10}$$

$$\hat{x}_q^* = x_q^* \tag{11}$$

$$x_q^* = f^{-1}(h_q(\mu_q^*)) \tag{12}$$

$$w_q^* = \sqrt{\frac{2}{f(x_q^*)}} \tag{13}$$

where

$$G(\mu) = \sum_q \int_0^{\mu_q} h_q(\alpha) d\alpha$$

and $h_q(\alpha)$ is the only positive solution of the equation:

$$\left(\frac{k_q \sqrt{2}}{\sqrt{f(x_q^*)}} + \lambda_q \right) \left(1 - f_d(x_q^*)\right) = \alpha$$

The proof is given in Appendix B.

4.2. Stability analysis of the equilibrium point

Our conjecture is that, under reasonable traffic conditions, *the equilibrium point can be made attractive by carefully designing $f_d(x)$, $f_m(x)$ and z.*

Even if we are unable to provide a general formal proof of our claim, we report a wide range of partially numerical and analytical results in support of our thesis.

We start analyzing the simplified case in which delays to propagate packets from the sources to the queue in (5) are neglected along with delays to propagate congestion signals in (8). In this way the dynamical system described by (4), (5), (7) and (8) becomes a more treatable pure ordinary differential system (with no delays). Furthermore we assume $z = 1$; hence, $\hat{x}_q(t) = x_q(t)$ and system solutions are unequivocally determined by the vector $(X(t), W(t))$ describing queues and windows dynamics. Under these assumptions we are able to formally prove local stability of the equilibrium point in the three previously considered scenarios.

4.2.1. Work-conserving server

The local asymptotic stability of the equilibrium point can be proved, in this case, by using the Lyapunov function technique.

Suppose that the system is at $t = 0$ in an initial state $(X(0), W(0))$ sufficiently close to the equilibrium point (X^*, W^*). We denote with $(X(t), W(t))$ the trajectory of the system and consider the following functional (Lyapunov function):

$$\mathcal{L}(X(t), W(t)) = \max_q (x_q - x_q^*)^2 + \beta \sum_q M_q (w_q - w_q^*)^2$$

which represents a sort of "distance" between the current state and the equilibrium point. Note that by definition: i) $\mathcal{L}(X(t), W(t)) \geq 0$; ii) $\mathcal{L}(X(t), W(t)) = 0$, if and only if $(X(t), W(t)) = (X^*, W^*)$. Since $\frac{d\mathcal{L}(X(t),W(t))}{dt} < 0$, for almost every $t > 0$ (as discussed in Appendix C), we can conclude that the "distance" between the current system state and the equilibrium point is reducing with time (i.e., the trajectory gets closer and closer to the equilibrium point).

Beyond local stability, for different parameters settings, starting from 100 randomly chosen initial conditions, in all cases we have numerically observed the convergence toward the equilibrium point of the solutions of the simplified (no delays) dynamical system of equations.

4.2.2. IQ switch

In case of the IQ switch, a formal proof can be done only in the special case of a 2×2 IQ switch by repeating arguments similar to the previous case and using the following Lyapunov function:

$$\mathcal{L}(X(t), W(t)) = \max_\pi (U(\pi) - U^*)^2 + \beta \sum_q M_q (w_q - w_q^*)^2 + \gamma \sum_q (x_q - x_q^*)^2$$

where π is one of the $P!$ possible matchings, $U(\pi)$ its corresponding weight, and U^* the weight of the MWM at the equilibrium. Also in this scenario, for several different parameters settings, starting from 100 randomly chosen initial conditions we have always observed the numerical convergence of the solutions toward the equilibrium point. The experiment was repeated both for 2×2 and 4×4 IQ switches.

4.2.3. Wireless scenario

In this case the local asymptotic stability of the system can be proved by linearizing the system around the equilibrium point and checking the stability of the linearized system. We report a sketch of the stability proof for the linearized system in Appendix D.

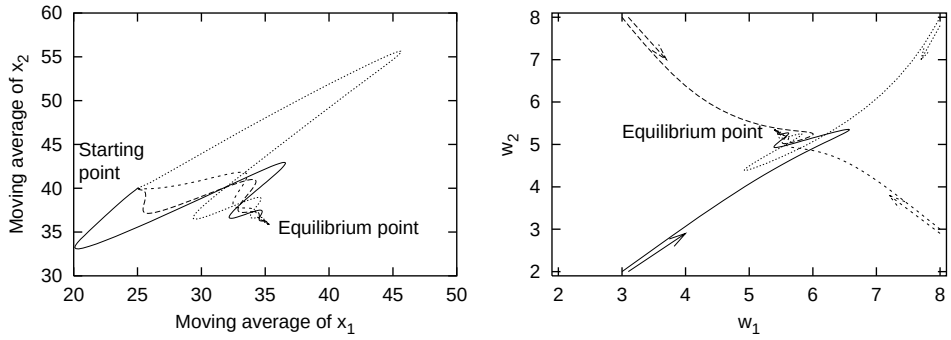

Figure 4: Dynamic behavior of a LQF server with two RED queues. The trajectories on the left graph show the moving average of the the queue lengths, whereas the trajectories on the right graph show the window sizes.

Also in this case, numerical experiments have shown that solutions converge to the equilibrium point starting from randomly chosen initial conditions, suggesting a global form of attractiveness for the equilibrium point.

4.3. Considering delays for the congestion signals

When considering the delays to propagate packets from sources to queues in (5), and to propagate congestion signals in (8), the problem of defining general conditions under which the equilibrium point is attractive becomes harder.

In the simpler case in which TCP sources interact with a single FIFO server implementing an AQM scheme, sufficient conditions for local stability have been obtained linearizing the system fluid equations in [18]. As a result, paper [18] provides guidelines for the design of AQM parameters based upon simple relations to the physical parameters of the system such as the number of interacting TCP connections, the round-trip time and the capacity of the queue. However, the same guidelines cannot be applied in our scenarios.

Fig. 4 shows the dynamics of a work conserving LQF server managing two RED queues ($Q = 2$). The trajectories were obtained by solving numerically the original system of delayed differential equations in (4)-(9), using an ad-hoc solver developed in C. Physical parameters were: $M_1 = M_2 = 100$, $r_1 = r_2 = 10$ ms, $\tau_1 = \tau_2 = 0.4$ and $\hat{\mu} = 10^5$ pkt/s. Packets were dropped according to RED loss profile. No marking was allowed. The RED parameters were: $\min_{th} = 100$, $\max_{th} = 500$ packets, $p_{\max} = 0.1$, $z = 10^{-4}$. Four trajectories corresponding to four different initial conditions are plotted in Fig. 4. All the trajectories converge to the equilibrium point.

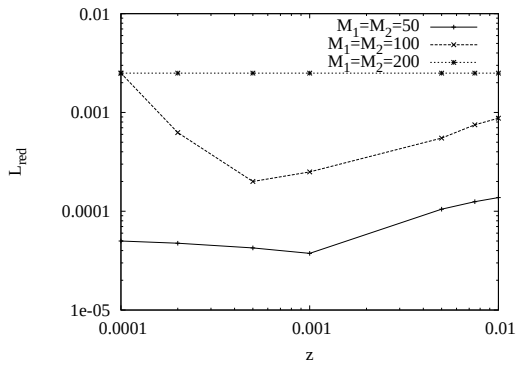

Figure 5: Design guidelines for two RED queues served by a LQF server. Stabilizing pairs (z, L_{red}) lie under the curves.

Under the same scenario, we investigated the settings of the RED parameters for which the system achieves an equilibrium point. In Fig. 5 we show the stability regions, for different RED parameters and number of TCP flows. Coherently with the approach followed in [18], we plot the stability regions in function of z and L_{red}, defined as $L_{red} = p_{max}/(\max_{th} - \min_{th})$. Given each specific z value, we implemented a binary search to find the maximum value of p_{max} for which an equilibrium point was reached in less than 3 seconds. Compared with Fig.12 of [18], the stability regions show a peculiar behavior: for large number of flows, any value of $p_{max} \in (0, 1)$ allows the system to converge, whereas for smaller number of flows, the maximum allowed p_{max} shows a non-monotonic behavior. *This shows that the design guidelines for a single FIFO queue cannot be applied in our scenarios.* The investigation and proper characterizations of the design criteria for our scenarios have been left for future investigation.

5. System performance and fairness

In this section we explicitly characterize the equilibrium point for the three previously defined scenarios, analyzing system performance and fairness.

Before proceeding, however, we need to agree on an acceptable definition of fair bandwidth allocation to TCP flows. This choice is rather critical in light of the fact that no global consensus exists in the networking community on what a fair allocation is. Our opinion is that, in the Internet context, a good reference model is constituted by the bandwidth share obtained by TCP connections traversing a FIFO buffer. In such a case, bandwidth is evenly distributed by the system to homogeneous connections (i.e., connections with the same round-trip time), while

bandwidth is distributed among inhomogeneous connections (i.e., connections with different round-trip times) proportionally to the inverse of the connection round trip time, thereby achieving a rough form of proportional fairness. In the following, we will qualify the above reference bandwidth allocation as the "fair allocation".

5.1. Work-conserving server

In this simple case, it is rather straightforward to obtain that the equilibrium point defined by (10)-(13) shows very surprising properties:

$$x_q^* = x^* \qquad w_q^* = w^* \qquad \forall q \tag{14}$$

$$\mu_q^* = (k_q w^* + \lambda_q)(1 - f_d(x^*)) \tag{15}$$

i.e., at the equilibrium queues have the same length, sources have the same average window size, and service is provided to regulated traffic aggregates proportionally to parameter [4] k_q.

The values for x^* and w^* can be explicitly computed; for example, in case $f_m(x_q) = 0$ and $\lambda_d = 0$; they are given by:

$$x^* = f_d^{-1}\Big(\frac{4 + \beta^2 - \sqrt{\beta^4 + 8\beta^2}}{4}\Big)$$

$$w^* = \frac{\beta + \sqrt{\beta^2 + 8}}{2}$$

where $\beta = \dfrac{\hat{\mu}}{\sum_q k_q}$.

The long-term per-flow throughput s_q is determined by the parameters at the equilibrium point through the simple relation

$$s_q = \frac{w^*}{r_q}(1 - f_d(x^*)) \tag{16}$$

Hence, the system bandwidth is distributed among connections proportionally to the inverse of the round-trip time, guaranteeing the same average share to homogeneous connections. *Thus, LQF provides the same long term bandwidth share among connections that we expect when adopting a conventional FIFO policy at the buffer!*

As final remark we notice that, since $x_q^* > 0$ (see proof of Theorem 1), it follows $\sum_q \mu_q^* = \hat{\mu}$, and consequently, the system is always able to efficiently exploit the available bandwidth.

[4] The fact that all queues are of the same length at the equilibrium immediately derives from the fact that all service rates μ_q are different from 0 at the equilibrium (see proof of Theorem 1). As a consequence sources experience the same marking/loss probability, and thus, have the same window size.

5.2. IQ switch

In this case, the analytical characterization of the equilibrium point requires the solution of a system of $2P^2 + 2P - 1$ non-linear equations. To simplify the analysis, we assume that all the Virtual Output Queues are fed by some regulated traffic sources (i.e., for every q, $k_q > 0$); we however emphasize that the analysis can be easily extended to the more general case.

First we point out that the equilibrium point must satisfy the following important property: *the weight $U(\pi)$ of any possible matching π, is always equal to U^* at the equilibrium*, i.e.:

$$U(\pi) = \sum_{q \in \pi} x_q^* = U^* \qquad \forall \pi$$

This property generalizes the property exhibited by LQF in the work-conserving queue.

As a consequence, X^* lies in the linear span of $\mathcal{M} \subset \mathbb{R}_+^{N^2}$, with $\mathcal{M} = \{I^1, I^2, \ldots, I^P, O^1, O^2, \ldots, O^P\}$, where I^p is a vector whose q-th element I_q^p is one if $q \in IQ(p)$ and null otherwise; and O^p be a vector whose element O_q^p is one if $q \in OQ(p)$, and null otherwise. Dimension of span($\mathcal{M}$) is $2P - 1$, hence X^* can be expressed as a linear combination of $2P - 1$ vectors selected within $\mathcal{M}$. Choosing the first $2P - 1$ independent vectors in $\mathcal{M}$, we can write:

$$X^* = \sum_{p=1}^{P} \alpha_p I^p + \sum_{p=1}^{P-1} \beta_p O^p \tag{17}$$

for some positive values of the parameters α_p and β_p.

On the other hand, rates μ_q^* and queue sizes x_q^* are deterministically related by the following systems of non-linear equations:

$$\left(\frac{k_q \sqrt{2}}{\sqrt{f(x_q^*)}} + \lambda_d \right) (1 - f_d(x_q^*)) = \mu_q^* \;\; 1 \leq q \leq Q \tag{18}$$

We notice that, since at the equilibrium point all the queues are non empty (i.e., $x_q^* > 0$, $\forall q$), the service rate vector at the equilibrium maximizes the global throughput i.e., $\sum_{q \in IQ(i)} \mu_q^* = \sum_{q \in OQ(j)} \mu_q^* = 1$ for every input i and output j.

The properties of the equilibrium point suggest that also in this case the system tends to evenly distributing the bandwidth among homogeneous TCP connections (at the equilibrium), while it tends to distribute the bandwidth among inhomogeneous connections proportionally to the inverse of their round-trip delay. This feeling is confirmed by our numerical experiments; we have focused on a 4×4

IQ switch, loaded with inhomogeneous connections according to the following class of scenarios, described by connection matrix $M = [M_{ij}]$ and RTT matrix $R = [R_{ij}]$:

$$M = \begin{pmatrix} m_0 & \alpha m_0 & \alpha^2 m_0 & \alpha^3 m_0 \\ \alpha^3 m_0 & m_0 & \alpha m_0 & \alpha^2 m_0 \\ \alpha^2 m_0 & \alpha^3 m_0 & m_0 & \alpha m_0 \\ \alpha m_0 & \alpha^2 m_0 & \alpha^3 m_0 & m_0 \end{pmatrix}$$

$$R = \begin{pmatrix} r_0 & \beta r_0 & \beta^2 r_0 & \beta^3 r_0 \\ \beta^3 r_0 & r_0 & \beta r_0 & \beta^2 r_0 \\ \beta^2 r_0 & \beta^3 r_0 & r_0 & \beta r_0 \\ \beta r_0 & \beta^2 r_0 & \beta^3 r_0 & r_0 \end{pmatrix}$$

where M_{ij} is the number of connections flowing from input i to output j, and R_{ij} is the corresponding average round-trip time; m_0, r_0, α and β are free positive parameters. We have tried several cases for different values of α and β ranging in the interval $[1, 3]$. In all cases the numerical results showed that relative bandwidth obtained by connections at the equilibrium is perfectly proportional to $1/\beta$, independently from the actual values of the other scenario parameters.

We emphasize that not always a perfectly "fair" (in the previously specified sense) distribution of the bandwidth is achieved in IQ switches. Bandwidth shares among TCP flows deviate from the "fair" distribution when traffic asymmetries among inputs or outputs ports are established (note that previous traffic patterns were completely symmetrical with respects to both inputs and outputs ports). We notice that, in the latter cases, forcing a "fair" distribution of bandwidth among TCP flows would cause a not complete exploitation of the switch bandwidth.

To better understand the behavior of the max-scalar policy, consider a traffic scenario comprising *homogeneous* TCP flows (i.e., the same RTT matrix R as before but with $\beta = 1$) and a different connection matrix M', in which the number of connections at input 4 is increased by parameter $\gamma > 1$ with respect to M:

$$M' = \begin{pmatrix} m_0 & \alpha m_0 & \alpha^2 m_0 & \alpha^3 m_0 \\ \alpha^3 m_0 & m_0 & \alpha m_0 & \alpha^2 m_0 \\ \alpha^2 m_0 & \alpha^3 m_0 & m_0 & \alpha m_0 \\ \gamma\alpha m_0 & \gamma\alpha^2 m_0 & \gamma\alpha^3 m_0 & \gamma m_0 \end{pmatrix}$$

The results are shown in the following matrix $T = [T_{ij}]$, where T_{ij} is the average throughput experienced at VOQ$_{ij}$:

$$T = \begin{pmatrix} \zeta & \zeta & \zeta & \zeta \\ \zeta & \zeta & \zeta & \zeta \\ \zeta & \zeta & \zeta & \zeta \\ \zeta/\gamma & \zeta/\gamma & \zeta/\gamma & \zeta/\gamma \end{pmatrix}$$

where $\zeta = \dfrac{1}{m_0(\alpha^2+1)(\alpha+1)}$.

We notice that flows traversing input port 4 are penalized in throughput with respect to other flows; this effect however has an easy explanation: connections traversing input 4 are bottlenecked at the input port where they obtain the maximum possible "fair" share. All the other connections evenly share the residual switch bandwidth. Note that a perfectly "fair" distribution of bandwidths is possible only at the cost of reducing the throughput of some connections (those not traversing input 4) without any beneficial effect on the other connections (those traversing input 4). Thus the system, in this case, distributes bandwidth according to a max-min fair scheme.

Several other numerical experiments (whose results are not reported for brevity) have confirmed that the *max-scalar policy at the equilibrium distributes bandwidth among the connections according to a weighed max-min fairness scheme in which connection weights are proportional to the inverse of the connection round-trip time*.

5.3. Wireless scenario

In this case, as already observed, queue services are provided, by the max-scalar policy, proportionally to the queue lengths; at the equilibrium:

$$\mu_q^* = \frac{x_q^*}{\|X^*\|_2}\hat{\mu}$$

The quantities x_q^* and w_q^* can be obtained solving the following system of $2q$ non-linear equations, $\forall q$:

$$\begin{cases} (k_q w_q^* + \lambda_q)(1 - f_d(x_q^*)) = \dfrac{x_q^*}{\|X^*\|_2}\hat{\mu} \\ w_q^* = \sqrt{\dfrac{2}{f(x_q^*)}} \end{cases}$$

Note that, differently from what happens for a work-conserving server, in this scenario the rule according to which bandwidth is subdivided among connection aggregates has an impact on the global system throughput. This is due to the fact that vectors which lie on the boundary of region $\mathcal{S}$ (i.e., vectors satisfying $\sum_q (\mu^*{}_q)^2 = \hat{\mu}$) do not correspond to the same global system throughput $\sum_q \mu_q^*$. Maximum system throughput is achieved when all the aggregates receive the same bandwidth ($\mu_q^* = \hat{\mu}/Q$ for every q), irrespectively of their parameters M_q and r_q. We notice, however, that in this case a significantly unfair distribution of throughputs to individual connections may result when parameters M_q, r_q or λ_q are

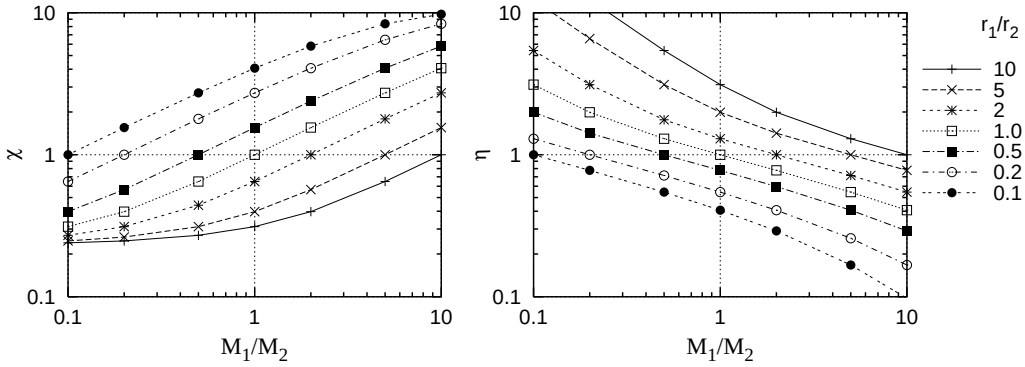

Figure 6: Wireless scenario with TCP traffic only: server unbalance χ (on the left) and fairness index η (on the right) vs. M_1/M_2, for different values of the ratio r_1/r_2.

strongly unbalanced. On the contrary the system that fairly distributes bandwidth to connections (i.e. obeys to the law $\frac{\mu_q^* - \lambda_q}{\mu_{q'}^* - \lambda_{q'}} = k_q/k_{q'}, \forall q, q'$) may be significantly inefficient in terms of system throughput.

We consider a wireless station hosting RED two queues (i.e., $Q = 2$), to ease the graphic presentation of numerical results. However all considerations apply to the more general case $Q > 2$. To obtain the solutions of the system of non-linear equations, we have developed an an-hoc solver based on GSL library [14].

We focus on two performance indexes. First, the *fairness index* η defined as the ratio between throughput-delay products of TCP flows belonging to first and the second aggregate $\eta = r_1 s_1 / r_2 s_2$, having defined s_q according to (16). Note that $\eta = 1$ corresponds to a "fair" distribution of bandwidth among TCP flows. Second, we consider the *service unbalance* χ among the two queues, defined as $\chi = \mu_1^* / \mu_2^*$.

We have fixed the service rate $\sqrt{\hat{\mu}} = 10^5$ packets/s, $M_1 = 100$ flows and $r_1 = 10$ ms. We have varied M_2 between 10 to 1000 flows, and r_2 between 1 and 100 ms, for a total of 49 different settings.

First we consider the case in which the wireless server is fed by regulated traffic, only i.e., $\lambda_1 = \lambda_2 = 0$. Fig. 6 shows η (on the right) and χ (on the left), in function of the ratio M_1/M_2, for different values of the ratio r_1/r_2.

Bandwidth is distributed between aggregates in a rather complex way, in this case; an aggregate gets more and more bandwidth by the system when the relative number of its flows increases. However, since χ exhibits a sub-linear dependence from parameter M_1/M_2, flows belonging to more numerous aggregates are penalized with respect to flows belonging to less numerous aggregates, as shown in

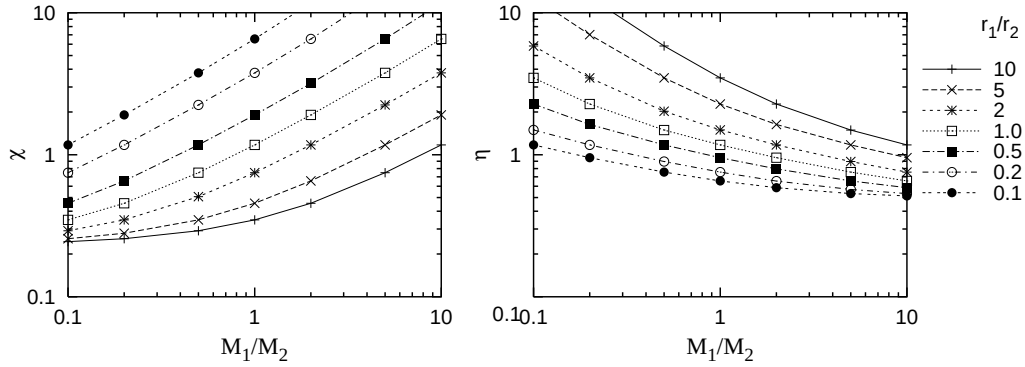

Figure 7: Wireless scenario with unregulated traffic ($\lambda_2 = \sqrt{\hat{\mu}}/3$): server unbalance χ (on the left) and fairness index η (on the right) vs. M_1/M_2, for different values of the ratio r_1/r_2.

Fig. 6. Bandwidth shares obtained by connections depend also on round trip-time: larger bandwidth is obtained by aggregates of connections with shorter round-trip times. However, since the flow bandwidth share increases sub-linearly with the inverse of round-trip time, flows with shorter round-trip get less than their "fair" share.

We consider now the case in which unregulated traffic arrives at queue 2 ($\lambda_2 = \sqrt{\hat{\mu}}/3$). Fig. 7 reports χ and η. With respect to the previous case, the presence of unregulated traffic at queue 2 induces a moderate perturbation of the relative connections share, penalizing flows belonging to aggregate 2.

Furthermore, we investigated the effect of $p_{\max}$, which is proportional to the slope of the RED loss profile, i.e. to the "aggressiveness" of the AQM scheme. We consider the original scenario with only regulated traffic, in which we set $r_1 = 10$ ms and $r_2 = 100$ ms, i.e. the case $r_1/r_2 = 0.1$ in Fig. 6. Now Fig. 8 shows both the server unbalance and the fairness index for different values of $p_{\max}$ and different values of M_1/M_2. Note that k_q represents also the minimum arrival rate (i.e. obtained for $w_q = 1$) at queue q. When M_1/M_2 is small, then $k_1 \approx k_2$; this condition corresponds to a symmetric scenario (even if heterogeneous) and the server serves both queues almost at the same rate and achieves the maximum fairness, independently from $p_{\max}$. However, for large values of M_1/M_2, $k_1 \approx 100k_2$ and the server serves the first queue most of the time. This unbalance is exacerbated by a more aggressive RED that penalizes the queue with smaller k_q, i.e. with smaller arrival rate. This is also confirmed by the fairness index, which varies by a factor 2 for different values of $p_{\max}$; smaller values of $p_{\max}$ appear to be preferable. We investigated also other scenarios with smaller r_1/r_2 ratio, and

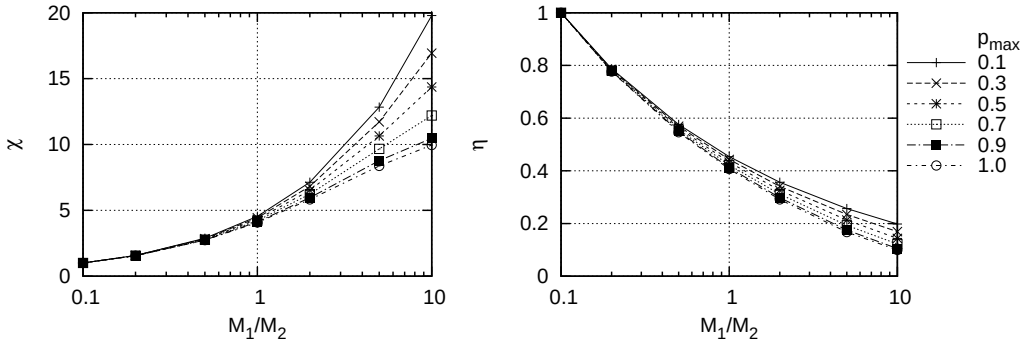

Figure 8: Wireless scenario with TCP traffic only: server unbalance χ (on the left) and fairness index η (on the right) vs. M_1/M_2, for different values of $p_{\max}$ in RED.

the effect of $p_{\max}$ becomes more negligible. We do not report here the detailed results for the sake of space.

In conclusion, in the wireless scenario the max-scalar policy exhibits a complex behavior reaching an operational point which is middle way between the maximum throughput operational point ($\mu_1^* = \mu_2^*$) and the point corresponding to a fair distribution of bandwidth to connections ($\frac{\mu_1^*-\lambda_1}{\mu_2^*-\lambda_2} = k_1/k_2$). So doing, a reasonable trade off is achieved between the conflicting requirements of optimizing global performance (system throughput) and fairly distributing bandwidths among TCP flows.

6. Validation of the model

As final step we validate the prediction of our model against a detailed simulator at packet level. To this end, we developed in OMNeT++[30] modules representing the systems of queues implementing the *max-scalar* scheduling policy. In simulations, we fed the queues with standard TCP new-Reno sources provided by the INET Framework [19].

First, we consider a simple scenario comprising a 1 Gbit/s work-conserving server managing three queues ($Q = 3$), and fed by 420 TCP flows. These TCP flows are distributed among the three queues as follows: $M_1 = 240$, $M_2 = 120$, $M_3 = 60$. Round-trip times of TCP flows have been randomly chosen according to a uniform distribution with support in the interval $[26, 30]$ ms. A RED AQM mechanism is implemented at the queues ($\min_{th} = 10$, $\max_{th} = 500$, $p_{max} = 0.05$, $z = 10^{-5}$). A comparison between model predictions (labeled TEO) and

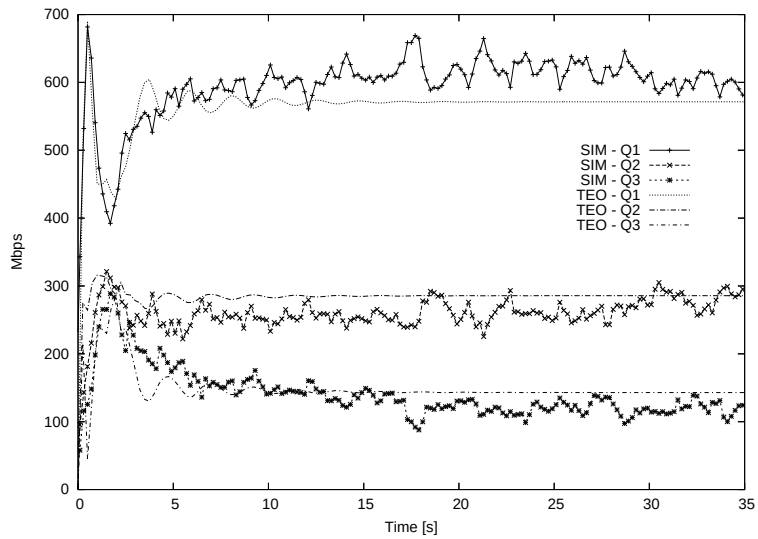

Figure 9: Scenario 1: bandwidth shares among TCP aggregates. Bandwidth obtained by aggregates, averaged over interval $[20, 35]$ s, are 604, 271 and 125 Mbit/s, respectively.

simulation results (labeled SIM) is reported in Fig. 9, where the bandwidth shares evolution for the three traffic aggregates are plotted. Simulation points are obtained by averaging the bandwidth obtained by aggregates within 200 ms windows. The good agreement between model predictions and simulation results confirms that bandwidth shares obtained aggregates are roughly proportional to parameters M_q.

As second scenario, we consider a case in which all server queues are fed by the same number of TCP flows (140); however flows traversing different queues have different round-trip times. Round-trip times of flows traversing queue 1 are uniformly distributed in the interval $[18, 22]$ ms; those traversing queue 2 are distributed in the interval $[36, 44]$ ms; those traversing queue 3 are distributed in the interval $[54, 66]$ ms. The same parameters of the previous scenario were used to tune the RED mechanism. Also in this case, a good agreement between model predictions and simulation is shown in Fig. 10. The server bandwidth is distributed to the aggregates proportionally to the inverse of their round-trip times.

As third scenario we have considered a 4×4 IQ switch with ports running at 1 Gbit/s. The analyzed traffic scenario, already considered in the previous section, is represented by matrices M and R (Sec.5.2) with $m_0 = 28$, $\alpha = 2$, $\beta = 1$ and $r_0 = 28$ ms. The same RED parameters of the previous scenarios were used. The throughput obtained by simulation, averaged over interval $[20, 35]$ s, is reported for

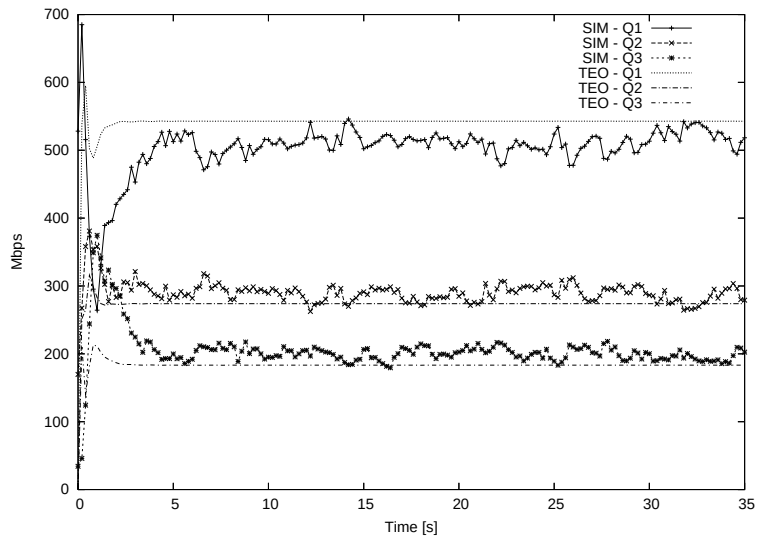

Figure 10: Scenario 2: bandwidth shares among TCP aggregates. Bandwidth obtained by aggregates, averaged over interval $[20, 35]$ s, are 530, 282 and 188 Mbit/s, respectively.

each input-output pair in the following matrix:

$$T_{SIM} = \begin{pmatrix} 73 & 141 & 268 & 518 \\ 524 & 71 & 139 & 266 \\ 270 & 531 & 67 & 132 \\ 133 & 257 & 526 & 84 \end{pmatrix} \qquad \text{Mbit/s}$$

which matches very well with the average throughput estimated by our model:

$$T_{TEO} = \begin{pmatrix} 67 & 133 & 267 & 533 \\ 533 & 67 & 133 & 267 \\ 267 & 533 & 67 & 133 \\ 133 & 267 & 533 & 67 \end{pmatrix} \qquad \text{Mbit/s}$$

These results validate the model and confirm that IQ switch bandwidth is efficiently exploited by the max-scalar policy as predicted by the model; moreover the long-term bandwidth shares are almost exactly proportional to M_q, as expected by the model.

At last, we emphasize that we have validated the model predictions against simulations in several other scenarios, which are not described in details for brevity. In all cases the simulation results have confirmed model predictions.

7. Conclusions

Max-scalar scheduling policies have previously been proposed to optimize the global system performance in several application contexts such as wireless networks, satellite networks and high-capacity router architectures.

Optimality of such scheduling policies was proved, however, only under assumptions of stationarity and admissibility for the traffic flowing through the system of queues. It is unclear how they behave in the case of either non stationary, or rate-adaptive traffic sources, that may induce temporary overloads of some system architectural elements.

In this paper we investigated how max-scalar scheduling policies behave under TCP traffic sources. To this end, we have described the average dynamics of both traffic sources and switch queues through a system of Delay Differential Equations (DDEs), whose properties were throughly analyzed.

Our findings were rather surprising and intriguing; the adoption of max-scalar scheduling policies along with carefully designed AQM schemes permits to efficiently exploit the bandwidth of complex systems such as either IQ switches or wireless stations without negatively affecting the fairness of TCP flows.

We recognize that research on max-scalar scheduling policies has been driven so far mainly by speculative interest, being such policies largely ignored in products implementations. Still, we believe that max-scalar scheduling policies offer interesting potentialities in application contexts where bandwidth over-provisioning has significant costs (like in wireless networks). For these reasons the results of our investigation, shedding some light on important aspects that are often neglected, can stimulate a debate on the implementability of such scheduling policies at nodes.

Complementary to our work, many open questions can be foreseen and are left for further investigation. First, a deeper characterization of the convergence region is needed to understand the range of AQM parameters that are compatible with a stable behavior of the system. Second, the sensitivity analysis of the equilibrium points is needed to design accurately the AQM scheme and meet a suitable fairness level. Third, the adopted methodology can be used to design new AQM schemes, tailored to support regulated traffic interacting with max-scalar scheduling policies.

References

[1] M. Ajmone Marsan, A. Bianco, P. Giaccone, E. Leonardi, F. Neri, "Packet-Mode Scheduling in Input-Queued Cell-Based Switch", *IEEE/ACM Transactions on Networking*, vol. 10, n. 5, Oct. 2002, pp. 666-678

[2] M. Ajmone Marsan, P. Giaccone, E. Leonardi, F. Neri, "On the stability of local scheduling policies in networks of packet switches with input queues",

IEEE Journal on Selected Areas in Communications, vol. 21, n. 4, May 2003, pp. 642-655

[3] M. Ajmone Marsan, M. Franceschinis, P. Giaccone, E. Leonardi, E. Schiattarella, A. Tarello, ”Using Partial Differential Equations to Model TCP Mice and Elephants in Large IP Networks”, *IEEE/ACM Transactions on Networking*, vol. 13, n. 6, Dec. 2005, pp. 1289-1301

[4] M. Andrews, L. Zhang, “Achieving Stability in Networks of Input-Queued Switches”, *IEEE/ACM Transactions on Networking*, vol. 11, n. 5, Oct. 2003, pp. 848-857

[5] G. Appenzeller, I. Keslassy, N. McKeown, “Sizing router buffers”, *ACM Sigcomm’04*, Portland, OR, USA, August 2004

[6] F. Baccelli, D. Hong, “Interaction of TCP Flows as Billiards”, *IEEE/ACM Transactions on Networking*, vol. 13, n. 4, Aug. 2005, pp. 841-853

[7] F. Baccelli, D. Hong, “Flow Level Simulation of Large IP Networks”, *IEEE Infocom 2003*, San Francisco, CA (USA), March 2003

[8] C. Barakat, P. Thiran, G.F. Iannaccone, C. Diot “Modelling Internet Backbone Traffic at Flow Level”, *IEEE Transactions on Signal Processing*, Vol 51, n. 8, Aug. 2003

[9] J. G. Dai, W. Lin, “Maximum Pressure Policies in Stochastic Processing Networks”, *Operations Research*, vol. 53, 2005, pp. 197-218

[10] J.G. Dai, B. Prabhakar, “The throughput of data switches with and without speedup”, *IEEE Infocom 2000*, Tel Aviv, Israel, Mar. 2000, pp. 556-564

[11] A. Eryilmaz, R. Srikant, “Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control”, *IEEE/ACM Transactions on Networking*, vol. 15,n. 6, Dec. 2007, pp. 1333-1344

[12] S. Floyd, V. Jacobson, “Random Early Detection Gateways for Congestion Avoidance”, *IEEE/ACM Transactions on Networking*, vol. 1, n. 4, Aug. 1993, pp. 397-413

[13] A. Goldsmith, S.A. Jafar, N. Jindal, S. Vishwanath, “Capacity Limits of MIMO Channels”, *IEEE JSAC*, vol. 21, n. 5, June 2003

[14] GNU Scientific Library, `http://www.gnu.org/s/gsl/`

[15] Y. Gu, Y. Liu, F. Lo Presti, V. Misra, D. Towsley, "Fluid Models and Solutions for Large-Scale IP Networks", *ACM Sigmetrics 03*, San Diego, CA (USA), June 2003

[16] Y. Gu, Y. Liu, D. Towsley, "On Integrating Fluid Models with Packet Simulation", *IEEE Infocom 04*, Hong Kong, China, March 2004

[17] C.V. Hollot, Y. Liu, V. Misra, D. Towsley, "Unresponsive flows and AQM performance," *IEEE Infocom 2003*, San Francisco, CA (USA), March 2003

[18] C.V. Hollot, V. Misra, D. Towsley, W.B. Gong, "Analysis and design of controllers for AQM routers supporting TCP flows", *IEEE Transactions on Automatic Control*, vol. 47, n. 6, Jun. 2002, pp.945-959

[19] INET Framework, `http://inet.omnetpp.org/`

[20] E. Leonardi, M. Mellia, M. Ajmone Marsan, F. Neri, "On the Throughput Achievable by Isolated and Interconnected Input-Queueing Switches under Multiclass Traffic", *IEEE Transactions on Information Theory*, Vol. 51, No. 3, Mar. 2005, pp.1167-1174

[21] X. Lin, N.B. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Wireless Networks", *IEEE/ACM Transactions on Networking*, vol. 14, n. 2, Apr. 2006, pp. 302-315

[22] Maple, `www.maplesoft.com`

[23] N. McKeown, A. Mekkittikul, V. Anantharam, J. Walrand, "Achieving 100% throughput in an input-queued switch", *IEEE Transactions on Communications*, vol. 47, n. 8, Aug. 1999, pp. 1260-1272

[24] N. McKeown, D. Wischik, et al., "Making router buffers much smaller (Parts I , II and III)", *ACM SIGCOMM Computer Communication Review*, vol. 35, n. 3, July 2005, pp.73-89

[25] S. Meyn, *Control Techniques for Complex Networks*, Cambridge University Press, 2007

[26] S. Misra, W.B. Gong, D. Towsley, "Fluid-Based Analysis of a Network of AQM Routers Supporting TCP Flows with an Application to RED", *ACM Sigcomm'00*, Stockholm, Sweden, August 2000

[27] M.J. Neely, E. Modiano, C.E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks", *IEEE JSAC*, vol. 23, n. 1, Jan. 2005, pp. 89-103

[28] M.J. Neely, E. Modiano, C.E. Rohrs, "Power Allocation and Routing in Multibeam Satellites with Time-Varying Channels", *IEEE/ACM Transactions on Networking*, vol. 11, n. 1, Feb. 2003, pp. 138-152

[29] M.J. Neely, E. Modiano, Li Chih-Ping, "Fairness and Optimal Stochastic Control for Heterogeneous Networks", *IEEE/ACM Transactions on Networking*, vol. 16, n. 2, Apr. 2008, pp.396-409

[30] "OMNeT ++ Discrete Event Simulation System", available at `http://www.omnetpp.org`

[31] R. Pan, B. Prabhakar, K. Psounis, D. Wischik, "SHRiNK: a method for enabling scaleable performance prediction and efficient network simulation", *IEEE/ACM Transactions on Networking*, vol. 13, n. 5, Oct. 2005, pp. 975-988

[32] K. Ross, N. Bambos, "Local Search Scheduling Algorithms for Maximal Throughput in Packet Switches", *IEEE Infocom 2004*, Hong Kong, China, Mar. 2004

[33] A. Shpiner, I. Keslassy, "Modeling the interactions of congestion control and switch scheduling", *Computer Networks*, vol. 55, n. 6, pp. 1257-1275, 2011

[34] L. Tassiulas, A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks", *IEEE Transactions on Automatic Control*, vol. 37, n. 12, Dec. 1992, pp. 1936-1948

[35] L. Tassiulas, "Scheduling and performance limits of networks with constantly changing topology", *IEEE Transactions on Information Theory*, vol. 43, n. 3, May 1997, pp. 1067-1073

[36] L. Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switches", *IEEE Infocom 1998*, San Francisco, CA, USA, Apr. 1998

[37] A. Todini, A. Baiocchi, D. Venturi, "Inadequacy of the Queue-Based Max-Weight Optimal Scheduler on Wireless Links with TCP Sources", *IEEE ICC 2009*, Dresden, Germany, Jun. 2009

[38] J. G. Dai "On Positive Harris Recurrence of Multiclass Queueing Networks: A Unified Approach Via Fluid Limit Models". *Annals of applied Probability*, Vol. 5, n.1, 1995.

Appendix A. A brief survey on fluid limits

Fluid limits methodology has been originally introduced by Dai [38] to establish the stability of cyclic networks of queues. Fluid limits have been extended and successfully applied for the characterization of throughput/delay properties of computer networks [6, 7, 9, 10].

The classical fluid limit approach applies to queuing systems with *infinite* buffer capacity subject to unregulated sources. Consider a system of infinite capacity queues whose evolution is driven by:

$$x_q(t) = x_q(\tau) + a_q(\tau, t) - \mu_q(\tau, t) \qquad \text{with } \tau < t$$

where $a_q(\tau, t)$ represents the number of arrivals at queue q in $[\tau, t)$, and $\mu_q(\tau, t)$ represents the number of customers served at queue q in $[\tau, t)$. The corresponding fluid limits are obtained by applying a fluid scaling to queue dynamics, i.e. by considering for any[5]. $r \in \mathbb{N}$, $x_q^r(t) = x_q(rt)/r$ and then taking the limit for large r:

$$\bar{x}_q(t) = \lim_{r \to \infty} x_q^r(t)$$

where the limit operator must be interpreted as a uniform limit over compact intervals for a properly defined subsequence of functions $x_q^{r_n}(t)$ (according to a weak Skorohod topology). Observe that if we assume that arrivals satisfy the strong law of large number, i.e., $\lim_{t \to \infty} a_q(\tau, t)/(t - \tau) = \lambda_q$ w.p.1 for some $\lambda_q \geq 0$ and every τ, then $\bar{x}_q(t)$ satisfies:

$$\bar{x}_q(t) = \bar{x}_q(\tau) + (t - \tau)\lambda_q - \bar{\mu}(\tau, t)$$

where $\bar{\mu}(\tau, t) = \lim_{r \to \infty} \mu(r\tau, rt)/r$. The importance of fluid limits resides in the fact that the study of the fluid limiting trajectories $\bar{x}_q(t)$ allows to gather insights on the stability of the original system of queues; in particular, if $\bar{x}_q(t) = 0$ for any $t > t_0$, the original queue is (rate) stable, otherwise it is unstable. We remark that that previous technique can be applied to both discrete time and continuous time queuing systems. In particular, [10] has derived the fluid limits equations for IQ switches under max-weight scheduling policies.

When we considering TC/IP networks, the previous approach cannot be directly applied because of the finite storage queuing systems. However fluidification can be still successfully applied reinterpreting the meaning of the fluid scaling operator according to the guidelines proposed in [6, 7].

[5] In this section, r is used as a mute variable and it is not related to r_q appearing in the previous sections

First, we define a sequence of scaled systems. The rth system is obtained by the original one by scaling up the number of TCP sources $M_q^r = rM_q$ and the transmission speed of queues $\mu_q^r = r\mu_q(t)$. By doing so, the per-flow share of the capacity remains constant with respect to r. AQM profiles are scaled as well in the rth system, according to $f^r(x) = f(rx)$. Finally, we define a rescaled version of the queue size $x_q^r(t) = x_q(t)/r$ and we analyze its behavior as r grows large: $\bar{x}_q(t) = \lim_{r\to\infty} x_q^r(t)$. Following the same approach in [6, 7], it can be shown that $\bar{x}_q(t)$ satisfies the fluid equations described in Sec. 3.1. Observe that fluid trajectories, in our case, describe the limiting dynamics of very high speed TCP/IP systems, since they are obtained scaling up the datarate of the bottleneck along with the number of TCP flows.

Appendix B. Proof of Theorem 1

The stationary equilibrium is independent from the temporal delays within the system of different equations. Hence, we neglect the temporal delays.

By combining (4) and (8), the equation describing the TCP dynamics becomes:

$$\frac{dw_q(t)}{dt} = \frac{1}{r_q} - \frac{w_q^2(t)}{2r_q} f(x_q(t)) \tag{B.1}$$

The equilibrium point, $(X^*, \hat{X}^*, W^*)$, by definition, satisfies the set of algebraic equations obtained by: (5), (7), (9), (B.1), setting to zero the time derivatives. As a consequence, $(X^*, \hat{X}^*, W^*)$ is an equilibrium point of the dynamical system iff it satisfies:

$$\frac{1}{r_q} = \frac{(w_q^*)^2}{2r_q} f(x_q^*) \tag{B.2}$$

$$(k_q w_q^* + \lambda_q)(1 - f_d(x_q^*)) = \mu_q^* \tag{B.3}$$

$$\hat{x}_q^* = x_q^* \tag{B.4}$$

$$\mu^* = \arg\max_{\alpha\in\mathcal{S}} \sum \alpha_q x_q^* \tag{B.5}$$

Note that, at the equilibrium point, for every q, it must be $f(x_q^*) > 0$ and $w_q^* > 0$, otherwise (B.2) cannot be satisfied; as a consequence, necessarily from assumption (iv) of the theorem, it results $x_q^* > 0$ and thus (B.3) (derived from (5)) is the only relevant case describing the fluid window evolution.

Observe that (B.5) can be rewritten in the following way, choosing $\mu^* \in \mathcal{S}$:

$$\sum_q (\mu_q^* - \eta_q) x_q^* \geq 0 \qquad \forall \eta \in \mathcal{S} \tag{B.6}$$

Now (B.3) and (B.2) allow to algebraically relate μ_q^* and x_q^*, indeed from (B.2):

$$w_q^* = \sqrt{\frac{2}{f(x_q^*)}}$$

and thus substituting in (B.3) we obtain:

$$\left(\frac{k_q\sqrt{2}}{\sqrt{f(x_q^*)}} + \lambda_q\right)\left(1 - f_d(x_q^*)\right) = \mu_q^* \tag{B.7}$$

The function on the left hand side is strictly decreasing with respect to its argument, $x_q^* \geq 0$, moreover for $x_q^* \to 0$ it tends to $+\infty$, while for $x_q^* = B_q$, by assumption (v), it is null. As a consequence for every value of μ_q^*, (B.7) always admits one and only one solution in x_q^*. We denote with $h_q(\mu_q^*)$ this solution. By construction $x_q^* = h_q(\mu_q^*)$, and $0 < h_q(\mu_q^*) \leq B_q$. At last, $h_q(\mu_q^*)$ is strictly decreasing with its argument. Considering again (B.6), μ^* satisfies:

$$\sum_q (\mu_q^* - \eta_q) h_q(\mu_q^*) \geq 0 \qquad \forall \eta \in \mathcal{S}$$

Define:

$$G(\mu) = \sum_q \int_0^{\mu_q} h_q(x)dx$$

where $G(\mu)$ by construction is a strictly concave function as it can be easily verified by computing its Hessian (we recall that $h_q()$ is a strictly decreasing function, and from assumption (iii) it is differentiable). Thus condition (B.6) can be rewritten as:

$$(\eta - \mu^*)\nabla G(\mu^*) \leq 0 \qquad \forall \eta \in \mathcal{S} \tag{B.8}$$

We conclude our proof, invoking the following lemma.

Lemma 1. *In order to satisfy condition (B.8) μ^* must be the only solution of the following optimization problem:*

$$\mu^* = \arg\max_{\mu \in \mathcal{S}} G(\mu) \tag{B.9}$$

First, observe that, since $G(\mu)$ is strictly concave and $\mathcal{S}$ is compact and convex, (B.9) admits one and only one solution.

Second, we prove that the satisfaction of condition (B.9) provides a necessary condition for the satisfaction of (B.8). Indeed, assume that μ^* does not satisfy (B.9), then there exists $\eta \in \mathcal{S}$ such that:

$$G(\eta) > G(\mu^*)$$

However, due to the concavity of $G(\mu)$:

$$G(\eta) \leq G(\mu^*) + (\eta - \mu^*)\nabla G(\mu^*)$$

from which $(\eta - \mu^*)\nabla G(\mu^*) > 0$, which contradicts (B.8).

Finally, the fact that the solution of (B.9) also satisfies (B.8) is an immediate consequence of the Karush Kuhn-Tucker conditions.

Appendix C. Stability of the equilibrium point: work conserving server

In this appendix we prove that $\frac{d\mathcal{L}(X(t),W(t))}{dt} < 0$ almost for every $t > 0$ around the equilibrium point. To simplify the calculations we suppose that the AQM scheme adopts a pure marking policy at the equilibrium. However the arguments reported in this proof can be extended in a straightforward way to the more general case.

Since $x_q(t)$ and $w_q(t)$ are by definition absolutely continuous functions, $\mathcal{L}(X(t), W(t)) = \max_q (x_q(t) - x^*)^2 + \beta \sum_q M_q (w_q(t) - w^*)^2$ is an absolutely continuous function, and thus it is differentiable almost for every $t > 0$. In the following we will show that whenever $\frac{d\mathcal{L}(X(t),W(t))}{dt}$ exists, it is negative.

We start by referring the results from Section 5.1; at the equilibrium point every queue has the same length, and every source have the same average window size, i.e., $\forall q$, $x_q^* = x^*$ and $w_q^* = w^*$. From (14) and (15), where $f_d(x^*) < 1$ and $w^* > 0$, it holds $\mu_q^* > 0$ for every q. We denote with $\mu_{\min}$ the minimum achieved rate at the equilibrium: $\mu_{\min} = \min_q \mu_q^*$. Now fix time t and suppose $\max_q \{|x_q(t) - x^*|, |w_q(t) - w^*|\} < \delta$ for some small $\delta > 0$. Let $\mathcal{Q}_0(t)$ be the set of queues which are, at time t, at maximum distance from the equilibrium queue sizes, i.e., $\mathcal{Q}_0(t) = \arg\max_q (x_q(t) - x^*)^2$.

If $|\mathcal{Q}_0(t)| = 1$, no problems of differentiability for $\mathcal{L}(X(t), W(t))$ arise, however if $|\mathcal{Q}_0(t)| > 1$, $\mathcal{L}(X(t), W(t))$ is not guaranteed to be differentiable at t. A sufficient and necessary condition for differentiability is that $\mathcal{Q}_0(t) = \mathcal{Q}_0(\tau)$ whenever $|\tau - t|$ is sufficiently small. This implies that all the queues with same length must have the same derivative, i.e., $\dfrac{dx_q(t)}{dt} = \dfrac{dx_{q'}(t)}{dt}$ for every q and q' belonging to $\mathcal{Q}_0(t)$ having the same length (moreover queues with different lengths must have the same derivative when taken in absolute value).

Let us now partition $\mathcal{Q}_0(t)$ into two subsets, $\mathcal{Q}_0^+(t)$ and $\mathcal{Q}_0^-(t)$, such that: $x_q(t) \geq x^*$ for any $q \in \mathcal{Q}_0^+(t)$ and $x_q(t) < x^*$ for any $q \in \mathcal{Q}_0^-(t)$. The services at those queues satisfy:

$$\sum_{q \in \mathcal{Q}_0^+(t)} \mu_q(t) = \hat{\mu} \qquad \text{and} \qquad \sum_{q \in \mathcal{Q}_0^-(t)} \mu_q(t) = 0$$

where $\hat{\mu}$ is the server capacity.
Case 1. First we assume that $|\mathcal{Q}_0^+(t)| < Q$ and $|\mathcal{Q}_0^-(t)| < Q$; then:

$$\sum_{q\in\mathcal{Q}_0^+(t)} \frac{dx_q(t)}{dt} = \sum_{q\in\mathcal{Q}_0^+(t)} \Big[k_q w_q(t) + \lambda_q\Big](1 - f(x_q(t))) - \hat{\mu}$$
$$\sum_{q\in\mathcal{Q}_0^-(t)} \frac{dx_q(t)}{dt} = \sum_{q\in\mathcal{Q}_0^-(t)} \Big[k_q w_q(t) + \lambda_q\Big](1 - f(x_q(t)))$$

As discussed above, all queues in $Q_0(t)$ must have the same derivative. Hence, after some calculations we obtain:

$$q \in \mathcal{Q}_0^+(t) \to |\mathcal{Q}_0^+(t)|\frac{dx_q(t)}{dt} < \sum_{q\in\mathcal{Q}_0^+(t)} (\delta k_q + \mu_q^*) - \hat{\mu}$$
$$q \in \mathcal{Q}_0^-(t) \to |\mathcal{Q}_0^-(t)|\frac{dx_q(t)}{dt} > \sum_{q\in\mathcal{Q}_0^-(t)} (-\delta k_q + \mu_q^*)$$

If we now define $\hat{k} = \max_q k_q$, and observe that

$$\hat{\mu} - \sum_{q\in\mathcal{Q}_0^+(t)} \mu_q^* \geq \mu_{\min} \text{ and } \hat{\mu} - \sum_{q\in\mathcal{Q}_0^-(t)} \mu_q^* \geq \mu_{\min}$$

we obtain:

$$q \in \mathcal{Q}_0^+(t) \to \frac{dx_q(t)}{dt} < \delta\hat{k} - \frac{\mu_{\min}}{|\mathcal{Q}_0^+(t)|}$$
$$q \in \mathcal{Q}_0^-(t) \to \frac{dx_q(t)}{dt} > -\delta\hat{k} + \frac{\mu_{\min}}{|\mathcal{Q}_0^-(t)|}$$

For enough small δ,

$$q \in \mathcal{Q}_0^+(t) \to \frac{dx_q(t)}{dt} < -\frac{\mu_{\min}}{2|\mathcal{Q}_0^+(t)|} < -\frac{\mu_{\min}}{2Q}$$
$$q \in \mathcal{Q}_0^-(t) \to \frac{dx_q(t)}{dt} > \frac{\mu_{\min}}{2|\mathcal{Q}_0^-(t)|} > \frac{\mu_{\min}}{2Q}$$

and we can claim:

$$\frac{d\max_{q'}(x_{q'}(t) - x^*)^2}{dt} = 2\frac{dx_q(t)}{dt}(x_q(t) - x^*) < -\frac{\mu_{\min}}{Q}|x_q(t) - x^*| \quad \text{(C.1)}$$

Now consider source q; for small δ we can exploit a first order Taylor's expansion of $w_q(t)$ around (w^*, x^*); from (B.1):

$$\frac{dw_q(t)}{dt} = \frac{-2w^* f(x^*)(w_q(t) - w^*) - f'(x^*)(w^*)^2(x_q(t) - x^*)}{2r_q} \tag{C.2}$$

and we can claim:

$$\frac{1}{2}\frac{d(w_q(t) - w^*)^2}{dt} = \frac{dw_q(t)}{dt}(w_q(t) - w^*) = \\ -\frac{2w^* f(x^*)(w_q(t) - w^*)^2}{2r_q} - \frac{f'(x^*)(w^*)^2}{2r_q}(w_q(t) - w^*)(x_q(t) - x^*) + o(\delta^2) \tag{C.3}$$

where $f'(x^*) = \left.\frac{df(x)}{dx}\right|_{x=x^*}$. Combining (C.1) and (C.3), we obtain:

$$\frac{d\mathcal{L}(X(t), W(t))}{dt} < -\frac{\mu_{\min}}{Q}|x_q(t) - x^*| - 2\beta w^* f(x^*)\sum_q \left[k_q(w_q(t) - w^*)^2\right] - \\ \beta f'(x^*)(w^*)^2 \sum_q \left[k_q(w_q(t) - w^*)(x_q(t) - x^*)\right] + o(\delta^2)$$

Note that the first two terms are negative, while the latter has an indefinite sign. However the whole is negative if β is chosen such that:

$$\delta f'(x^*)(w^*)^2 \sum_q k_q < \frac{\mu_{\min}}{\beta Q}$$

Case 2. Now we consider the case $|\mathcal{Q}_0^+(t)| = Q$ or $|\mathcal{Q}_0^-(t)| = Q$ i.e., all the queues at time t are of the same length $x_q(t) = x(t)$. Repeating similar considerations as before, we obtain:

$$\frac{dx_q(t)}{dt}(x(t) - x^*) \leq \frac{1}{Q}(1 - f(x(t)))(x(t) - x^*)\sum_q \left[k_q(w_q(t) - w^*)\right]$$

From which, exploiting a first order Taylor's expansion of $f(x)$ around x^* and then

(C.2), it follows that:

$$\frac{d\mathcal{L}(X(t),W(t))}{dt} = \frac{2}{Q}\sum_q \Big[k_q(w_q(t)-w^*)\Big](x(t)-x^*)(1-f(x^*)) - \frac{2}{Q}f'(x^*)(x(t)-x^*)^2\sum_q k_q - 2\beta w^* f(x^*)\sum_q \Big[k_q(w_q(t)-w^*)^2\Big] - \beta f'(x^*)(w^*)^2\Big[\sum_q k_q(w_q(t)-w^*)\Big](x(t)-x^*) + o(\delta^2)$$

can be made always negative choosing

$$\beta > \frac{2(1-f(x^*))}{Qf'(x^*)(w^*)^2}$$

Appendix D. Stability of the equilibrium point: wireless station

Denoting with $\Delta w_q(t) = w_q(t) - w_q^*$ and $\Delta x_q(t) = x_q(t) - x_q^*$, the linearized dynamical system of equations comprises $2Q$ equations. The first set of Q equations are obtained linearizing (B.1):

$$\frac{d\Delta w_q(t)}{dt} = -\frac{2w_q^* f(x_q^*)\Delta w_q}{2r_q} - \frac{(w_q^*)^2 f'(x_q^*)\Delta x_q}{2r_q}$$

The other Q equations are obtained linearizing the queue dynamics equations:

$$\frac{d\Delta x_q(t)}{dt} = k_q(1-f_d(x_q^*))\Delta w_q - (k_q w_q^* + \lambda_q){f_d}'(x_q^*)\Delta x_q - \frac{1}{2\|X^*\|_2}\Delta x_q + \frac{1}{2\|X^*\|_2}\sum_{q'\neq q}\Delta x_{q'}$$

where $f'(x_q^*) = \left.\frac{df(x)}{dx}\right|_{x=x_q^*}$ and $f_d'(x_q^*) = \left.\frac{df_d(x)}{dx}\right|_{x=x_q^*}$.

The asymptotic stability of the equilibrium point $(0,0)$ in the above linearized system of equations can be studied using standard techniques. The first step consists in rewriting the above system of equations in its standard form:

$$\frac{dY(t)}{dt} = AY(t)$$

where $Y(t)$ is the vector state

$$(\Delta w_1(t), \Delta x_1(t), \Delta w_2(t), \Delta x_2(t) \ldots \Delta w_Q(t), \Delta x_Q(t))$$

A sufficient and necessary condition for $(0, 0)$ to be asymptotically stable is that matrix A is stable, i.e., all eigenvalues of A have real part strictly negative.

We have verified the stability of matrix A computing the characteristic polynomial of A and applying the standard Routh-Hurtwits criterion; we were able to evaluate the characteristic polynomial symbolically through Maple [22] but we do not report the details for brevity.

Authors' biographies

Paolo Giaccone received the Dr.Ing. and Ph.D. degrees in telecommunications engineering from the Politecnico di Torino, Torino, Italy, in 1998 and 2001, respectively. He is currently an Assistant Professor in the Department of Electronics, Politecnico di Torino. During the summer of 1998, he was with the High Speed Networks Research Group, Lucent Technology-Bell Labs, Holmdel, NJ. During 2000-2001 and in 2002 he was with the Information Systems Networking Lab, Electrical Engineering Dept., Stanford University, Stanford, CA. His main area of interest is the design of network algorithms, the theory of interconnection networks, and the performance evaluation of telecommunication networks through simulations and analytical methods.

Emilio Leonardi is currently an Associate Professor at the Dipartimento di Elettronica of Politecnico di Torino. He received a Dr.Ing. degree in Electronics Engineering in 1991 and a Ph.D. in Telecommunications Engineering in 1995 both from Politecnico di Torino. He participated in several national and European projects. He has been scientific coordinator of the European 7-th FP STREP project NAPA-WINE on P2P streaming applications. He participated to the program committees of several conferences including: IEEE Infocom, and ACM MobiHoc. He was guest editor of two special issues of IEEE Journal of Selected Areas of Communications focused on high speed switches and routers. He is currently associate editor of IEEE Transactions on Parallel and Distributed Systems. His research interests are in the field of: performance evaluation of wireless networks, P2P systems, caching systems. queueing theory.

Fabio Neri received the Dr.Ing. and Ph.D. degrees in electrical engineering from the Politecnico di Torino, Torino, Italy, in 1981 and 1987, respectively. He was a Full Professor in the Department of Electronics, Politecnico di Torino. He passed away, unexpectedly in 2011. His research interests were in the fields of performance evaluation of communication networks, high-speed and all-optical

networks, packet-switching architectures, discrete event simulation, and queueing theory.