

# On the behavior of optimal scheduling algorithms under TCP sources

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

**Abstract**—We focus on the dynamical interaction between regulated Additive Increase Multiplicative Decrease (AIMD) traffic sources and the *max-scalar* scheduling policy, which has been proved to be optimal in terms of throughput in case of stationary unregulated traffic sources.

We describe the average dynamics of both sources and queues dynamics through a system of Ordinary Differential Equations (ODEs), which is numerically solved.

## I. INTRODUCTION AND PREVIOUS WORK

<sup>1</sup>In recent years a significant effort has been devoted by the 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], [3], [5], [9], [11], [13], [14], [16], [17]. This problem dates back to the early '90, when Tassiulas and Ephremides, in their seminal work [19], have first shown that the maximization of the throughput in a *networks 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 current queues state.

It is worth noticing that the scheduling policy proposed in [19], the so-called *max-scalar policy*, and its later extensions [1], [5], [9], [11], [13], [14], [17], do not require any a priori knowledge of the long-term traffic, thereby appearing amenable for implementation in contexts in which traffic is highly dynamic and unpredictable.

Optimality of the *max-scalar policy* and its extension 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 *max-scalar 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 very recently the attention has been turned to the analysis of the interaction between optimal dynamical 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 Transport Control Protocol (TCP) and dynamically adapt the sending rate to the estimated traffic congestion level according to a Additive Increase Multiplicative Decrease (AIMD) scheme.

<sup>1</sup>This work was funded by PRIMO FIRB of the Italian Ministry for Education, Research and University.

In this paper, we describe the average dynamics of both sources and queues dynamics through a system of Ordinary Differential Equations (ODEs), which is numerically solved.

In [6], [15] the behavior of max-scalar policies under regulated sources has been analyzed. However, the rate adaptation algorithms considered in [6], [15] significantly differ from the AIMD source behavior considered here, 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 algorithm driven only by losses and delay information.

## II. SYSTEMS OF INTERACTING QUEUES

We consider a systems of  $Q$  discrete time, interacting queues; this provides an abstract model for several different communication scenarios such as the system of transmission queues at a wireless access point, or the virtual output queueing (VOQ) system in a input queued (IQ) switch.

We assume packets to be of fixed size. Queues evolve as follows:

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

where  $x_q(n)$  represents the queue length,  $a_q(n)$  represents the number of packets arriving at the queue,  $\mu_q(n)$  represents the amount of service received by queue  $q$  during time  $(n, n+1]$  and  $[x]^+$  represents  $\max(0, x)$ . The set of achievable queues service rates is subject to a set of physical constraints. 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  $\mathcal{S} \subset \mathbf{R}_+^Q$ , for every  $n$ .

Just as matter of example we consider three possible scenarios:

- **Work conserving queue:** in such simple case, the sum of service rates is bounded by the transmission capacity  $C$  of the queue:  $\mathcal{S} = \{\mu : \mu_q \geq 0 \text{ and } \sum_q \mu_q \leq C\}$ .
- **IQ switch:** the system of VOQs in a  $P \times P$  input queued switch comprises  $Q = P^2$  interacting queues. In this case, the sum of services provided to all the virtual output queues either residing at the same input card or directed to the same output card is limited to one packet per slot. Let us denote with  $IQ(i)$ , ( $1 \leq i \leq P$ ), the set of VOQs at input  $i$  and with  $OQ(j)$ , ( $1 \leq j \leq P$ ) the set of queues directed to output  $j$ . It results that  $\mathcal{S}$  is defined as the set

of  $\mu$  which satisfy the following constraints:

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

- **Wireless scenario:** in this case  $\mathcal{S}$  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 transmission rate  $\mu_q$  depends only on the power  $P_q$  used for transmitting information from  $q$  [14]. 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,  $\mathcal{S}$  is defined as the convex region defined by all the vectors  $\mu = (\mu_1(P_1), \dots, \mu_q(P_q), \dots, \mu_N(P_N))$  being  $\sum_q P_q \leq P_{tot}$ .

The geometry of region  $\mathcal{S}$  depends on the specification of functions  $\mu_q(P_q)$  which in their turn depend on the physical layer specification [14]. In this paper, just as matter of example, we assume  $\mathcal{S}$  to be a circular region, defined by following constraints:

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

being  $C$  a positive constant.

#### A. Optimal Policy under unregulated traffic sources

Under unregulated traffic sources, the problem of the definition of the optimal scheduling policy and the associated throughput region in complex systems of interacting queues under dynamic scheduling policies, has attracted significant attention in the last decade from the research community since the pioneering work [19]. 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 achieve maximum throughput if max-scalar scheduling,  $\mathcal{P}_{MS}$ , policy is applied. According to  $\mathcal{P}_{MS}$ , at each time slot  $n$ , the service vector is selected as follows:

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

The result in [19] has been generalized and adapted to different application contexts in the last years. As matter of example we just briefly recall some of the related works. In the switching context, several studies have been aimed at the definition of the stability region in Input-Queued (IQ) switching architectures built around a buffer-less crossbar: papers [1], [9], [11], [16], [17] 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 [5], [10]; in [2], [3], [9], finally, the problem of the definition of the stability region in networks of IQ switches has been considered. In the context of the satellite and wireless networks, generalizations of  $\mathcal{P}_{MS}$  have been recently proposed and shown to achieve the maximum

throughput in [13], [14], [18]. Finally, recently [4] generalizes the result in [19] under more general exogenous arrival processes applying a different analytical technique called fluid models.

All the previous works, however, have considered unregulated stationary traffic sources.

#### B. Interaction with Regulated sources

Only very recently [6], [15], the attention has been turned to the analysis of the interaction between optimal dynamical policies and adaptive traffic sources. Papers [6], [15] have shown that  $\mathcal{P}_{MS}$  well behaves in presence of regulated sources, so guaranteeing an acceptable degree of fairness to flows also in case of temporary overload. However, these papers have focused on congestion control mechanisms significantly different from TCP, which require that the traffic sources to strictly interact with the network and gather detailed and updated information about the queues status.

Let  $\rho_q(n)$  be the aggregate arrival rate at queue  $n$ , equal to the overall sending rate at the corresponding sources. In [6] sources were assumed to adjust  $\rho_q(n)$  on the base of the instantaneous queue size  $x_q(n)$ :

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

being  $\alpha_q$ ,  $K$  and  $\gamma_q$  opportune positive constants. Similarly in [15] the rate of sources must be dynamically adapted on the basis of the instantaneous queues lengths. According to one of the proposals in [15]:

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

being  $V$  a control parameter and  $\rho_{max}$  the maximum allowed source sending rate.

### III. MODEL FOR AIMD SOURCES

All the above congestion control schemes significantly differ from TCP/IP, running in the Internet. Aim of our work is to study the behavior of TCP-based congestion control mechanism in network 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 source and queue dynamics, we adopt a continuous time fluid approach [12] in which both sources and queues average dynamics are described by deterministic ordinary differential equations.

For simplicity, we assume that all  $M_q$  TCP sources experience the same round trip time  $r_q$ . We further neglect the effects of queueing delays on round trip time, identifying  $r_q$  with only the propagation delay component along the path. This assumption appears justified by the consideration that in modern large bandwidth networks often propagation delay represents the dominant component of the round trip time.

We suppose that each queue  $q$  implements an active queue management mechanism, like RED [7], according to which the packet loss probability  $p_q(t)$  is related to buffer level  $x_q(t)$  according to the relation  $p_q(t) = f(x_q(t))$ , being  $f(x)$  a

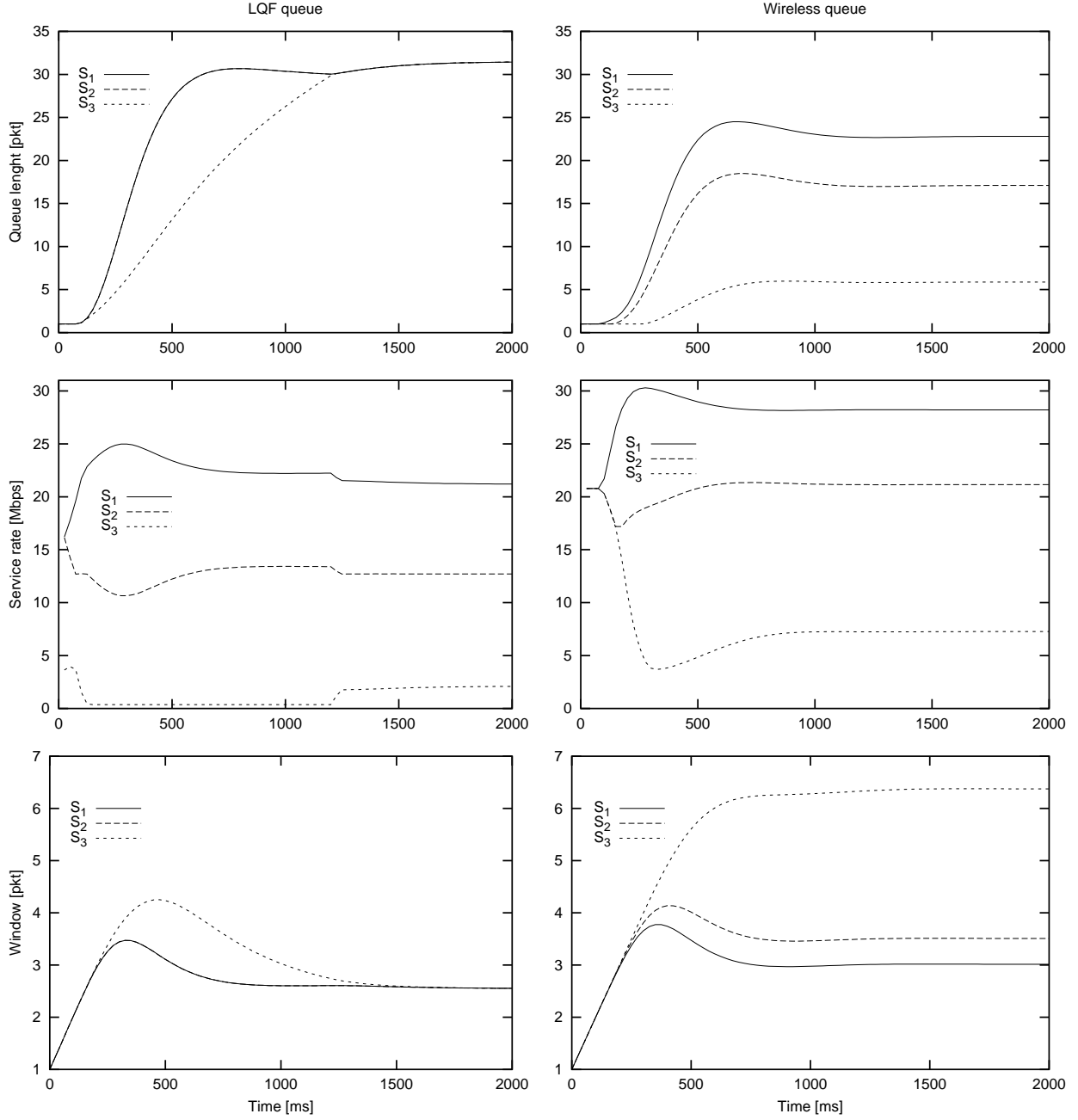


Fig. 1. Performance observed for LQF and wireless server, in the case of three classes of sources  $S_1$ ,  $S_2$  and  $S_3$

continuous strictly increasing function. Recall that  $\mu_q(t)$  is the service rate of queue  $q$  at time  $t$ . Let  $w_q(t)$  represent the average window of the TCP sources feeding  $q$ .

To simplify the analysis, we consider only long-lived sources whose behavior can be modeled, approximatively, as an ideal AIMD source. We neglect the effects of both time-outs and slow starts phases. Note that short-lived flows can be successful assimilated to unregulated flows [8]. The fluid evolution of the average window size  $w_q(t)$  is driven by the

classical well known AIMD fluid equation:

$$\frac{dw_q(t)}{dt} = \frac{1}{r_q} - \frac{w_q(t)w_q(t - r_q)}{2r_q} p_q(t - (r_q - \tau_q))$$

where  $\tau_q$  is the propagation delay between sources and queue  $q$ . The fluid evolution of queue lengths  $x_q(t)$  is driven by the following equation:

$$\frac{dx_q(t)}{dt} = \frac{M_q}{r_q} w_q(t - \tau_q) (1 - p_q(t)) - \mu_q(t) 1_{\{x_q(t) > 0\}}$$

where the first term on the rate represents the average aggre-

gate arrival rate at queue  $q$ ; indeed,  $M_q w_q / r_q$  represents the overall average sending rate of the  $M_q$  sources.

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

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

being  $\mathcal{S}$  the set of feasible service vectors. We suppose  $\mathcal{S}$  to be a compact convex sub-set of  $\mathbb{R}_+^Q$ . Note that  $\mu(t)$  is the vector which maximizes the scalar product with the vector  $X(t) = (x_1(t), \dots, x_q(t))$ , inside  $\mathcal{S}$ . Hence, from a geometric point of view,  $\mu(t)$  is the point (or is selected among the points) where the perpendicular to  $\mu(t)$  is tangent to  $\mathcal{S}$ .

If we now apply the policy to the previous three scenarios:

- Work conserving queue:  $\mathcal{P}_{MS}$  selects simply the queue with the largest queue size. This policy is usually referred as Longest Queue First (LQF) policy.
- Wireless queue: it can be easily shown that  $\mathcal{P}_{MS}$  provides service according to:

$$\mu_q(t) = \frac{x_q(t)}{\|X(t)\|_2} C$$

- IQ switch:  $\mathcal{P}_{MS}$  selects the queues corresponding to the maximum weight matching on the bipartite graph built as follows: an edge connects node  $i$  to node  $j$  if VOQ  $q$  present at input  $i$  and directed to output  $j$  stores at least one packet; the weight of this edge is set equal to  $x_q$ .

#### IV. NUMERICAL RESULTS

Consider the case in which a single server provides a service rate equal to 3.6 Mbps. We assume that each packet is 1500 bytes long. Three classes of sources are present:  $S_1, S_2, S_3$  with  $M_1 = 10$ ,  $M_2 = 6$  and  $M_3 = 1$ . All the queues implement the same AQM scheme with loss function:  $f(x) = x/X_{max}$ , where  $X_{max}$  is the maximum storage capacity. We set  $X_{max} = 100$  packets and  $r_1 = r_2 = r_3 = 100$  ms. For simplicity, we show the results only for LQF and wireless systems. Figure 1 compares the transient behavior observed in both systems for each class of sources; the graphs are obtained by evaluating numerically the fluid equations describing the system dynamics.

LQF tends to equalize the queue lengths and, indeed, all the queue lengths converge to the same value. The initial transient phase on the queue length is different for  $S_3$  since just one source is not able to feed its queue enough to follow the growth of the other queues; indeed, the service rate received by  $S_3$  is very small. Since  $S_1$  and  $S_2$  experience larger loss rate, they slow down and allow  $S_3$  to reach their queue length. Hence, at the end of the transient period, all the queue lengths and window sizes are equal.

For the wireless queue, the final queue sizes are different for each class of sources. The service rates, by the policy definition, are proportional to the queue size. The final effect is that larger windows are observed for classes with larger number of sources.

Finally, both systems converge to a stationary behavior, whose characterization is the aim of our future work. Similar results have been observed considering other scenarios.

#### 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. Andrews, L. Zhang, "Achieving Stability in Networks of Input-Queued Switches", *IEEE INFOCOM 2001*, Anchorage, Alaska, USA, Apr. 2001, pp. 1673-1679
- [4] J. G. Dai, W. Lin, "Maximum Pressure Policies in Stochastic Processing Networks", *Operations Research*, vol. 53, 2005, pp. 197-218
- [5] 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
- [6] A. Eryilmaz, R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control", *INFOCOM 2005*, March 2005, Miami Fl (USA), pp 1794 - 1803
- [7] S.Floyd, V.Jacobson, "Random Early Detection Gateways for Congestion Avoidance", *IEEE/ACM Transactions on Networking*, vol. 1, n. 4, pp. 397-413, August 1993.
- [8] C.V.Hollot, Y.Liu, V.Misra, D.Towsley, "Unresponsive flows and AQM performance," *IEEE Infocom 2003*, San Francisco, CA, USA, March 2003.
- [9] E. Leonardi, M. Mellia, M. Ajmone Marsan, F. Neri, "On the Throughput Achievable by Isolated and Interconnected Input-Queueing Switches under Multiclass Traffic", *IEEE INFOCOM 2002*, New York, NY, USA, June 2002
- [10] E. Leonardi, M. Mellia, F. Neri, M. Ajmone Marsan, "On the Stability of Input-Queued Switches with Speedup", *IEEE/ACM Trans. on Networking*, vol. 9, n. 1, Feb. 2001, pp. 104-118
- [11] N. McKeown, A. Mekkittikul, V. Anantharam, J. Walrand, "Achieving 100% throughput in an input-queued switch", *IEEE Trans. on Communications*, vol. 47, n. 8, Aug. 1999, pp. 1260-1272
- [12] 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 2000*, Stockholm, Sweden, August 2000.
- [13] M.J. Neely, E. Modiano, C.E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks", *IEEE INFOCOM 2003*, vol. 1, Mar. 2003, pp. 745-755
- [14] M.L. Neely, E. Modiano, C.E. Rohrs "Power Allocation and Routing in Multibeam Satellites with Time-Varying Channels", *IEEE/ACM Trans. on Networking*, vol. 11, n. 1, Feb. 2003, pp. 138-152
- [15] M.J. Neely, E. Modiano, Li Chih-Ping "Fairness and optimal stochastic control for heterogeneous networks", *INFOCOM 2005*, March 2005, Miami Fl (USA), pp 1723 - 1734
- [16] K. Ross, N. Bambos, "Local Search Scheduling Algorithms for Maximal Throughput in Packet Switches", *IEEE INFOCOM 2004*, Mar. 2004
- [17] L. Tassiulas, "Linear complexity algorithms for maximum throughput in radio networks and input queued switches", *IEEE INFOCOM 1998*, San Francisco, CA, USA, Apr. 1998
- [18] 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
- [19] L. Tassiulas, A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multi-hop radio networks", *IEEE Trans. on Automatic Control*, vol. 37, n. 12, Dec. 1992, pp. 1936-1948