Delay-Optimal Scheduling in Bandwidth-Sharing Networks

Maaike Verloop, Rudesindo Núñez-Queija, Sem Borst

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

1 Introduction

Over the past few years, the processor-sharing discipline has emerged as a useful paradigm for evaluating the flow-level performance of elastic data transfers competing for bandwidth on a single bottle-neck link. Bandwidth-sharing networks as considered by Massoulié & Roberts [2] provide a natural extension for modeling the dynamic interaction among competing elastic flows that traverse several links.
Bonald & Massoulié [1] showed that a wide class of α-fair bandwidth-sharing policies as introduced by Mo & Walrand [3] achieve stability in such networks under the simple (and necessary) condition that no individual link is overloaded. While stability is arguably the most fundamental performance criterion, flow-level delays and throughputs are obviously crucial metrics too. Although useful approximations and bounds have been obtained, the latter performance metrics have largely remained intractable in all but a few special cases. In particular, it is not well understood to what extent the flow-level delays and throughputs achieved by common bandwidth-sharing schemes and fairness notions leave potential room for improvement.
Spurred by findings that flow sizes show huge variability, several studies have examined the scope for improvement from size-based scheduling mechanisms such as SRPT and LAS, see for instance Rai *et al.* [4], and demonstrated potentially significant gains over processor-sharing. Yet, the merits of size-based scheduling remain a matter of debate, since the implementation of such mechanisms in core routers presents a major challenge and because the exact gains crucially depend on the performance metric that is adopted.

With the exception of a few papers [6], nearly all studies have considered a single-node scenario, even though there are various indications that priority mechanisms in networks may give rise to starvation effects with possibly disastrous consequences. Recently, it was shown that size-based scheduling disciplines such as SRPT and LAS may unnecessarily fail to achieve stability, even at arbitrarily low loads [5]. Consequently, these disciplines can not render optimal performance, and do not yield any meaningful benchmark for the scope for improvement over standard bandwidth-sharing mechanisms. Our objective here is to determine optimal scheduling strategies for a simple linear network so as to assess the effectiveness of standard allocation policies based on α-fair bandwidth sharing.

2 Stochastic Optimality

We will examine optimal scheduling in a linear bandwidth-sharing network consisting of L nodes each with unit service rate. There are $L + 1$ classes of users, where class-i users require service at node i only, $i = 1, \dots, L$ and class-0 users require service at all L nodes simultaneously. Class-i users arrive according to a Poisson process with parameter λ_i and have exponentially distributed service requirements with mean $1/\mu_i$, $i = 0, \dots, L$.
We seek scheduling policies that in some appropriate sense minimize the total number of users in the network. We only allow (possibly preemptive) policies that have no knowledge available of the remaining service requirements (this excludes e.g., SRPT). In order to minimize the total number of users, in the short run an "optimal" policy must maximize the total output rate of the system, but at the same time it needs to achieve a high degree of service parallelism, to achieve good performance over long intervals (and, in particular, ensure stability when possible). When there is no conflict between these two objectives, there

exists a **stochastically** optimal policy:

Proposition 1 *If $\sum_{i=1}^{L}\mu_i \leq \mu_0$, then the number of users is minimized (at any time) by giving preemptive priority to class 0. Similarly, if $\sum_{i=1}^{L}\mu_i \geq \mu_0$ and $\sum_{i=1,i\neq j}^{L}\mu_i \leq \mu_0$ for all j, then in all nodes full service must be allocated to class 0, unless all other classes are present as well. In both cases, if class 0 is empty, all other classes with at least one user present are served at full rate.*

3 Switching Curve

The case that is not yet covered is when $\sum_{i=1,i\neq j}^{L}\mu_i \geq \mu_0$ for some $j = 1,\ldots,L$. Then no stochastically optimal policy exists, as can be argued by comparing short-term and long-term advantages of different scheduling rules. We will therefore be concerned with policies that minimize the *mean* number of jobs at any time and restrict ourselves to the case $L = 2$ with $\mu_1 > \mu_0$. It can be shown that if users of both classes 1 and 2 are present it is optimal to serve them at full rate (as is intuitively clear). The (less obvious) structure of the optimal allocation when only users of classes 0 and 1 are present is described in the next proposition. We use the random variable $N_i(t)$ to denote the number of class-i users at time t.

Proposition 2 *There exists a switching curve $h(\cdot)$ such that, when $N_2(t) = 0$, it is optimal to serve class 0 at full rate if $N_1(t) \leq h(N_0(t))$ and to serve class 1 at full rate otherwise.*

In general no exact characterization of $h(\cdot)$ exists, but even when achievable, it may be too complex for practical purposes. For this reason we investigate a related fluid model which can be seen as the limit of our model considered at a large time scale. The next proposition characterizes the optimal fluid policy. We use $n_i(t)$ to denote the (scaled) length of the queue of class i at time t.

Proposition 3 *Assume $\rho_1 \leq \rho_2$ and $n_2(t) = 0$. If $\mu_1, \mu_2 \geq \mu_0$, it is optimal to serve class 0 at rate $1-\rho_2$ (hence keeping $n_2(t)$ equal to zero) whenever $n_1(t) \leq \frac{\rho_2-\rho_1}{1-\rho_0-\rho_2} n_0(t)$ and fully serve class 1 otherwise. If $\mu_1 \geq \mu_0 \geq \mu_2$, then the corresponding condition is $n_1(t) \leq \frac{\mu_2}{\mu_1+\mu_2-\mu_0}\frac{\mu_1}{\mu_0}\frac{\rho_2-\rho_1}{1-\rho_0-\rho_2} n_0(t)$.*

Through numerical experiments we observed a good match between these limiting switching curves and the exact optimal strategies in the stochastic model. We also made a numerical comparison between the (weighted) mean delay across the full spectrum of α-fair strategies and the optimal policies. In all our experiments we observed that (i) the differences within the class of α-fair allocations are not significant, and (ii) they compare well with the optimal strategies. Apparently, α-fair mechanisms succeed in *dynamically* adjusting the bandwidth allocation in an efficient manner, without knowledge of the distributions.

References

[1] Bonald, T., Massoulié, L. (2001). Impact of fairness on Internet performance. In: *Proc. ACM Sigmetrics/Performance 2001 Conf.*, 82–91.

[2] Massoulié, L. Roberts, J.W. (2000). Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems* **15** 185–201.

[3] Mo, J., Walrand, J. (2000). Fair end-to-end based congestion control. *IEEE/ACM Trans. Netw.* **8**, 556–567.

[4] Rai, I.A., Urvoy-Keller, G., Biersack, E.W. (2003). Analysis of LAS scheduling for job size distributions with high variance. In: *Proc. ACM Sigmetrics 2003 Conf.*, 218–228.

[5] Verloop, I.M., Borst, S.C., Núñez Queija, R. (2005). *Stability of size-based scheduling disciplines in resource-sharing networks.* To appear in: *Proc. Performance 2005.*

[6] Yang, S., De Veciana, G. (2004). Enhancing both network and user performance for networks supporting best-effort traffic. *IEEE/ACM Transactions on Networking* **12**, 349–360.