Stability of processor sharing networks with simultaneous resource requirements

Jennie Hansen, Cian Reynolds and Stan Zachary

Heriot-Watt University
Edinburgh

We study the phenomenon of *entrainment* in processor sharing networks, whereby, while individual network resources have sufficient capacity to met demand, the requirement for simultaneous availability of resources means that a network may nevertheless be unstable. We show that instability occurs through poor control, and that, for a variety of network topologies, only small modifications to controls are required in order to ensure stability. For controls which possess a natural monotonicity property, we give some new results for the classification of the corresponding Markov processes, which lead to conditions both for stability and for instability.

Keywords: processor sharing networks, stability, entrainment.

AMS 2000 subject classification: Primary: 60K20; Secondary: 60K25.

Short title. Stability of processor sharing networks

1 Introduction

Modern communications networks, such as the Internet, are able at any time to share their resources, for example bandwidth, among those calls or connections currently in progress. Such calls may require simultaneous capacity from several resources in the network. For example, "streaming" applications require, for their duration, a reservation of bandwidth on each network link over which they connect. It may then happen that while each resource in the network, considered in isolation, has sufficient capacity to service the demand placed on it, the control of the network is such that the requirement for *simultaneous* availability of capacity ensures that over time demand cannot be met, and that the network is unstable, that is, that the number of calls present in the network tends to infinity. This is the phenomenon of *entrainment* and has been previously studied in this context by various authors, notably Bonald and Massoulié (2001), de Veciana *et al* (2001), Kelly and Williams (2004). In particular these authors have considered a broad class of so-called *fair-sharing* control strategies—see below. Bonald and Massoulié and de Veciana *et al* have shown that here the instability problems referred to above do not arise, i.e. that provided the various network resources individually have sufficient capacity, under any fair-sharing control the network will remain stable. In the present paper we study the phenomenon of entrainment in more detail and characterise some of those conditions which give rise to its occurrence. In particular we show that it frequently arises through poor control whenever there are no calls of certain classes in the network, and that only minor adjustments to control strategies are required in order to avoid it. Thus very flexible management schemes, including those utilising significant prioritisation, may be safely implemented. We further establish, in Section 3, some new results for the stability of Markov chains whose transitions rates possess a natural monotonicity property, yielding both necessary and sufficient conditions for the stability of a wide class of network controls.

Since network parameters—in particular call arrival rates—are unlikely to be known in advance, it is of particular concern to identify controls whose stability is robust with respect to variations in these parameters.

We take as our model the following, which is essentially that introduced by Roberts and Massoulié (2000). Let R denote the finite set of possible call, or connection, types. We denote the state of the network at time t by $\boldsymbol{n}(t) = (n_r(t), \, r \in R)$, where $n_r(t)$ is the number of calls of each type r in progress at that time. Calls of each type $r \in R$ arrive at the network as a Poisson process with rate ν_r and have *sizes* which are exponentially distributed with mean μ_r^{-1}. Arrival processes and call sizes are all independent. (As usual, the above distributional assumptions make for simplicity of analysis. However, it seems likely that all the results of this paper remain qualitatively correct for other distributions of interarrival times and call sizes, subject only to these distributions having finite means, and to the same independence assumptions.) When the state of the network is $\boldsymbol{n}$, calls of each type r are allocated *in total* a *bandwidth* $b_r(\boldsymbol{n}) \geq 0$. We assume always that $b_r(\boldsymbol{n}) = 0$ whenever $n_r = 0$. The process $\boldsymbol{n}(\cdot)$ is thus Markov with state space $\mathbb{Z}_+^{|R|}$ and transition rates given by, for all $r \in R$,

$$\boldsymbol{n} \to \begin{cases} \boldsymbol{n} + \boldsymbol{e}_r & \text{at rate } \nu_r, \\ \boldsymbol{n} - \boldsymbol{e}_r & \text{at rate } \mu_r b_r(\boldsymbol{n}), \end{cases} \tag{1}$$

where $\boldsymbol{e}_r$ denotes the vector whose rth component is 1 and whose other components are 0. For each $\boldsymbol{n}$ define also $\boldsymbol{b}(\boldsymbol{n}) = (b_r(\boldsymbol{n}), \, r \in R)$. We shall refer to $\boldsymbol{b} = (\boldsymbol{b}(\boldsymbol{n}), \, \boldsymbol{n} \in \mathbb{Z}_+^{|R|})$ as the *control* for the Markov process $\boldsymbol{n}(\cdot)$. The set of *feasible* controls $\boldsymbol{b}$ is defined by a set of capacity constraints

$$\sum_{r \in R} a_{jr} b_r(\boldsymbol{n}) \leq c_j, \qquad j \in J, \tag{2}$$

indexed in a finite set J, where each $a_{jr} \in \{0, 1\}$. Here c_j may be thought of as the capacity of resource j, and a call of type r utilises resource j if and only if $a_{jr} = 1$.

In order to allow some results to be stated with sufficient generality, it is convenient to allow the possibility that, for any j, we may have $c_j = \infty$ (corresponding to the effective nonexistence of the resource constraint j). However, we assume, without loss of generality, that, for all $r \in R$, there exists some $j \in J$ with $c_j < \infty$ and $a_{jr} = 1$.

Our interest is in identifying, for fixed values of the parameters ν_r, μ_r, c_j, and a_{jr}, $r \in R$, $j \in J$ (which we shall regard as defining a given network) those controls $\boldsymbol{b}$ which are both feasible and *stable*, where we now take the latter to mean that the corresponding Markov process $\boldsymbol{n}(\cdot)$ is positive recurrent. A control $\boldsymbol{b}$ for which $\boldsymbol{n}(\cdot)$ is null recurrent or transient is referred to as *unstable*.

For any feasible control $\boldsymbol{b}$, for any $\boldsymbol{n}$, we shall say that a resource $j \in J$ is *saturated* if the corresponding constraint (2) is satisfied with equality. Following Bonald and Massoulié (2001), we shall further say that a feasible control $\boldsymbol{b}$ is *Pareto efficient* if, for all $\boldsymbol{n}$ and for all r such that $n_r > 0$, there exists $j \in J$ with $a_{jr} = 1$ such that j is saturated (so that $b_r(\boldsymbol{n})$ may not be increased without either decreasing $b_{r'}(\boldsymbol{n})$ for some $r' \neq r$ or else violating the constraints (2).) It is sometimes helpful to consider controls which are not feasible, so we note in particular that the requirement of Pareto efficiency includes that of feasibility.

For each $r \in R$ we define $\kappa_r = \nu_r / \mu_r$, which may be thought of as the rate at which "work" of type r arrives at the network. Many, but not all, stability results depend on

the parameters ν_r and μ_r only through the corresponding κ_r. As observed by Bonald and Massoulié (2001), a necessary and sufficient condition for the existence of a stable feasible control is given by

$$\sum_{r \in R} a_{jr} \kappa_r < c_j, \qquad j \in J. \tag{3}$$

For the necessity of this condition, observe that if it is violated for some j, then since, for any feasible control $\boldsymbol{b}$, we then have $\sum_{r \in R} a_{jr}(\kappa_r - b_r(\boldsymbol{n})) \geq 0$, standard arguments—see, for example, Asmussen (2003, Chapter 1, Proposition 5.4) or the argument of Kelly and Williams (2004)—show that $\boldsymbol{b}$ cannot be stable. For the sufficiency of the condition (3), define, for any vector $\hat{\boldsymbol{b}} = (\hat{b}_r, r \in R)$, the corresponding *complete partitioning* control $\boldsymbol{b}$ by $b_r(\boldsymbol{n}) = \hat{b}_r$ whenever $n_r > 0$. Note that this defines a process $\boldsymbol{n}(\cdot)$ which corresponds to $|R|$ independent queues and is such that, in each queue r, arrivals occur at rate ν_r and departures (when $n_r > 0$) at rate $\mu_r \hat{b}_r$. Thus a necessary and sufficient condition for the stability of this control is that $\hat{b}_r > \kappa_r$ for all $r \in R$. In particular we may choose $\hat{\boldsymbol{b}}$ such that the corresponding complete partitioning control is stable and feasible if and only if the condition (3) holds. Further, given such a $\hat{\boldsymbol{b}}$, we may clearly define a Pareto efficient control $\boldsymbol{b}'$ such that $b'_r(\boldsymbol{n}) \geq \hat{b}_r$ for all $\boldsymbol{n}$ and for all r such that $n_r > 0$. The corresponding process $\boldsymbol{n}'(\cdot)$ may then be coupled to the process $\hat{\boldsymbol{n}}(\cdot)$ corresponding to $\hat{\boldsymbol{b}}$ in such a way that $n'_r(t) \leq \hat{n}_r(t)$ for all r and for all $t \geq 0$. Hence the condition (3) is also sufficient for the existence of *some* stable Pareto efficient control.

It is also clear that, in the case $|J| = 1$ of a single resource constraint, the condition (3) is sufficient to ensure the stability of any Pareto efficient control. This is not in general true when $|J| > 1$, as is shown by Example 1.1 below, which is a simplification of one given by Bonald and Massoulié (2001).

Example 1.1. Suppose that $R = \{1, 2\}$, $J = \{1, 2\}$ and that the matrix $A = (a_{jr}, j \in J, r \in R)$ is given by

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Thus calls of type 1 are constrained by resource 1 only, while calls of type 2 are constrained by both resources 1 and 2. Suppose, without loss of generality, that $c_2 \leq c_1$. We assume the condition (3), which here becomes

$$\kappa_1 + \kappa_2 < c_1, \qquad \kappa_2 < c_2. \tag{4}$$

Consider the Pareto efficient control in which complete priority is given to calls of type 1. Then the long-run fraction of time in which the network is empty of calls of type 1—and so resource 1 is available for use by calls of type 2—is given by $1 - \kappa_1/c_1$. Since, when this resource is available, calls of type 2 are processed at rate c_2, standard arguments for the stability of a single-server queue now show that the control is stable if and only if $\kappa_2 < c_2(1 - \kappa_1/c_1)$. This is a condition which is generally more restrictive than (4) above. When it is violated we have the phenomenon of entrainment referred to above; that is, from (4), each resource in the network, considered in isolation, has sufficient capacity, but the given Pareto efficient control is nevertheless unstable.

In Example 1.1 above instability may be considered as occurring because such high priority is given to calls of type 1 as to ensure that, when these are emptied from the system, resource 1 is thereafter consistently underutilised. As we shall show later, this problem

3

may in general be avoided by choosing Pareto efficient controls $\boldsymbol{b}$ such that, for each r, $b_r(\boldsymbol{n})$ is usually small whenever n_r is small. In particular this property is possessed by the class of *(weighted) fair-sharing* controls introduced by Mo and Walrand (2000) in a generalisation of various classes considered earlier (see, for example, Kelly *et al* (1998)). These are defined as follows: for $\alpha > 0$, $\alpha \neq 1$, and weights w_r, $r \in R$, a *weighted α-fair* control, or bandwidth allocation, is given by taking, for each $\boldsymbol{n}$, $\boldsymbol{b}(\boldsymbol{n})$ to maximise the concave function

$$\sum_{r \in R} w_r n_r^\alpha \frac{b_r(\boldsymbol{n})^{1-\alpha}}{1-\alpha}, \tag{5}$$

subject to the capacity constraints (2). This class is further extended to each of the cases $\alpha = 0, 1, \infty$ by taking the limit of the α-fair control as α tends to each of these values. (For $\alpha = 1$ this is equivalent to replacing the quantity defined in (5) by $\sum_{r \in R} n_r \log b_r(\boldsymbol{n})$.) For the cases $\alpha = 1$ and $\alpha = \infty$, De Veciana *et al* (2001) use Lyapunov function techniques to show that the condition (3) is sufficient for the stability of (weighted) α-fair controls. Bonald and Massoulié (2001) show that this result holds for general α by using fluid limits and appealing to a result of Dai (1995) for multiclass queueing networks. (In fact it is not certain that Dai's result is directly applicable to the present networks with their simultaneous resource requirements; however, Bonald and Massoulié's approach is essentially equivalent to showing that their function f defined by

$$f(\boldsymbol{n}) = \sum_{r \in R} w_r \mu_r^{-1} \kappa_r^{-\alpha} \frac{n_r^{\alpha+1}}{\alpha+1}$$

is a suitable Lyapunov function for establishing the sufficiency of (3) for the stability of α-fair controls, and this requires only a small extension to their argument.) We note in particular that the stability of the fair-sharing controls is robust with respect to parameter variation, subject of course to (3) being satisfied.

The resource underutilisation of Example 1.1 may be further understood as resulting from the nonsmooth nature of the call arrival process. Consider the analogous *fluid model* in which "work" of each type r arrives steadily at rate κ_r and is processed at rate $b_r(\boldsymbol{n})$, where each n_r is now the volume of work of type r in the network and where $\boldsymbol{b}$ is again subject to constraints of the form (2). Then, under the condition (4) of Example 1.1, it is easy to see that every Pareto efficient control is stable, in the sense here that the total volume of work in the system eventually becomes and remains zero. For our stochastic model, the possible modified control discussed above, in which $b_1(\boldsymbol{n})$ is kept small whenever n_1 is small and n_2 is large, may be seen as a smoothing operation forcing the behaviour of the stochastic model to follow more closely that of the fluid model. However, as Example 1.2 below shows, instability may also occur in ways such that even the corresponding control for the analogous fluid model is also unstable.

For any control $\boldsymbol{b}$ and any function f on $\mathbb{Z}_+^{|R|}$, define the function $D_{\boldsymbol{b}} f$ on $\mathbb{Z}_+^{|R|}$ by

$$D_{\boldsymbol{b}} f(\boldsymbol{n}) = \sum_{r \in R} \left[\nu_r \left(f(\boldsymbol{n} + \boldsymbol{e}_r) - f(\boldsymbol{n}) \right) + \mu_r b_r(\boldsymbol{n}) \left(f(\boldsymbol{n} - \boldsymbol{e}_r) - f(\boldsymbol{n}) \right) \right]. \tag{6}$$

(Since, for $\boldsymbol{n}$ and r such that $n_r = 0$, we have also $b_r(\boldsymbol{n}) = 0$, there is no problem arising from the lack of a formal definition of $f(\boldsymbol{n} - \boldsymbol{e}_r)$ in this case. Further, $D_{\boldsymbol{b}}$ may be thought of as the generator of the Markov process $\boldsymbol{n}(\cdot)$ under the control $\boldsymbol{b}$.)

Example 1.2. Consider the network defined by $R = \{1, 2, 3\}$, $J = \{1, 2, 3\}$,

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

with $\nu_r = \nu$, $\mu_r = 1$ for all r and $c_j = c$ for all j. Assume that the condition (3) is satisfied, i.e. that $2\nu < c$. Consider any control $\boldsymbol{b}$ such that, for all $\boldsymbol{n} \neq \boldsymbol{0}$, we have $b_r(\boldsymbol{n}) = c$ for some r such that $n_r > 0$ and $b_s(\boldsymbol{n}) = 0$ for $s \neq r$. Then $\boldsymbol{b}$ is Pareto efficient. However, for the function $f(\boldsymbol{n}) := \sum_{r \in R} n_r$ and for all $\boldsymbol{n} \neq \boldsymbol{0}$, we have $D_{\boldsymbol{b}} f(\boldsymbol{n}) = 3\nu - c$. Standard martingale arguments—again see Asmussen (2003, Chapter 1, Proposition 5.4)—now show that $\boldsymbol{b}$ is unstable whenever $3\nu \geq c$. In this example the instability is not simply the result of poor control for $\boldsymbol{n}$ close to the boundary of $\mathbb{Z}_+^3$, and is equally present in the analogous fluid model. We return to this example in Section 3.

Our aim in the remainder of paper is to identify more general classes of stable controls, to provide insight into how the phenomenon of entrainment rises, and to show how controls may be modified if necessary so as to avoid it. In Section 2 we consider a number of fairly simple network topologies, and use Lyapunov function techniques to show that, for these, any Pareto efficient control $\boldsymbol{b}$ is stable, provided *only* that, as discussed above and for suitable r, $b_r(\boldsymbol{n})$ is modified so as to be small whenever n_r is small. In Section 3 we consider controls which possess a natural monotonicity property, likely to be satisfied in any applications. We introduce some new analytical techniques to prove some fairly general results for the stability of the associated Markov chains. These give sufficient conditions for stability, which, for many classes of control, are also close to being necessary. These results are applicable to a substantial class of priority-based controls.

2   Simple network topologies

In this section we consider a number of network topologies and in each case show that, under the condition (3), *any* Pareto efficient control is stable provided only that it is suitably modified for values of $\boldsymbol{n}$ close to the boundary of $\mathbb{Z}_+^{|R|}$.

The results of this section are based on simple Lyapunov function techniques, in particular Foster's criterion. Proposition 2.1 below states the specialisation of this to the present problem—see, for example, Asmussen (2003, Chapter 1, Proposition 5.3(ii)), noting that here jumps of the process $\boldsymbol{n}(\cdot)$ may only occur between neighbouring states, and also that the usual uniformisation argument translates statements for discrete-time processes to the present continuous time setting.

Proposition 2.1. *Given any control $\boldsymbol{b}$, suppose that there exists a positive function f on $\mathbb{Z}_+^{|R|}$, a finite subset F of $\mathbb{Z}_+^{|R|}$ and some $\epsilon > 0$ such that $D_{\boldsymbol{b}} f(\boldsymbol{n}) \leq -\epsilon$ for all $\boldsymbol{n} \notin F$. Then $\boldsymbol{b}$ is stable.*

In considering the various network topologies of this section, it is convenient to define, for any $a \geq 1$, the function g_a on $\mathbb{Z}_+$ by

$$g_a(n) = \begin{cases} \dfrac{a}{2} + \dfrac{n^2}{2a}, & \text{if } n < a, \\ n, & \text{if } n \geq a, \end{cases}$$

We first consider again the very simple Example 1.1 of the Introduction, but with a general control $\boldsymbol{b}$.

Example 2.1. Let R, J and A be as in Example 1.1 and assume again that $c_2 \leq c_1$ and also the condition (4). We show that, given any $\delta > 0$, there is some $a \geq 1$ (depending on the parameters ν_r, μ_r and c_j and δ) such that a sufficient condition for the stability of any Pareto efficient control $\boldsymbol{b}$ is given by

$$\kappa_2 - b_2(\boldsymbol{n}) \leq -\delta, \qquad \text{whenever } n_1 < a \text{ and } \boldsymbol{n} \notin F, \tag{7}$$

for some finite set F (necessarily including $\{\boldsymbol{n} \colon n_1 < a, n_2 = 0\}$). (Note that, from (4), condition (7) may be satisfied for any $\delta \in (0, c_2 - \kappa_2]$.)

Observe first that, from (4) and the saturation of the resource 1 whenever $\boldsymbol{n} \neq \boldsymbol{0}$, there exists $\delta' > 0$ such that, for any Pareto efficient control $\boldsymbol{b}$,

$$\sum_{r=1}^{2} (\kappa_r - b_r(\boldsymbol{n})) \leq -\delta', \qquad \text{for all } \boldsymbol{n} \neq \boldsymbol{0}. \tag{8}$$

For any fixed a, define the function f_a on $\mathbb{Z}_+^2$ by

$$f_a(\boldsymbol{n}) = \frac{g_a(n_1)}{\mu_1} + \frac{n_2}{\mu_2}.$$

Elementary calculations show that, from (6), for any Pareto efficient control $\boldsymbol{b}$ satisfying condition (7) for some finite F, and for any $\boldsymbol{n} \notin F$,

$$D_{\boldsymbol{b}} f_a(\boldsymbol{n}) = \min\left(\frac{n_1}{a}, 1\right) [\kappa_1 - b_1(\boldsymbol{n})] + \kappa_2 - b_2(\boldsymbol{n}) + \frac{1}{2a} h_{1,a}(\boldsymbol{n})$$

$$\leq -\min\left(\frac{n_1}{a}, 1\right) \delta' - \max\left(1 - \frac{n_1}{a}, 0\right) \delta + \frac{1}{2a} h_{1,a}(\boldsymbol{n}) \tag{9}$$

$$\leq -\min(\delta, \delta') + \frac{1}{2a} h_{1,a}(\boldsymbol{n}), \tag{10}$$

where the function $h_{1,a}$ is given by

$$h_{1,a}(\boldsymbol{n}) = \begin{cases} \kappa_1 + b_1(\boldsymbol{n}), & \text{if } 0 \leq n_1 < a, \\ b_1(\boldsymbol{n}), & \text{if } n_1 = a, \\ 0, & \text{if } n_1 > a, \end{cases} \tag{11}$$

and where (9) follows from (7) and (8). Since $b_1(\boldsymbol{n}) \leq c_1$ for any feasible control $\boldsymbol{b}$, it follows from (10) and (11) that a may be chosen sufficiently large that, for any Pareto efficient control $\boldsymbol{b}$ satisfying (7) for some finite F,

$$D_{\boldsymbol{b}} f_a(\boldsymbol{n}) \leq -\frac{1}{2} \min(\delta, \delta') \qquad \text{for all } \boldsymbol{n} \notin F,$$

and so, by Proposition 2.1, $\boldsymbol{b}$ is stable.

Thus, for this example, any Pareto efficient control is stable provided only that, outside of some finite set F, it is suitably modified for values of $\boldsymbol{n}$ such that n_1 is small. However, in practice the parameters ν_r in particular are unlikely to be known, and so it is desirable to choose controls whose stability is robust. One such possibility is to choose any Pareto efficient control which assigns complete priority to calls of type 2 whenever $n_1 < a$ for

some a. Provided only that the condition (4) is satisfied, the above result shows that this will be stable provided a is sufficiently large. The precise value of a required depends on the slack in the inequalities (4), but calculations for the "worst case", in which calls of type 1 have complete priority whenever $n_1 \geq a$, show that in general a need only be small.

Note in particular that the various fair-sharing controls defined in the Introduction always satisfy the condition (7) for some δ and for some sufficiently large F, and hence (as already remarked) are always stable provided that the condition (4) is satisfied.

We now extend the above example to each of two more general network topologies. In each case it is again the case that only small modifications, identified below, are required to Pareto efficient controls in order to ensure their stability.

Example 2.2. Consider the network with $R = \{1, \ldots, |R|\}$, and in which each call type r requires service from a dedicated resource of capacity c_r together with service from a resource which is shared by all call types and has capacity c_0. The constraints (2) defining the feasible controls $\boldsymbol{b}$ are thus

$$\sum_{r \in R} b_r(\boldsymbol{n}) \leq c_0, \tag{12}$$

$$b_r(\boldsymbol{n}) \leq c_r, \qquad r \in R. \tag{13}$$

We assume, without loss of generality, that $c_0 < \infty$ and that

$$c_0 \leq \sum_{r \in R} c_r. \tag{14}$$

As usual we assume the condition (3), which here becomes

$$\sum_{r \in R} \kappa_r < c_0, \tag{15}$$

$$\kappa_r < c_r, \qquad r \in R. \tag{16}$$

The case $|R| = 2$ with $c_1 = \infty$ is the earlier Example 2.1 (with c_0 here corresponding to c_1 there). However, in the general case described above a little more care is needed in the conditions for the stability of Pareto efficient controls.

For any $\boldsymbol{n} \in \mathbb{Z}_+^{|R|}$, define $n_{\min} = \min_{r \in R} n_r$. We show briefly that, given any $\delta > 0$, there is again some $a \geq 1$ (depending on the parameters ν_r, μ_r and c_j and δ) such that a sufficient condition for the stability of any Pareto efficient control $\boldsymbol{b}$ is given by

$$\sum_{r\,:\,n_r \geq a'} (\kappa_r - b_r(\boldsymbol{n})) \leq -\delta, \qquad \text{for all } a' \in [1, a] \text{ and } \boldsymbol{n} \text{ such that } n_{\min} < a, \, \boldsymbol{n} \notin F, \tag{17}$$

for some finite set F. (The existence of a $\delta > 0$ such that the condition (17) may be satisfied is guaranteed by (15) and (16). Further, it is not difficult to see that in the case $R = 2$, this condition reduces to that of Example 2.1 taken together with a similar condition in which the roles of calls of types 1 and 2 are interchanged.)

The proof of this result is similar to that of Example 2.1. Note first that it follows from (14) that, for any Pareto efficient control, the resource 0 is necessarily saturated for any $\boldsymbol{n}$ such that $n_{\min} \geq 1$. Hence, from (16), there exists $\delta' > 0$ such that, again for any Pareto efficient control $\boldsymbol{b}$,

$$\sum_{r \in R} (\kappa_r - b_r(\boldsymbol{n})) \leq -\delta' \qquad \text{for all } \boldsymbol{n} \text{ such that } n_{\min} \geq 1. \tag{18}$$

7

For any fixed a, define the function f_a on $\mathbb{Z}_+^{|R|}$ by

$$f_a(\boldsymbol{n}) = \sum_{r \in \mathbb{R}} \frac{g_a(n_r)}{\mu_r}.$$

Elementary calculations show that, from (6), for any Pareto efficient control $\boldsymbol{b}$ satisfying the condition (17) for some finite F, and for any $\boldsymbol{n} \notin F$,

$$
\begin{aligned}
D_{\boldsymbol{b}} f_a(\boldsymbol{n}) &= \sum_{r \in R} \min\left(\frac{n_r}{a}, 1\right) [\kappa_r - b_r(\boldsymbol{n})] + \frac{1}{2a} \sum_{r \in R} h_{r,a}(n_r), \\
&= \frac{1}{a} \sum_{a'=1}^{a} \sum_{r:\, n_r \geq a'} (\kappa_r - b_r(\boldsymbol{n})) + \frac{1}{2a} \sum_{r \in R} h_{r,a}(n_r) \\
&\leq -\min(\delta, \delta') + \frac{1}{2a} \sum_{r \in R} h_{r,a}(n_r),
\end{aligned}
\tag{19}
$$

where, for each r, the function $h_{r,a}$ is given by (11) with the index r replacing the index 1, and where (19) follows from (17) for $\boldsymbol{n}$ such that $n_{\min} < a$ and from (18) for $\boldsymbol{n}$ such that $n_{\min} \geq a$. It now follows as in Example 2.1 that a may be chosen sufficiently large that, for any Pareto efficient control $\boldsymbol{b}$ satisfying (17) for some finite F,

$$D_{\boldsymbol{b}} f_a(\boldsymbol{n}) \leq -\frac{1}{2} \min(\delta, \delta') \qquad \text{for all } \boldsymbol{n} \notin F,$$

and so, again by Proposition 2.1, the control $\boldsymbol{b}$ is stable.

Thus, in order to ensure the stability of a general Pareto efficient control $\boldsymbol{b}$, it is only necessary to appropriately modify $\boldsymbol{b}(\boldsymbol{n})$ for $\boldsymbol{n}$ such that $n_{\min} < a$. In particular it follows from the above result that a Pareto efficient control $\boldsymbol{b}$ whose stability is reasonably robust is given by requiring that, for some a, for all $\boldsymbol{n}$ (such that $n_{\min} < a$) and for all $a' \in [1, a]$, calls of types r such that $n_r \geq a'$ collectively have complete priority over calls of the remaining types, that is, that

$$\sum_{r:\, n_r \geq a'} b_r(\boldsymbol{n}) = \min\left(c_0, \sum_{r:\, n_r \geq a'} c_r\right).$$

For a sufficiently large, depending on the slack in the inequalities (15) and (16), any such control is stable.

Now consider further the case $R = 2$. Note that the present topology is completely general for a network with two call types. Let $\boldsymbol{b}$ be any Pareto efficient control such that

$$\lim_{n_1 \to \infty} b_1(n_1, n_2) = c_1 \qquad \text{for all } n_2, \tag{20}$$

$$\lim_{n_2 \to \infty} b_2(n_1, n_2) = c_2 \qquad \text{for all } n_1. \tag{21}$$

Then it is straightforward that, for $\delta < \min_{r=1,2}(c_r - \kappa_r)$, and for any a, the condition (17) is satisfied for F sufficiently large, and so $\boldsymbol{b}$ is stable.

In particular the conditions (20) and (21) are again satisfied by the various fair-sharing controls. The present conditions are of course considerably more general. However the development of corresponding results for general networks with $R \geq 3$ remains a challenging problem.

Our final network topology is a simple "backbone" structure.

Example 2.3. Consider a network with resource set $J = \{1, \ldots, k\}$, where resource j has capacity c_j as usual. The set of call types is given by $R = \{0, 1, \ldots, k\}$ where calls of each type $r = 1, \ldots, k$ require service from the single resource $j = r$, while calls of type 0 require service from each of the resources $1, \ldots, k$.

The constraints (2) defining the feasible controls $\boldsymbol{b}$ are given by

$$b_0(\boldsymbol{n}) + b_r(\boldsymbol{n}) \leq c_r, \qquad r = 1, \ldots k. \tag{22}$$

Again we assume the condition (3), which here becomes

$$\kappa_0 + \kappa_r < c_r, \qquad r = 1, \ldots k. \tag{23}$$

The state of the network is thus here denoted by $\boldsymbol{n} = (n_0, n_1, \ldots, n_k)$. For any such $\boldsymbol{n}$ define $\hat{n}_{\max} = \max(n_1, \ldots, n_k)$.

This example again generalises that of Example 2.1 with the call type 0 here playing the role of the call type 2 in that example. In a fairly straightforward generalisation of the result of that example, it is here the case that, given any $\delta > 0$, there is again an $a \geq 1$ such that a sufficient condition for the stability of any Pareto efficient control $\boldsymbol{b}$ is given by

$$\kappa_0 - b_0(\boldsymbol{n}) \leq -\delta, \qquad \text{whenever } \hat{n}_{\max} < a \text{ and } \boldsymbol{n} \notin F, \tag{24}$$

for some finite set F

To see this, suppose first that $\mu_r = 1$ for all $r = 0, 1, \ldots, k$ and that the condition (24) is satisfied. For any $a \geq 1$, define the function f_a on $\mathbb{Z}_+^{|R|}$ by $f_a(\boldsymbol{n}) = n_0 + g_a(\hat{n}_{\max})$. Then, as in the case of Example 2.1, it follows straightforwardly from the conditions (23) and (24) that there exists $\delta' > 0$ and a sufficiently large that $D_{\boldsymbol{b}} f_a(\boldsymbol{n}) \leq -\delta'$ for all $\boldsymbol{n} \in \mathbb{Z}_+^{|R|} \setminus F$ such that additionally $\hat{n}_{\max} = n_r$ for a *single* value of $r = 1, \ldots, k$. It is now easy to see that this latter restriction may be removed (possibly at the expense of increasing a) by suitably smoothing the function f in the neighbourhood of those $\boldsymbol{n}$ such that $\hat{n}_{\max} = n_r$ for two or more values of $r = 1, \ldots, k$. The desired result thus follows in this case; for general μ_r only routine modifications to the above argument are required.

Hence we again have that any Pareto efficient control requires only slight modification—for those $\boldsymbol{n}$ such that $\hat{n}_{\max} < a$—in order to be stable. A robust Pareto efficient control is given, for example, by assigning complete priority to calls of type 0 whenever $\hat{n}_{\max} < a$, the necessary value of a depending on the slack in the inequalities (23).

Examples 2.1–2.3 above all consider fairly simple network topologies. In the analogous fluid model defined in the Introduction, it is easily seen that, for each of these topologies, condition (3) is sufficient for the stability of *any* Pareto efficient control. (In each case this follows, for example, by using the same Lyapunov function as for the stochastic model, except that the function g_a may be replaced by the identity function.) The examples illustrate a principle which seems likely to be true for more general network topologies, namely that when a control is such that it is stable for the fluid model, then there is a closely approximating control which is stable for the corresponding stochastic model.

In the next section we consider stability criteria for quite general network topologies, applicable typically to controls where there is some prioritisation among call types.

3 Stability of monotonic controls

Many controls likely to be of practical application possess a simple monotonicity property (see below). In this section we study stability for a wide class of such controls, giving sufficient conditions for stability, which, for many classes of control, are also close to being necessary (see Remark 3.2 below which further discusses the applicability of the results of this section). We require first the following quite general lemma.

We shall say that a control $\boldsymbol{b}$ is *bounded* if, for all r, $b_r(\boldsymbol{n})$ is bounded in $\boldsymbol{n}$.

Lemma 3.1. *Let $\boldsymbol{b}$ be any bounded control and, as usual, let $\boldsymbol{n}(\cdot)$ denote the Markov process for the corresponding network. Then, for all $r \in R$,*

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t b_r(\boldsymbol{n}(u)) \, du \leq \kappa_r \quad a.s. \tag{25}$$

Proof. For all $r \in R$, the compensated process $n_r^*(\cdot)$ defined by

$$n_r^*(t) = n_r(t) - n_r(0) + \int_0^t (b_r(\boldsymbol{n}(u)) - \kappa_r) \, du$$

is a zero-mean martingale. Since also the transition rates of the Markov process $\boldsymbol{n}(\cdot)$ are bounded, it follows that, for some constant M and for all $t \geq 0$, we have $\mathbf{E}(n_r^*(t)^2) \leq Mt$. Thus, for $1/2 < \alpha < 1$, the process $(n_r^*(t)/t^\alpha)_{t>0}$ is an $\mathcal{L}^2$-bounded supermartingale, and so, as $t \to \infty$, converges almost surely to some finite random variable. It follows that

$$\frac{n_r(t)}{t} + \frac{1}{t} \int_0^t b_r(\boldsymbol{n}(u)) \, du \to \kappa_r \quad \text{a.s.}$$

Since the process $n_r(\cdot)$ is positive, the result (25) now follows. $\qquad \square$

Remark 3.1. In the case where the bounded control $\boldsymbol{b}$ is stable, the Markov process $\boldsymbol{n}(\cdot)$ is positive recurrent, and so we have the stronger result that, for all $r \in R$,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t b_r(\boldsymbol{n}(u)) \, du = \mathbf{E}_\pi b_r = \kappa_r \quad \text{a.s.} \tag{26}$$

Here $\mathbf{E}_\pi b_r$ denotes the expectation of the function b_r with respect to the stationary distribution π of $\boldsymbol{n}(\cdot)$, the first equality in (26) follows from the ergodic theorem, and the second is simply the assertion that, under stationarity (and readily deducible from the balance equations defining π), the expected arrival and departure rates for calls of type r are equal.

We shall say that a bounded control $\boldsymbol{b}$ is *monotonic* if, for all r and for all $\boldsymbol{n}$,

$$b_r(\boldsymbol{n}) \text{ is increasing in } n_r \text{ (with } n_s \text{ fixed for all } s \neq r), \tag{27}$$

$$b_r(\boldsymbol{n}) \text{ is decreasing in } n_s \text{ (with } n_{s'} \text{ fixed for all } s' \neq s), \text{ for all } s \neq r. \tag{28}$$

Note that, depending on the network structure, this property is natural in many applications. For instance, for the structure of Example 2.2, it is possessed by all the fair-sharing controls, and also by any other reasonable Pareto efficient control. For more complex network structures, controls may be coupled to monotonic controls to establish stability results using the results given below. See also Example 3.1 below. We note further that a

related but somewhat different definition of monotonicity is used by Bonald and Proutière (2004).

For any monotonic control $\boldsymbol{b}$, and for each $S \subseteq R$ (including the case where S is the empty set $\emptyset$), define the function $\boldsymbol{b}^S \colon \mathbb{Z}_+^{|S|} \to \mathbb{R}_+^{|R|}$ by

$$b_r^S(\boldsymbol{n}_S) = \lim_{n_s \to \infty \, \forall \, s \notin S} b_r(\boldsymbol{n}), \qquad r \in S, \tag{29}$$

$$b_r^S(\boldsymbol{n}_S) = \lim_{n_r \to \infty} \lim_{n_s \to \infty \, \forall \, s \notin S \cup \{r\}} b_r(\boldsymbol{n}), \qquad r \notin S, \tag{30}$$

where $\boldsymbol{n}_S = (n_s, \, s \in S)$. Note that, by monotonicity, this function is well-defined. In particular, in (29), the order within $R \setminus S$ in which the limits are taken is irrelevant; however, in (30), the final limit to be taken must be that as $n_r \to \infty$.

For any monotonic control $\boldsymbol{b}$ and any S as above, we shall say that $\boldsymbol{b}^S$ is *stable* if the application of the control $(b_s^S, \, s \in S)$ to calls in the network of types $s \in S$ (with ν_s, μ_s, $s \in S$, as usual) yields a positive recurrent Markov process $\boldsymbol{n}^S(\cdot) = (n_s^S(\cdot), \, s \in S)$. (This $|S|$-dimensional process may be thought of as that which results when the number of calls of each type $r \notin S$ is infinite.) When $\boldsymbol{b}^S$ is stable we shall let π^S denote the stationary distribution of $\boldsymbol{n}^S(\cdot)$ (or, where necessary, the probability function of this distribution); we shall further define, for each $r \in R$,

$$\mathbf{E}_{\pi^S} b_r^S = \sum_{\boldsymbol{n}_S \in \mathbb{Z}_+^{|S|}} \pi^S(\boldsymbol{n}_S) b_r^S(\boldsymbol{n}_S) \tag{31}$$

to be the expected value of b_r^S under this distribution. In the case where S is the empty set $\emptyset$, we have that $\boldsymbol{b}^\emptyset = (b_r^\emptyset, \, r \in R)$ is a vector of constants. We make the natural convention that $\boldsymbol{b}^\emptyset$ is always stable; the distribution $\pi^\emptyset$ is concentrated on a single point, and we have $\mathbf{E}_{\pi^\emptyset} b_r^\emptyset = b_r^\emptyset$ for all $r \in R$.

Our main result of this section is Theorem 3.1 below. The first part is similar in spirit to results of Borovkov (1998, Chapter 8) for asymptotically spatially homogeneous Markov processes. However, the application of those results here would require that the right side of (30) is invariant under interchange of the limits in that expression, a condition which is not in general satisfied for our monotonic controls. Rather the monotonicity itself provides sufficient structure to obtain the results of Theorem 3.1.

Theorem 3.1. *Suppose that the control $\boldsymbol{b}$ is monotonic and that $S \subseteq R$ is such that $\boldsymbol{b}^S$ is stable.*

  (i) If $r \notin S$ is such that

$$\mathbf{E}_{\pi^S} b_r^S > \kappa_r, \tag{32}$$

    then $\boldsymbol{b}^{S \cup \{r\}}$ is stable.

  (ii) If $r \notin S$ is such that

$$\mathbf{E}_{\pi^S} b_r^S < \kappa_r, \tag{33}$$

    then $\boldsymbol{b}^{S \cup \{r\}}$ is unstable.

Remark 3.2. Given the stability of $\boldsymbol{b}^S$ for some $S \subset R$ (recall that, as already remarked, $\boldsymbol{b}^\emptyset$ is always stable), Theorem 3.1 gives criteria for determining the stability or otherwise of $\boldsymbol{b}^{S \cup \{r\}}$ for any $r \notin S$, except only in the case of equality in (32) or (33) (where the natural

conjecture is that $\boldsymbol{b}^{S\cup\{r\}}$ is unstable—see also the remarks at the end of Example 3.1). Recursive application of the theorem thus yields sufficient conditions both for the stability and the instability of monotonic controls. However, note that, for example in the case $R = \{1,2\}$, $\boldsymbol{b}^{\{1\}}$ and $\boldsymbol{b}^{\{2\}}$ may both be unstable, while $\boldsymbol{b} = \boldsymbol{b}^{\{1,2\}}$ is stable, as is the case for fair-sharing controls here. In such circumstances Theorem 3.1 does not settle the question of the stability of $\boldsymbol{b}$. Rather its primary application is to controls in which there is a sufficient hierarchy of prioritisation as to permit the recursive application of the first part of the theorem to at least establish the stability of $\boldsymbol{b}^S$ for $S = R\setminus\{r\}$ for some $r \in R$. The theorem then also (in general) settles the question of the stability of $\boldsymbol{b}$ itself. For an illustration of the application of the theorem, see Example 3.1 below.

Proof of Theorem 3.1. Since, for given $S \subset R$ and $r \notin S$, the stability of $\boldsymbol{b}^{S\cup\{r\}}$ corresponds to the positive recurrence of the Markov process $\boldsymbol{n}^{S\cup\{r\}}(\cdot)$ defined above (in which the number of calls of each type $s \notin S\cup\{r\}$ is effectively held at infinity), it is sufficient to prove the results (i) and (ii) in the case where $S = R\setminus\{r'\}$ for some r', and we henceforth assume this. (The primary advantage of doing so is that we avoid some unpleasant notational complexity.) We identify any $\boldsymbol{n} \in \mathbb{Z}_+^{|R|}$ with the pair $(\boldsymbol{n}_S, n_{r'})$ where $\boldsymbol{n}_S = (n_s,\, s \in S)$. Recall that then, for each such $\boldsymbol{n}_S$, we have $b_{r'}^S(\boldsymbol{n}_S) = \lim_{n_{r'}\to\infty} b_{r'}(\boldsymbol{n}_S, n_{r'})$.

Suppose first that the condition (32) holds. We require to show that $\boldsymbol{b}$ is stable. The underlying idea is that the monotonicity of $\boldsymbol{b}$ and stability of $\boldsymbol{b}^S$ ensure that the components $(n_s(\cdot),\, s \in S)$ of the process $\boldsymbol{n}(\cdot)$ become and remain small, and the condition (32) then ensures that, except in some finite region A, the remaining component $n_{r'}(\cdot)$ of this process is decreasing at a rate which is bounded away from zero; thus the process $\boldsymbol{n}(\cdot)$ spends, in the long term, a nonzero proportion of time within A.

It follows from (32) and the monotonicity of $\boldsymbol{b}$ that we can choose a finite set $A = \{\boldsymbol{n}\colon n_r \le \bar{n}_r,\, r \in R\} \subset \mathbb{Z}_+^{|R|}$ and a positive function $\bar{b}_{r'}$ on $\mathbb{Z}_+^{|S|}$ such that

$$\bar{b}_{r'} \text{ is decreasing in each of its arguments,} \tag{34}$$

$$\sum_{\boldsymbol{n}_S\in\mathbb{Z}_+^{R-1}} \pi^S(\boldsymbol{n}_S)\bar{b}_{r'}(\boldsymbol{n}_S) > \kappa_{r'}, \tag{35}$$

$$b_{r'}(\boldsymbol{n}) \ge \bar{b}_{r'}(\boldsymbol{n}_S) \qquad \text{for all } \boldsymbol{n} \notin A. \tag{36}$$

For example, given $\bar{n}_r,\, r \in R$, we may choose

$$\bar{b}_{r'}(\boldsymbol{n}_S) = \mathbf{I}(n_s \le \bar{n}_s,\, s \in S)\, b_{r'}(\boldsymbol{n}_S, \bar{n}_{r'}), \qquad \boldsymbol{n}_S \in \mathbb{Z}_+^{|S|}. \tag{37}$$

The condition (34) then follows from the monotonicity of $\boldsymbol{b}$; since also $\boldsymbol{b}$ is nonnegative the condition (36) follows trivially from (37), except for $\boldsymbol{n}$ such that $n_s \le \bar{n}_s$ for $s \in S$ and $n_{r'} > \bar{n}_{r'}$, in which case (36) again follows from the monotonicity of $\boldsymbol{b}$; finally the condition (35) follows from (32) and the monotone convergence theorem by choosing $\bar{n}_r$, $r \in R$, all sufficiently large, since, for any $\boldsymbol{n}_S$,

$$\lim_{\bar{n}_r\to\infty\,\forall r\in R} \mathbf{I}(n_s \le \bar{n}_s,\, s \in S)\, b_{r'}(\boldsymbol{n}_S, \bar{n}_{r'}) = \lim_{\bar{n}_{r'}\to\infty} b_{r'}(\boldsymbol{n}_S, \bar{n}_{r'}) = b_{r'}^S(\boldsymbol{n}_S).$$

It also follows from the monotonicity of the control $\boldsymbol{b}$ that we can couple the corresponding Markov process $\boldsymbol{n}(\cdot)$ to a process $\boldsymbol{n}^S(\cdot) = (n_s^S(\cdot),\, s \in S)$ on $\mathbb{Z}_+^{|S|}$ with control $\boldsymbol{b}^S$ in such

a way that $n_s(t) \leq n_s^S(t)$ for all $t > 0$ and for all $s \in S$. Since also the process $\boldsymbol{n}^S(\cdot)$ has stationary distribution π^S, it follows from (34) that, for the function $\bar{b}_{r'}$ defined above,

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t \bar{b}_{r'}(\boldsymbol{n}_S(u)) \, du \geq \lim_{t \to \infty} \frac{1}{t} \int_0^t \bar{b}_{r'}(\boldsymbol{n}^S(u)) \, du = \mathbf{E}_{\pi^S} \bar{b}_{r'}, \tag{38}$$

where $\boldsymbol{n}_S(\cdot) = (n_s(\cdot), s \in S)$ and where the final equality above follows by the ergodic theorem.

It now follows from Lemma 3.1 that, for some $M > 0$,

$$\kappa_{r'} \geq \limsup_{t \to \infty} \frac{1}{t} \int_0^t b_{r'}(\boldsymbol{n}(u)) \, du$$

$$\geq \limsup_{t \to \infty} \frac{1}{t} \int_0^t b_{r'}(\boldsymbol{n}(u)) \, \mathbf{I}((\boldsymbol{n}(u) \notin A) \, du$$

$$\geq \limsup_{t \to \infty} \frac{1}{t} \int_0^t \bar{b}_{r'}(\boldsymbol{n}_S(u)) \, \mathbf{I}((\boldsymbol{n}(u) \notin A) \, du \tag{39}$$

$$\geq \limsup_{t \to \infty} \frac{1}{t} \int_0^t \bar{b}_{r'}(\boldsymbol{n}_S(u)) \, du - \lim_{t \to \infty} \frac{M}{t} \int_0^t \mathbf{I}((\boldsymbol{n}(u) \in A) \, du \tag{40}$$

$$\geq \mathbf{E}_{\pi^S} \bar{b}_{r'} - \lim_{t \to \infty} \frac{M}{t} \int_0^t \mathbf{I}((\boldsymbol{n}(u) \in A) \, du, \tag{41}$$

where the inequality (39) follows from (36), the inequality (40) follows since, from (34), the function $\bar{b}_{r'}$ is necessarily bounded (note that, since A is finite, the limit in the final term in (40) always exists), and the final inequality (41) follows from (38). Thus, from (41) and (35),

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \mathbf{I}((\boldsymbol{n}(u) \in A) \, du > 0.$$

Since A is finite it now follows from the ergodic theorem that the Markov process $\boldsymbol{n}(\cdot)$ is positive recurrent and so $\boldsymbol{b}$ is stable.

Now suppose instead that the condition (33) holds. We show that the Markov process $\boldsymbol{n}(\cdot)$ corresponding to $\boldsymbol{b}$ is transient (and hence $\boldsymbol{b}$ is unstable). The underlying idea here is that whenever $n_{r'}(\cdot)$ is very large, the process $\boldsymbol{n}(\cdot)$ again behaves approximately as if it were controlled by $\boldsymbol{b}^S$, and thus, from (33), we may expect that $\lim_{t \to \infty} n_{r'}(t) = \infty$ a.s. To make this rigorous we again use the monotonicity of $\boldsymbol{b}$ to couple the process $\boldsymbol{n}(\cdot)$ to a process $\hat{\boldsymbol{n}}(\cdot)$ whose control is sufficiently close to that of $\boldsymbol{b}^S$ that we may show that $\lim_{t \to \infty} \hat{n}_{r'}(t) = \infty$ a.s., and for which the coupling ensures that also $\lim_{t \to \infty} n_{r'}(t) = \infty$ with strictly positive probability.

Given $\bar{n}_{r'} \in \mathbb{Z}_+$ (fixed, to be chosen later), define a Markov process $\hat{\boldsymbol{n}}(\cdot) = (\hat{n}_r(\cdot), r \in R)$ as follows: for each $s \in S$, the component process $\hat{n}_s(\cdot)$ has state space $\mathbb{Z}_+$ as usual, while $\hat{n}_{r'}(\cdot)$ has state space $\mathbb{Z}$; for each $r \in R$, transitions $n_r \to n_r + 1$ occur at rate ν_r as usual, while transitions $n_r \to n_r - 1$ occur at rate $\mu_r \hat{b}_r(\boldsymbol{n})$, where

$$\hat{b}_s(\boldsymbol{n}) = b_s(\boldsymbol{n}_S, \bar{n}_{r'}), \qquad s \in S, \tag{42}$$

$$\hat{b}_{r'}(\boldsymbol{n}) = b_{r'}^S(\boldsymbol{n}_S). \tag{43}$$

Observe that the process $\hat{\boldsymbol{n}}(\cdot)$ has uniformly bounded transition rates which are independent of $n_{r'} \in \mathbb{Z}$. Suppose that

$$\boldsymbol{n}(0) = \hat{\boldsymbol{n}}(0), \qquad n_{r'}(0) > \bar{n}_{r'}. \tag{44}$$

Define the random time $T = \min\{t > 0: n_{r'}(t) < \bar{n}_{r'}\}$. It follows from the monotonicity of $\boldsymbol{b}$ that, for all $\boldsymbol{n}$ such that $n_{r'} \geq \bar{n}_{r'}$,

$$\hat{b}_s(\boldsymbol{n}) \geq b_s(\boldsymbol{n}) \geq b_s^S(\boldsymbol{n}_S), \qquad s \in S. \tag{45}$$

and hence that we may couple the processes $\hat{\boldsymbol{n}}(\cdot)$, $\boldsymbol{n}(\cdot)$ and the $|S|$-dimensional process $\boldsymbol{n}^S(\cdot)$ with control $(b_s^S, s \in S)$ in such a way that

$$\hat{n}_s(t) \leq n_s^S(t), \qquad \text{for all } s \in S \text{ and for all } t \geq 0 \tag{46}$$

and

$$\hat{n}_s(t) \leq n_s(t), \qquad \text{for all } s \in S \text{ and for all } 0 \leq t \leq T. \tag{47}$$

Since $\boldsymbol{n}^S(\cdot)$ is assumed positive recurrent with stationary distribution π^S, it follows from (45) or (46) that $\hat{\boldsymbol{n}}_S(\cdot) = (\hat{n}_s(\cdot), s \in S)$ is similarly positive recurrent with stationary distribution $\hat{\pi}_S$ say. Further, as $\bar{n}_{r'} \to \infty$, the control $\hat{\boldsymbol{b}}$ converges pointwise in each of its components to $\boldsymbol{b}^S$. Hence elementary arguments (e.g. consideration of the times of return to $\boldsymbol{0}$ of the process $\boldsymbol{n}^S(\cdot)$, coupled with the use of (46) and the ergodic theorem) show that, again as $\bar{n}_{r'} \to \infty$, $\hat{\pi}_S$ converges in distribution to π^S. Since also $b_{r'}^S$ is bounded, it now follows from (33) that we may choose the constant $\bar{n}_{r'}$ sufficiently large that

$$\mathbf{E}_{\hat{\pi}_S} b_{r'}^S < \kappa_{r'}, \tag{48}$$

(where, analogously to (31), $\mathbf{E}_{\hat{\pi}_S} b_{r'}^S$ is the expectation of $b_{r'}^S$ with respect to $\hat{\pi}_S$).

It further follows from the monotonicity of $\boldsymbol{b}$ and from (47) that, for all $0 \leq t \leq T$,

$$\hat{b}_{r'}(\hat{\boldsymbol{n}}(t)) = b_{r'}^S(\hat{\boldsymbol{n}}_S(t))$$
$$\geq b_{r'}^S(\boldsymbol{n}_S(t)) \tag{49}$$
$$\geq b_{r'}(\boldsymbol{n}(t)), \tag{50}$$

where (49) follows from (47) and the monotonicity of $\boldsymbol{b}$ while (50) follows from the definition of $\boldsymbol{b}^S$ and, again, the monotonicity of $\boldsymbol{b}$. It now follows from (50) that we may couple also the components r' of the processes $\hat{\boldsymbol{n}}(\cdot)$ and $\boldsymbol{n}(\cdot)$ in such a way that

$$\hat{n}_{r'}(t) \leq n_{r'}(t), \qquad \text{for all } 0 \leq t \leq T. \tag{51}$$

As noted above, the process $\hat{\boldsymbol{n}}_S(\cdot) = (\hat{n}_s(\cdot), s \in S)$ has stationary distribution $\hat{\pi}_S$, while the process $\hat{n}_{r'}(\cdot)$ may be viewed as a Markov additive process modulated by the remaining components $\hat{\boldsymbol{n}}_S(\cdot)$ of $\hat{\boldsymbol{n}}(\cdot)$. From (48), the expectation of the increments of $\hat{n}_{r'}(\cdot)$ between those times at which $\hat{\boldsymbol{n}}(\cdot)$ returns to any fixed state is strictly positive. It follows from the standard theory of Markov additive processes that $\lim_{t \to \infty} \hat{n}_{r'}(t) = \infty$ a.s., and further that, under the condition (44),

$$\mathbf{P}(\hat{n}_{r'}(t) \geq \bar{n}_{r'} \text{ for all } t \geq 0, \lim_{t \to \infty} \hat{n}_{r'}(t) = \infty) > 0,$$

and hence, from (51), that also

$$\mathbf{P}(n_{r'}(t) \geq \bar{n}_{r'} \text{ for all } t \geq 0, \lim_{t \to \infty} n_{r'}(t) = \infty) > 0.$$

Hence the process $\boldsymbol{n}(\cdot)$ is transient as required. $\qquad\square$

We illustrate the use of the above result with a simple example.

Example 3.1. Consider again the network of Example 1.2, in which $R = J = 3$. As previously observed a necessary and sufficient condition for the existence of *some* stable control is given by $2\nu < c$. Further, if $3\nu < c$, then Proposition 2.1 with the Lyapunov function f given by $f(\boldsymbol{n}) = \sum_{r=1}^{3} n_r$ shows that *any* Pareto efficient control is stable. Suppose now that $2\nu < c$ and that the Pareto efficient control $\boldsymbol{b}$ is such that, for $r = 1, 2$, $b_r(\boldsymbol{n})$ is independent of n_3 and

$$b_1(\boldsymbol{n}) + b_2(\boldsymbol{n}) = c \qquad \text{for all } \boldsymbol{n} \text{ such that } \max(n_1, n_2) > 0. \tag{52}$$

Thus in particular calls of types 1 and 2 collectively have complete priority over calls of type 3. Although we do not in this example require any further monotonicity conditions on $\boldsymbol{b}$, it follows from the requirement of Pareto efficiency that the control $\boldsymbol{b}^{\{1,2\}} \colon \mathbb{Z}_+^2 \to \mathbb{R}_+^3$ is well-defined as before, being obtained from $\boldsymbol{b}$ by letting $n_3 \to \infty$. It follows from (52) that the condition $2\nu < c$ is necessary and sufficient to ensure that $\boldsymbol{b}^{\{1,2\}}$ is stable. We use (a slight modification of) Theorem 3.1 to investigate the stability of $\boldsymbol{b}$. The stationary distribution $\pi^{\{1,2\}}$ on $\mathbb{Z}_+^2$ induced by $\boldsymbol{b}^{\{1,2\}}$ is here just that of the process $(n_1(\cdot), n_2(\cdot))$. Further, since, from (52), $n_1(\cdot) + n_2(\cdot)$ is Markov, with a stationary distribution which is geometric and independent of any more detailed specification of $\boldsymbol{b}^{\{1,2\}}$, it follows that

$$\pi^{\{1,2\}}(0, 0) = 1 - \frac{2\nu}{c}. \tag{53}$$

It follows from (2), (52) and the Pareto efficiency of $\boldsymbol{b}$ that

$$b_3^{\{1,2\}}(0, 0) = c, \qquad b_3^{\{1,2\}}(n_1, 0) = 0 \quad \text{for all } n_1 \geq 1, \qquad b_3^{\{1,2\}}(0, n_2) = 0 \quad \text{for all } n_2 \geq 1 \tag{54}$$

We thus have that

$$\mathbf{E}_{\pi^{\{1,2\}}} b_3^{\{1,2\}} = \sum_{(n_1,n_2)\in\mathbb{Z}_+^2} \pi^{\{1,2\}}(n_1, n_2) b_3^{\{1,2\}}(n_1, n_2)$$
$$\geq c - 2\nu, \tag{55}$$

with equality if and only if $b_3^{\{1,2\}}(n_1, n_2) = 0$ for all (n_1, n_2) such that $\min(n_1, n_2) \geq 1$. But this latter condition holds if and only if, for all (n_1, n_2) such that $\min(n_1, n_2) \geq 1$, we have $\min\big(b_1^{\{1,2\}}(n_1, n_2), b_2^{\{1,2\}}(n_1, n_2)\big) = 0$, i.e. in the case of the control considered in Example 1.2 in which maximum resource is always allocated to calls of one type, and in which we have already observed that we have stability if and only if $3\nu < c$. Otherwise we have strict inequality in (55).

Now note that, although $\boldsymbol{b}$ does not here satisfy all the conditions for monotonicity given earlier, the assumption that $b_1(\boldsymbol{n})$ and $b_2(\boldsymbol{n})$ are independent of n_3 ensures that Theorem 3.1 continues to apply, indeed in a slightly improved form, to show that the condition $\mathbf{E}_{\pi^{\{1,2\}}} b_3^{\{1,2\}} > \nu$ is necessary and sufficient for the stability of $\boldsymbol{b}$. (For the sufficiency, note that the proof of part (i) of the theorem, with $S = \{1, 2\}$ and $r' = 3$, goes through as before, except that the coupling between $\boldsymbol{n}(\cdot)$ and process $\boldsymbol{n}^S(\cdot)$ is now obtained with equality, and so we no longer require the condition (34) in order to obtain (38). Similar obvious simplifications apply to the proof of part (ii), which here becomes a fairly standard argument and in particular delivers null recurrence—and hence instability—in the case $\mathbf{E}_{\pi^{\{1,2\}}} b_3^{\{1,2\}} = \nu$.)

15

Suppose now that c and $\boldsymbol{b}$ are held fixed and that ν is allowed to vary. The obvious coupling argument shows that if $\boldsymbol{b}$ is stable for any ν then it is also stable for any $\nu' < \nu$. The above adaptation of Theorem 3.1, together with (55), shows that there is some critical parameter λ (depending on the detailed specification of $\boldsymbol{b}^{\{1,2\}}$ and hence $\pi^{\{1,2\}}$) such that $1/3 \leq \lambda \leq 1/2$ and that $\boldsymbol{b}$ is stable if $\nu < \lambda c$ and unstable if $\nu > \lambda c$. For the control of Example 1.2 we already know that $\lambda = 1/3$; otherwise for the case $\nu = c/3$ we have strict inequality in (55) and hence stability; simple continuity arguments now give $\lambda > 1/3$ in this case.

Acknowledgement

References

[1] Asmussen, S. (2003). *Applied Probability and Queues.* Springer, New York.

[2] Bonald, T. and Massoulié, L. (2001). Impact of fairness on Internet performance. In *Proceedings of ACM SIGMETRICS 2001.* Cambridge, Massachusetts.

[3] Bonald, T. and Proutière, A. (2004). On Stochastic Bounds for Monotonic Processor Sharing Networks *Queueing Systems*, **47**, 81–106.

[4] Borovkov, A. A. (1998). *Ergodicity and Stability of Stochastic Processes.* Wiley, Chichester.

[5] Dai, J. (1995). On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Annals of Applied Probability*, **5**, 49–77.

[6] de Veciana, G., Lee, T.-J. and Konstantopoulos, T. (2001). Stability and performance analysis of networks supporting elastic services. *IEEE/ACM Trans. on Networking*, **9**, 2–14.

[7] Kelly, F. P., Maulloo, A. K. and Tan, D. K. H. (1998). Rate control in communication networks: shadow prices, proportional fairness and stability. *J. Operational Research Society*, **49**, 237–252.

[8] Kelly, F. P. and Williams, R. J. (2004). Fluid model for bandwidth sharing. *Annals of Applied Probability*, **14**, 1055–1083.

[9] Mo, J. and Walrand, J. (2000). Fair end-to-end window-based congestion control. *IEEE/ACM Trans. on Networking*, **8**, 556–567.

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