Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Assessing the efficiency of resource allocations in bandwidth-sharing networks

I.M. Verloop, R. Núñez Queija

REPORT PNA-E0702 JULY 2007

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Assessing the efficiency of resource allocations in bandwidth-sharing networks

ABSTRACT
Resource allocation in bandwidth-sharing networks is inherently complex: The distributed nature of resource allocation management prohibits global coordination for efficiency, i.e., aiming at full resource usage at all times. In addition, it is well recognized that resource efficiency may be conflicting with other critical performance measures such as flow delay. Without a notion of optimal (or “near-optimal”) behavior, the performance of resource allocation schemes can not be assessed properly. In previous work, we showed that optimal workload-based (or queue-length based) strategies have certain structural properties (they are characterized by so-called switching curves), but are too complex in general to be determined exactly. In addition, numerically determining the optimal strategy often requires excessive computational effort. This raises the need for simpler strategies with “near-optimal” behavior that can serve as a sensible bench-mark to test resource allocation strategies. We focus on flows traversing the network, sharing the resources on their common path with (independently generated) cross-traffic. Assuming exponentially distributed flow sizes, we show that in many scenarios optimizing the "drain time" under a fluid scaling gives a simple linear switching strategy that accurately approximates the optimal strategy. When two nodes on the flow path are equally congested, however, the fluid scaling is not appropriate, and the corresponding strategy may not even ensure stability. In such cases we show that the appropriate scaling for efficient workload-based allocations follows a square-root law. Armed with these, we then assess the potential gain that any sophisticated strategy can achieve over standard alpha-fair strategies, which are representations of common distributed allocation schemes, and confirm that alpha-fair strategies perform excellently among non-anticipating policies. In particular, we can approximate the optimal policy with a weighted alpha-fair strategy.

2000 Mathematics Subject Classification: 68M20, 60K25, 90B18
Keywords and Phrases: network efficiency; fair resource allocation; bandwidth-sharing networks; distributed resource management; flow delays; size-based scheduling; switching strategies; linear network; fluid scaling; square-root scaling; weighted alpha-fair strategies; proportional fair
Note: This research is financially supported by The Netherlands Organization for Scientific Research (NWO).

Assessing the efficiency of resource allocations in bandwidth-sharing networks

Maaike Verloop[⋆] , Rudesindo Núñez-Queija[⋆,‡]

[⋆]CWI, The Netherlands
[‡]TNO Information and Communication Technology, The Netherlands

Abstract

Resource allocation in bandwidth-sharing networks is inherently complex: The distributed nature of resource allocation management prohibits global coordination for efficiency, i.e., aiming at full resource usage at all times. In addition, it is well recognized that resource efficiency may be conflicting with other critical performance measures such as flow delay. Without a notion of optimal (or "near-optimal") behavior, the performance of resource allocation schemes can not be assessed properly. In previous work, we showed that optimal workload-based (or queue-length based) strategies have certain structural properties (they are characterized by so-called switching curves), but are too complex in general to be determined exactly. In addition, numerically determining the optimal strategy often requires excessive computational effort. This raises the need for simpler strategies with "near-optimal" behavior that can serve as a sensible bench-mark to test resource allocation strategies.

We focus on flows traversing the network, sharing the resources on their common path with (independently generated) cross-traffic. Assuming exponentially distributed flow sizes, we show that in many scenarios optimizing the "drain time" under a fluid scaling gives a simple linear switching strategy that accurately approximates the optimal strategy. When two nodes on the flow path are equally congested, however, the fluid scaling is not appropriate, and the corresponding strategy may not even ensure stability. In such cases we show that the appropriate scaling for efficient workload-based allocations follows a square-root law. Armed with these, we then assess the potential gain that any sophisticated strategy can achieve over standard α-fair strategies, which are representations of common distributed allocation schemes, and confirm that α-fair strategies perform excellently among non-anticipating policies. In particular, we can approximate the optimal policy with a weighted α-fair strategy.

Key words: network efficiency, fair resource allocation, bandwidth-sharing networks, distributed resource management, flow delays, size-based scheduling, switching strategies, linear network, fluid scaling, square-root scaling, weighted alpha-fair strategies, proportional fair.

1 Introduction

Document transport in the Internet is regulated by distributed packet-based congestion control mechanisms, usually relying on one of the many incarnations of TCP (Transmission Control Protocol). By dividing a document into smaller parts (packets) the entire file is not transported as a single unity. Instead, parts of it reside at different nodes along the transmission path. The "instantaneous transfer rate" of the entire document can be thought of as being equal to the minimum transfer rate along the entire path. As TCP is a distributed mechanism, individual flows dynamically adjust to congestion along the path and the actual transmission rate for any flow fluctuates over time. Over somewhat longer time scales, these fluctuations average out and the effective transfer rate may be determined as a time average. The dynamics of the transfer

rates for TCP and similar mechanisms have been extensively studied under various mathematical approaches. Using a fluid representation for packets flowing from their origins to destinations, a variety of performance measures such as convergence, fairness and random effects have been investigated [15, 4, 3]. The class of α-fair allocations were shown to capture a wide range of distributed allocation mechanisms such as TCP, the proportional fair allocation and the max-min fair allocation [19]. Overviews of mathematical models for TCP can be found in [14, 26].

A common assumption in the majority of these papers is that the number of active data flows is fixed. In order to study the dynamics in the number of flows, a common approach is to assume that TCP dynamics occur on a much faster time scale, so that the rate allocation can be assumed to adapt instantly after a change in the number of flows. The flow-level performance of α-fair allocations was first studied in [7] and more recently, including several other allocation mechanisms, in [8]. A further powerful approach in studying the complex dynamics is by investigating different asymptotic regimes, see for example the large-network scaling in [13] or the heavy-traffic scaling in [16].

It is worth emphasizing that all these models essentially differ from traditional queueing networks, in that each job (i.e., document transfer) can be seen to claim resources from several nodes simultaneously instead of "hopping" from node to node. The effect of simultaneous resource sharing will be even more pronounced in the future with increasingly popular applications such as peer-to-peer overlays, which generate extremely long transfers.

Performance analysis of bandwidth-sharing networks is crucially different from and arguably more difficult than traditional queueing networks developed in the 60s for computer communication networks. As a consequence, closed-form analysis is (currently) elusive, except in a few specific cases. This significantly complicates the task of designing efficient allocation mechanisms. As a measure for efficiency we generically choose the (average) number of active flows. In previous work, it was shown that blindly applying size-based scheduling strategies, which are known to have certain optimality properties when there is a single resource [25, 22], are not optimal in general and may not even guarantee maximum stability [30]. Such strategies can therefore not serve as a sensible benchmark to compare the performance of other – implementable – scheduling strategies (for example α-fair allocations). For that reason we studied the structure of optimal (size-oblivious) strategies in a linear network, which can be shown to be characterized by certain "switching regimes" [31]: as the numbers of flows vary, the optimal allocation dynamically switches between several priority rules. In certain specific cases, the optimal switching allocation may degenerate to a static priority rule. Optimal policies may in general not be distributed, or require knowledge of the statistical properties of the flows and, thus, not be likely to serve for actual implementation. However, as mentioned above, such policies do provide useful benchmarks to compare against for any implementable strategy and estimate the scope left for further improvement.

Although this sheds some light on the structure of the optimal policies, an exact characterization of the switching curves is in general not possible. To gain a better understanding of the functional form of the optimal switching curves, we therefore set out to study these in asymptotic regimes. In [29] this was done for a highly loaded network. In this paper, we study the underloaded case after scaling the state space. The approach is similar to that developed for re-entrant lines, see for example [17] and [24]. The dynamics of re-entrant lines and bandwidth-sharing networks as considered here are however somewhat different. For stability considerations under a unifying framework, see [11].

Using a linear scaling for both the state space and time, leads to simple linear switching curves. Often this approach indeed finds close approximations to the optimal policies. For some scenarios however, applying a linear scaling may result in a policy that not only is far from optimal, it may in fact be unstable. In that case the diffusion scaling leads to policies that approximate the optimal policy.

Through numerical experiments we include comparisons of the optimal policies (when numerically feasible), the class of α-fair allocations (with specific attention to the proportional fair policy and TCP) and simple strategies characterized by either linear, square-root or constant switching curves. We then check whether a weighted α-fair allocation can approximate optimal performance by choosing appropriate weights.
The remainder of our paper is organized as follows. Section 2 describes the model and presents several preliminary results. We establish cases under which strict priority rules achieve optimality, describe the general structure of the optimal policy and discuss new stability results that are useful in the subsequent analysis. In Section 3 we derive optimal strategies under a linear scaling approach that are close to optimality in many cases. This scaling turns out to be inappropriate when two nodes along a flow path have the same load. In Section 4 we show that in such cases a square-root scaling is the right choice and we discuss how this can be understood from the Central Limit Theorem. Section 5 contains numerical evidence that the scaling approach yields sensible benchmarks. It also shows that α-fair strategies, and proportional fair allocations in particular, perform well in general and can even approach the optimal policy when choosing the best parameters. We conclude the paper with a short summary and ideas for on-going research in Section 6.

2 Model description and preliminary results

Although our approach is applicable to more general settings, for conciseness we discuss our ideas in the simplest possible setting of two nodes, see Figure 1.
We consider a linear network with 2 nodes. There are three traffic classes, where class i requires service at node i only, $i = 1, 2$, while class 0 requires service at both nodes simultaneously. Some of the results mentioned here translate directly to linear networks with more than 2 nodes. Class-i users arrive as a Poisson process of rate λ_i, and have generally distributed service requirements, B_i, with mean $1/\mu_i$. Let the traffic load of class i be $\rho_i := \lambda_i \mathbb{E}(B_i)$, thus the load at node i is $\rho_0 + \rho_i$. The conditions $\rho_0 + \rho_i < 1$, $i = 1, 2$, are necessary but in general not sufficient for stability [30].

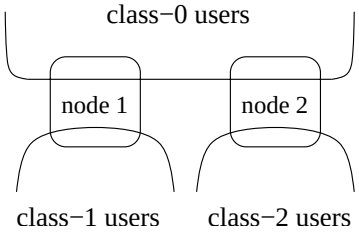


Figure 1: Linear network with 2 nodes.

We denote by $\bar{\Pi}$ the class of all non-anticipating (possibly preemptive) policies. For a given policy $\pi \in \bar{\Pi}$, denote by $N_i^\pi(t)$ the number of class-i users at time t. Define N_i^π as a random variable with the corresponding steady-state distribution (when it exists).
The central objective is to minimize the mean total number of users in the system. Because of Little's law, this is equivalent to minimizing the mean overall sojourn time.

2.1 Priority rules & optimality

For completeness we first briefly discuss scenarios that lead to optimal strict priority rules. Recall that in single-server multi-class systems with exponentially distributed service requirements with

mean $1/\mu_i$, the μ-rule, which amounts to giving priority to the users with the highest service rate, μ_i, is known to stochastically minimize the number of users. The rationale behind this rule is that it maximizes the output rate at all times. In our network context, besides trying to maximize the total output rate of the system, we must take into account that when serving class 1 while class 0 is present and class 2 is empty, leaves node 2 under-utilized. In general, there can be a trade-off between maximizing the output rate and using the full capacity in every node whenever there is a backlog at that node. When B_i is exponentially distributed such that $\mu_0 \geq \mu_i$, for both $i = 1, 2$, these two objectives are not conflicting and the policy that stochastically minimizes the total number of users at every point in time among all non-anticipating policies degenerates to a static priority rule. This is summarized in the next two propositions, which extend to linear networks with more than two nodes as well [31]. In [29] these results are extended to heavy traffic scenarios with general service requirements.

Proposition 2.1 *Denote by π^* the policy that gives preemptive priority to class 0 whenever it is backlogged. Assume B_i is exponentially distributed with mean $1/\mu_i$ and $\mu_1 + \mu_2 \leq \mu_0$. Then π^* stochastically minimizes the total number of users among all policies in $\bar{\Pi}$.*

Proposition 2.2 *Denote by π^{**} the policy that serves classes $i = 1, 2$, whenever they are both backlogged. Otherwise class 0 is served. When class 0 is non-backlogged, all other classes with a backlog are served simultaneously. Assume B_i is exponentially distributed with mean $1/\mu_i$ and $\mu_1+\mu_2 \geq \mu_0 \geq \max\{\mu_1, \mu_2\}$. Then π^{**} stochastically minimizes the total number of users among all policies in $\bar{\Pi}$.*

If the mean size of class-0 users increases beyond that of at least one of the two other classes, i.e. $\mu_0 < \mu_i$ for at least one $i = 1, 2$, no strict priority rule is optimal. It may still be better to sometimes serve class 0 even if that does not maximize the departure rate in the short run. Doing so, may create the potential to serve classes 1 and 2 simultaneously in the future and therefore offer a higher degree of parallelism. Hence as the number of users varies, the system will dynamically switch between several priority rules. The next section provides the general structure of the optimal policy.

2.2 General structure of the optimal policy

We focus on the uncovered case, that is exponential service requirements with $\mu_0 < \mu_i$ for at least one $i = 1, 2$. As may be expected, when there are users of both classes 1 and 2 present, serving them will be optimal, since $\mu_0 \leq \mu_1 + \mu_2$ [31]. When there are only users of classes 0 and 1 present (no class-2 users) and $\mu_1 < \mu_0$, serving class 0 seems appropriate, since it uses the full capacity in both nodes and it maximizes the total output rate. However, when $\mu_0 < \mu_1$, there is no obvious choice: Serving class 0 is work-conserving, but the total output rate of the system is not maximized. In contrast, serving class 1 will maximize the total output rate, but leaves node 2 unused. It is easy to see that the latter can indeed lead to unnecessary instability. For example, if $\mu_i > \mu_0$ for both $i = 1, 2$, then giving priority to classes 1 and 2 myopically maximizes the total departure rate, but is unstable when $\rho_0 > (1 - \rho_1)(1 - \rho_2)$.
Since a stochastically optimal policy may in general not exist, we focus on the average-optimal policy instead, i.e., the policy that minimizes $\mathbb{E}(N_0^\pi + N_1^\pi + N_2^\pi)$ over all policies $\pi \in \bar{\Pi}$. The next proposition states that the optimal policy can be characterized by a switching curve that determines which class should be served. The proof can be found in [31].

Proposition 2.3 *Assume the service requirements are exponentially distributed with $\mu_1 + \mu_2 > \mu_0$. If both classes 1 and 2 are non-empty, then the expected average-optimal stationary policy serves these classes simultaneously. When class 3-i is empty, $i = 1, 2$, class 0 is served if $\mu_0 \geq \mu_i$,*

otherwise the optimal policy is characterized by a switching curve $h_i(\cdot)$, i.e. class 0 is served if and only if $N_i(t) < h_i(N_0(t))$.

Proposition 2.3 determines the structure of the optimal policy, but does not explicitly characterize the optimal switching curve. To gain some further understanding, let us compare two different switching policies, say with switching curves $(h_1(N_0), h_2(N_0))$ and $(g_1(N_0), g_2(N_0))$, while $h_i(N_0) \leq g_i(N_0)$ for all N_0, $i = 1, 2$. Clearly, in the short run, a lower switching curve is better when $\mu_0 < \mu_1$, since the output rate will increase for some states (and remain the same for all other states). In the long run, however, a higher switching curve may actually pay off: when starting in the same state, a higher curve empties the system faster, see Lemma 1 below, and has therefore less strict stability conditions, see Corollary 2.4.

Lemma 1 *For a given switching policy with switching curves $h_1(n_0)$ and $h_2(n_0)$, denote by $W_j^h(t)$ the workload of class j at time t. Let $h_i(n_0) \leq g_i(n_0)$ for all n_0, $i = 1, 2$. If $W_0^g(0) \leq W_0^h(0)$ and $W_0^g(0) + W_i^g(0) \leq W_0^h(0) + W_i^h(0)$, for $i = 1, 2$, then*

$$W_0^g(t) \leq_{st} W_0^h(t) \tag{1}$$

$$W_0^g(t) + W_i^g(t) \leq_{st} W_0^h(t) + W_i^h(t), \quad \text{for } i = 1, 2, \tag{2}$$

for all $t \geq 0$. (The symbol $\leq_{st}$ denotes the usual stochastic ordering.)

The proof can be found in Appendix A.

Corollary 2.4 *Let $h_i(n_0) \leq g_i(n_0)$ for all n_0, $i = 1, 2$. If the system is stable under the policy with switching curves $h_i(n_0)$, it is also stable under the policy with switching curves $g_i(n_0)$.*

The switching curve needs to find the right balance between these short and long run effects. In the remainder of the paper, we try to find the optimal switching curve under an appropriate scaling of the state space.

2.3 Preliminary stability results

Later in the paper it will be convenient to obtain the stability condition for the policy π^{***}. It is defined as the policy that gives preemptive priority to class 2. When class 2 is empty, class 0 receives preemptive priority. Class 1 is only served whenever there is capacity left unused. This strict priority rule has switching curves $h_1(n_0) = \infty$ and $h_2(n_0) = 0$. Stability is still ensured for general service requirements, as is stated in the following lemma. In fact, it states stability for more general strategies that are work conserving in node 2. The proof is in Appendix B. It essentially uses that the behavior of classes 0 and 2 are autonomously determined by the dynamics within node 2. In order for class 1 to be unstable two things are necessary: the work of class 1 must grow unboundedly *and* for a non-negligible portion of time, class 2 is served while class 1 is not present. Obviously, these two things can not both be true.

Lemma 2 *Any non-idling policy which gives strict priority to class 0 over class 1 when class 2 is empty, is stable under the standard conditions: $\rho_0 + \rho_i < 1$, $i = 1, 2$. In particular this holds for policy π^{***}.*

Let us denote the workloads and queue lengths under π^{***} by W_i^{***} and N_i^{***}. For general distributions, we can determine the mean workload for class 2 from the Pollaczek-Khintchine formula: $\mathbb{E}(W_2^{***}) = \frac{\lambda_2 \mathbb{E}({B_2}^2)}{2(1-\rho_2)}$. Class 0 sees its service being interrupted by busy periods of class 2 so that [27]:

$$\mathbb{E}(W_0^{***}) = \frac{\lambda_0 \mathbb{E}({B_0}^2) + \lambda_2 \mathbb{E}({B_2}^2)}{2(1-\rho_0-\rho_2)} - \frac{\lambda_2 \mathbb{E}({B_2}^2)}{2(1-\rho_2)}.$$

For exponential service requirements, the mean queue length is obtained from $\mathbb{E}(N_i^{***}) = \mu_i \mathbb{E}(W_i^{***})$. For class 1 there are no expressions available for the mean workload and the mean queue length. Determining these requires solving a boundary value problem [10], as is the case for policy π^{**} as well.

3 Linear scaling

In Section 2.2 we established the existence of a switching curve for exponential service requirements, however a parametric characterization of the curve could not be given. In this section we will scale the system linearly and derive for the so obtained fluid model the optimal policy and, in contrast to the stochastic model, obtain an exact expression for the optimal switching curve.

We consider a sequence of systems indexed by a superscript n. The workload and the number of class-i users in the n-th system at time t are denoted by $W_i^n(t)$ and $N_i^n(t)$ respectively. The initial queue length depends on n such that $\lim_{n\to\infty} \frac{1}{n} N_i^n(0) = a_i$. We will be interested in the fluid limits, where time is also scaled linearly:

$$\lim_{n\to\infty} \frac{N_i^n(nt)}{n} =: n_i(t) \quad \text{and} \quad \lim_{n\to\infty} \frac{W_i^n(nt)}{n} =: w_i(t).$$

Note that $w_i(t) = \frac{n_i(t)}{\mu_i}$, as explained in Remark 4.1. We refer to [9, 23] for further details on fluid limits.

Denote by $S_i(n(t-\Delta), nt) \geq 0$ the total capacity that is allocated to class i during the time interval $(n(t-\Delta), nt)$. Restricting to strategies that have proper limits $\lim_{n\to\infty} \frac{1}{n} S_i(n(t-\Delta), nt) \equiv s_i(t-\Delta, t)$ is reasonable from practical considerations since it only excludes hysteretic control strategies. Similarly, we may assume that $s_i(t) := \lim_{\Delta\to 0} \frac{1}{\Delta} s_i(t-\Delta, t)$ is well defined as well (with probability 1). Naturally, $S_0(n(t-\Delta), nt) + S_i(n(t-\Delta), nt) \leq n\Delta$, hence $s_0(t) + s_i(t) \leq 1$. From $W_i^n(nt) = W_i^n(n(t-\Delta)) + A_i(n(t-\Delta), nt) - S_i(n(t-\Delta), nt)$, we obtain: $w_i(t) - w_i(t-\Delta) = \rho_i \Delta - s_i(t-\Delta, t)$. After dividing by Δ and letting $\Delta \to 0$ we conclude that the fluid processes w_i are described by the following differential equations:

$$\begin{aligned} \frac{\mathrm{d}w_i}{\mathrm{d}t}(t) &= \rho_i - s_i(t), \text{ for } i = 0, 1, 2, \\ w_i(t) &\geq 0, \text{ for } i = 0, 1, 2, \\ s_0(t) + s_i(t) &\leq 1, \text{ for } i = 1, 2, \\ s_i(t) &\geq 0, \text{ for } i = 0, 1, 2. \end{aligned}$$

In principle, $s_i(t)$ may be a random process. However, since the input processes have been replaced by their averages, there is no gain in considering stochastic allocations $s_i(t)$. Therefore, the evolution of the fluid model involves no randomness, which renders it more tractable. We now proceed to derive the optimal clearing policy for the fluid model, starting from any initial state. Let Π be the set of all policies that satisfy $s_0(t) + s_i(t) \leq 1$, for $i = 1, 2$, $s_j(t) \leq \rho_j$ when $n_j(t) = 0$ and $s_j(t) \geq 0$ for $j = 0, 1, 2$. We use the following two definitions for an optimal policy.

- Policy $\overline{\pi}$ is called path-wise optimal if $\sum_{i=0}^2 n_i^{\overline{\pi}}(t) \leq \sum_{i=0}^2 n_i^{\pi}(t)$ for all $t \geq 0$, $\pi \in \Pi$.

Path-wise optimal policies do not necessarily exist, in which case we use the following criterion.

- Policy $\overline{\pi}$ is called average-optimal if $\int_0^T \sum_{i=0}^2 n_i^{\overline{\pi}}(t)\mathrm{d}t \leq \int_0^T \sum_{i=0}^2 n_i^{\pi}(t)\mathrm{d}t$ for all $\pi \in \Pi$, with T such that the system is empty at time T.

We see that a path-wise optimal policy, if it exists, is automatically average-optimal.

Remark 3.1 *Note that the fluid model retains the non-work-conserving property of the original model. In node i, exactly $1 - s_i(t)$ is left for class 0, so that $s_0(t) \leq 1 - s_i(t)$ for $i = 1, 2$. For an optimal strategy, we have $s_0(t) = 1 - s_i(t)$ for at least one i, but not necessarily for both. It is therefore possible that capacity is lost in one of the nodes.*

We now set out to determine the optimal policies for the fluid model, distinguishing between path-wise and average optimality. Without loss of generality we will assume that $\rho_1 \leq \rho_2$ and $\rho_0 + \rho_2 < 1$. The latter condition is needed for stability and guarantees that the fluid model will empty (and then remains empty if controlled optimally).

3.1 Path-wise optimal policies

Whenever the stochastic model allowed for stochastic optimization, i.e., in the cases that there was no conflict between maximizing the output rate and fully using all resources, one would expect that the fluid model allows for a path-wise optimal policy, which is confirmed by the next proposition.

Proposition 3.2 *Assume $\rho_1 \leq \rho_2$ and $\rho_0 + \rho_2 < 1$. A path-wise optimal policy for the fluid model can be found in the following scenarios. In all cases $s_i = 1 - s_0$ if $n_i > 0$ and $s_i = \min(\rho_i, 1 - s_0)$ if $n_i = 0$, for $i = 1, 2$.*
If $\mu_1 + \mu_2 \leq \mu_0$ then the optimal policy is:

$s_0 = 1$ *if* $n_0 > 0$,
and $s_0 = \rho_0$ *if* $n_0 = 0$.

If $\mu_1 + \mu_2 \geq \mu_0 \geq \mu_1, \mu_2$, then the optimal policy is:

$s_0 = 0$ *if* $n_1, n_2 > 0$,
$s_0 = 1 - \rho_2$ *if* $n_0 > 0, n_1 > 0, n_2 = 0$,
$s_0 = 1 - \rho_1$ *if* $n_0 > 0$ *and* $n_1 = 0$,
and $s_0 = \rho_0$ *otherwise.*

If $\mu_2 \geq \mu_0 \geq \mu_1$, then the optimal policy is:

$s_0 = 0$ *if* $n_2 > 0$,
$s_0 = 1 - \rho_2$ *if* $n_0 > 0$ *and* $n_2 = 0$,
and $s_0 = \rho_0$ *otherwise.*

The first two policies correspond to policies π^* and π^{**} respectively, which were optimal in the stochastic model (Propositions 2.1 and 2.2). The third case corresponds to strategy π^{***}. Recall from Lemma 2 that π^{***} is stable in the stochastic model under the standard conditions.

3.2 Average-optimal policies

With the exception of the third case in Proposition 3.2, all cases for which we only determined a characterization of average-optimal policies in the stochastic model, lead to average-optimal policies in the fluid model, as is confirmed by the next proposition.

Proposition 3.3 *Assume $\rho_1 \leq \rho_2$ and $\rho_0 + \rho_2 < 1$. An average-optimal (but not path-wise optimal) policy for the fluid model can be found in the following situations. In all cases $s_i = 1 - s_0$ if $n_i > 0$ and $s_i = \min(\rho_i, 1 - s_0)$ if $n_i = 0$, for $i = 1, 2$.*
If $\mu_1 \geq \mu_0 \geq \mu_2$, then the optimal policy is:

$s_0 = 0$ *if* $n_1 \geq \frac{\mu_2}{\mu_1+\mu_2-\mu_0} \times \frac{\mu_1}{\mu_0} \times \frac{\rho_2-\rho_1}{1-\rho_0-\rho_2} n_0$ *or if* $n_1 > 0$ *and* $n_2 > 0$,
$s_0 = \rho_0$ *if* $n_0 = n_1 = 0$,
$s_0 = 1-\rho_1$ *if* $n_0 > 0$ *and* $n_1 = 0$,
and $s_0 = 1-\rho_2$ *otherwise.*

If $\mu_1, \mu_2 \geq \mu_0$, *then the optimal policy is:*

$s_0 = 0$ *if* $n_1 \geq \frac{\rho_2-\rho_1}{1-\rho_0-\rho_2} n_0$ *or if* $n_2 > 0$,
and $s_0 = 1-\rho_2$ *otherwise.*

For illustration, these strategies are depicted in Figures 2 and 3 (the case when both $n_1 > 0$ and $n_2 > 0$ is not shown since then $s_0 = 0$). If $\mu_1 \geq \mu_0 \geq \mu_2$, there is a linear switching curve in the left plane of Figure 2 ($n_2 = 0$) above which class 1 must be served with full capacity. Below that curve, class 0 receives the fraction $1-\rho_2$ that is left from keeping class 2 empty. On the horizontal axis, class 0 would receive $1-\rho_1$, which forces class 2 to increase. The plane on the right shows that when class 1 is empty, it receives exactly its average and remains empty. Class 2 increases while class 0 is emptied.

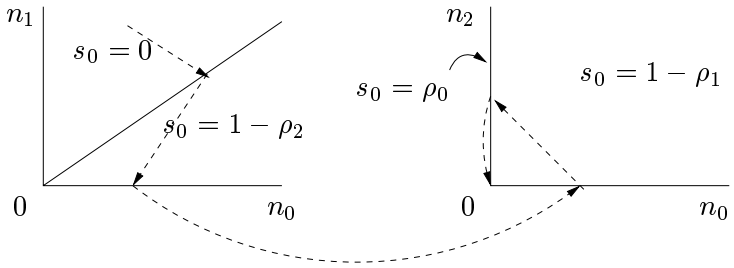


Figure 2: Optimal capacity allocation when $n_2 = 0$ (left) and $n_1 = 0$ (right); if $\mu_1 \geq \mu_0 \geq \mu_2$.

Similarly, if $\mu_1, \mu_2 \geq \mu_0$, there is a linear switching strategy in the left plane of Figure 3. The plane on the right shows that when class 1 is empty, it will remain empty. Class 0 receives no capacity unless class 2 is empty as well.

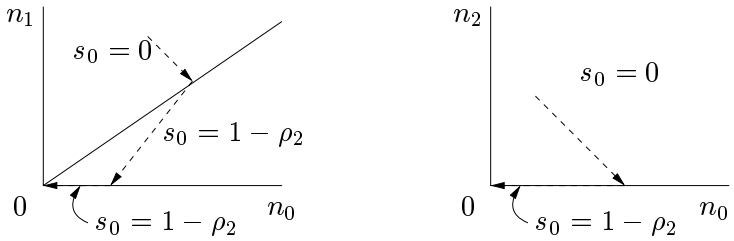


Figure 3: Optimal capacity allocation when $n_2 = 0$ (left) and $n_1 = 0$ (right); if $\mu_1, \mu_2 \geq \mu_0$.

The above two policies may be translated to the stochastic model as strategies with switching curves $h_i(N_0) = c_i N_0$. For the case $\mu_1 \geq \mu_0 \geq \mu_2$, we have $c_1 = \frac{\mu_2}{\mu_1+\mu_2-\mu_0} \times \frac{\mu_1}{\mu_0} \times \frac{\rho_2-\rho_1}{1-\rho_0-\rho_2}$ and $c_2 = \infty$. Note that c_1 depends on the traffic loads as well as the service rates. When $\mu_1, \mu_2 \geq \mu_0$, we have $c_1 = \frac{\rho_2-\rho_1}{1-\rho_0-\rho_2}$ and $c_2 = 0$. Now c_1 only depends on the traffic loads. This can be explained from the fact that the optimal fluid trajectory does not leave the $n_2 = 0$ plane when $\mu_1 \geq \mu_0 \geq \mu_2$. It is evident that the minimization of $\int_0^T (n_0(t)+n_1(t))\mathrm{d}t = \mu_0 \int_0^T (w_0(t)+\frac{\mu_1}{\mu_0} w_1(t))$ can only depend on μ_0 and μ_1 through their ratio. Combining this with a linear switching curve and the fact that $n_i(t) = \mu_i w_i(t)$, it must be that the optimal switching curve in terms of the (n_0, n_1) process is independent of μ_0 and μ_1.

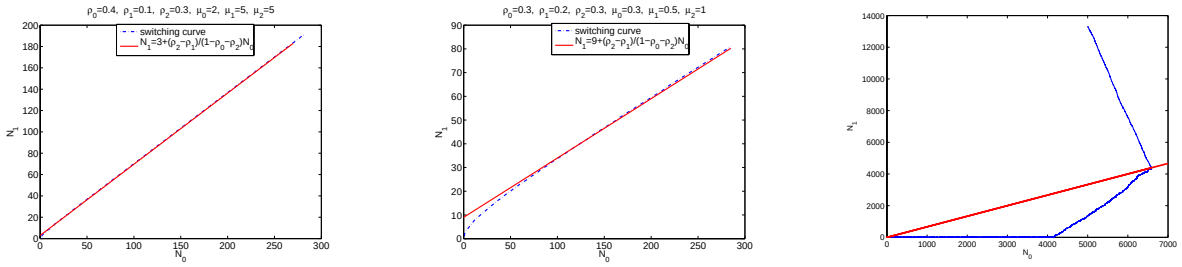


Figure 4: a) and b) Switching curves under the optimal policy in the stochastic model and the fluid approximation, when $\rho_1 < \rho_2$, c) Trajectory in the stochastic model with a linear switching curve.

The optimal switching curves for the stochastic model can be computed numerically by solving the dynamic programming equations. When $\rho_1 \neq \rho_2$, the optimal switching curves in the fluid model give a good approximation for the optimal switching curves in the stochastic model, see Figure 4 a) and b). However, this is not the case when $\rho_1 = \rho_2$, as will be explained in Section 4. In Figure 4 c) a trajectory of the stochastic model is plotted. We chose $\rho_0 = 0.4, \rho_1 = 0.1, \rho_2 = 0.3$ and $\mu_0 = 2$, $\mu_1 = \mu_2 = 5$. The switching curve is chosen as in Proposition 3.3, hence $h_1(N_0) = \frac{\rho_2 - \rho_1}{1-\rho_0-\rho_2} N_0$ and $h_2(N_0) = 0$. The starting state is $N_0 = 5000, N_1 = 13333$ and $N_2 = 0$. We see that the number of class-1 users decreases and the number of class-0 users increases until the trajectory hits the linear switching curve. From that moment on, both class 1 and class 0 decrease linearly in time. At the same time the number of class-2 users remains close to zero. This coincides with the dynamics of the fluid model.

3.3 Proofs of Propositions 3.2 and 3.3

In this section we present the proofs of Propositions 3.2 and 3.3. For the original model it was proved that it is optimal to serve classes 1 and 2 simultaneously, whenever both are present [31]. The fluid model inherits this property.

Observation 1 (Class 1 or 2 backlogged) *Assume $\mu_1 + \mu_2 \geq \mu_0$. Suppose at time t the state is $\boldsymbol{w}(t) = \boldsymbol{w}$. When $w_1, w_2 > 0$, then $s_1(t) = s_2(t) = 1$.*
Furthermore, when there is no backlog of either class 1 or class 2, an optimal policy always keeps at least one of these classes empty. Hence, if $w_i = 0$ and $w_j > 0$, $i, j = 1, 2$, then $s_i(t) = \rho_i$ and $s_j(t) \geq \rho_i$. If $w_i = w_j = 0$ and $\rho_i \leq \rho_j$, then $s_i(t) = \rho_i$ and $s_j(t) \geq \rho_i$.

Observation 1 fully characterizes the optimal policy in states where both classes 1 and 2 are backlogged. We therefore only need to consider the following two cases: no backlog of class 1 and no backlog of class 2.
No backlog of class 1: Suppose the system is in state $\boldsymbol{w}(t) = \boldsymbol{w}$, with $w_1 = 0$. Observation 1 implies that in the $w_1 = 0$-plane we have $s_1(t) = \rho_1$, so class 2 receives at least capacity ρ_1. Since $\rho_1 < \rho_2$, once the optimal trajectory has entered the $w_1 = 0$-plane, it will stay in this plane from then on, and the time until reaching the origin is the same for every non-idling policy. Therefore, the optimal way to allocate the remaining $1 - \rho_1$ capacity between classes 0 and 2 is as follows: when $\mu_0 \leq \mu_2$ it is optimal to fully prioritize class 2, but when $\mu_0 \geq \mu_2$ it is optimal to prioritize class 0 over class 2 (see Figures 5 a) and b) for the corresponding optimal trajectories).
No backlog of class 2: Suppose the system is in state $\boldsymbol{w}(t) = \boldsymbol{w}$, with $w_2 = 0$ and $w_1 > 0$. Observation 1 implies that in the $w_2 = 0$-plane we have $s_2(t) = \rho_2$ as long as $w_1 > 0$. We are

now left with finding the optimal way to allocate the remaining $1-\rho_2$ capacity between classes 0 and 1.
When $\mu_0 \geq \mu_1$, allocating the remaining capacity fully to class 0 is work-conserving in both nodes and maximizes the departure rate. It can be proved that this is indeed optimal.
When $\mu_0 < \mu_1$, giving full priority to class 1 maximizes the departure rate. This however leaves $1 - s_2(t) = 1 - \rho_2$ capacity in node 2 unutilized. As soon as $w_1 = 0$ we go to the $w_1 = 0$-plane and are faced with an unnecessarily high workload in node 2. As in the stochastic model, the trade-off between serving class 0 or 1 arises and it turns out to be optimal to give first priority to class 1 and then to switch to class 0. Let the switching point be denoted by $\boldsymbol{b} = (b_0, b_1)$, see Figure 5 c). In order to obtain the average-optimal policy, we only need to determine the optimal switching point. We do this by calculating the costs belonging to the trajectory that turns at $\boldsymbol{b}$. The time it takes to move from $\boldsymbol{w}$ to $\boldsymbol{b}$ is equal to $T(\boldsymbol{w}, \boldsymbol{b}) = \frac{b_0 - w_0}{\rho_0}$ during which the length of the queue is on average $\frac{w_0+b_0}{2}\mu_0 + \frac{w_1+b_1}{2}\mu_1$. The time it takes to move from $\boldsymbol{b}$ to $\bar{\boldsymbol{w}}$ is equal to $T(\boldsymbol{b}, \bar{\boldsymbol{w}}) = \frac{b_1}{\rho_2 - \rho_1}$ during which the length of the queue is on average $\frac{b_0+\bar{w}_0}{2}\mu_0 + \frac{b_1}{2}\mu_1$, with $\bar{w}_0 = b_0 - \frac{1-\rho_0-\rho_2}{\rho_2-\rho_1} b_1$. The costs for switching point $\boldsymbol{b}$ can now be written as:

$$K(\boldsymbol{w}, \boldsymbol{0}) = T(\boldsymbol{w}, \boldsymbol{b})(\frac{w_0 + b_0}{2}\mu_0 + \frac{w_1 + b_1}{2}\mu_1) + T(\boldsymbol{b}, \bar{\boldsymbol{w}})(\frac{b_0 + \bar{w}_0}{2}\mu_0 + \frac{b_1}{2}\mu_1) + K(\bar{\boldsymbol{w}}, \boldsymbol{0}), \qquad (3)$$

with $b_0 \in [w_0, w_0 + w_1 \frac{\rho_0}{1-\rho_1}]$, $b_1 = w_1 - T(\boldsymbol{w}, \boldsymbol{b})(1 - \rho_1)$ and $K(\bar{\boldsymbol{w}}, \boldsymbol{0})$ the costs belonging to the optimal trajectory that moves from $\bar{\boldsymbol{w}}$ to $\boldsymbol{0}$. The term $K(\bar{\boldsymbol{w}}, \boldsymbol{0})$ depends on the values of μ_0 and μ_2.
When $\mu_0 \leq \mu_2$, it takes $T(\bar{\boldsymbol{w}}, \boldsymbol{0}) = \frac{\bar{w}_0}{1-\rho_0-\rho_2}$ to reach the origin. Under the optimal policy there is no backlog of class 2. The average queue length of class 0 is $\bar{w}_0/2$, hence

$$K(\bar{\boldsymbol{w}}, \boldsymbol{0}) = T(\bar{\boldsymbol{w}}, \boldsymbol{0})\mu_0 \frac{\bar{w}_0}{2}.$$

When $\mu_0 \geq \mu_2$ an optimal trajectory will look like the path in Figure 5 b). It starts in $\bar{\boldsymbol{w}}$ and it hits the vertical axis in the point $\boldsymbol{d} = (0, 0, d_2)$. The time it takes to empty the system of class 0 is equal to $T(\bar{\boldsymbol{w}}, \boldsymbol{d}) = \frac{\bar{w}_0}{1-\rho_0-\rho_1}$. At that time the total work of class 2 has increased from 0 to d_2 and after that it decreases again to 0. Note that d_2 is equal to $d_2 = T(\bar{\boldsymbol{w}}, \boldsymbol{d})(\rho_2 - \rho_1)$. We can conclude that

$$K(\bar{\boldsymbol{w}}, \boldsymbol{0}) \quad = \quad T(\bar{\boldsymbol{w}}, \mathbf{d})\mu_0 \frac{\bar{w}_0}{2} + T(\bar{\boldsymbol{w}}, \boldsymbol{0})\mu_2 \frac{d_2}{2}.$$

It can now be checked that when minimizing the costs given by (3) over $\boldsymbol{b}$, the optimal $\boldsymbol{b}$ lies on the linear switching curve as stated in Proposition 3.3. See [28, Section 5] for more details.

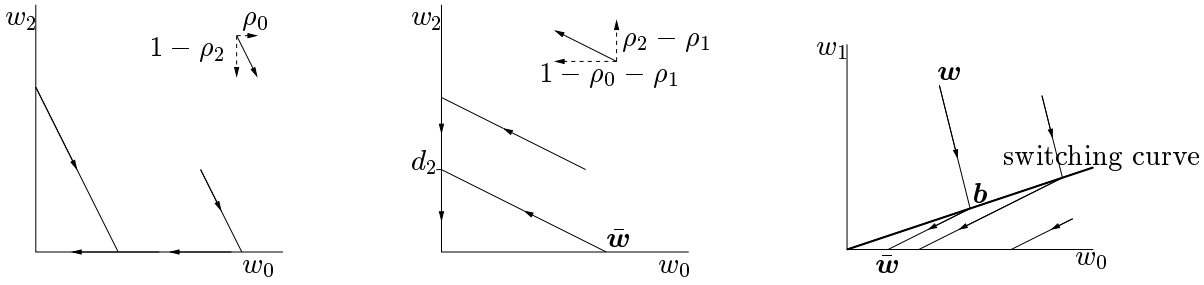


Figure 5: Optimal trajectories in the fluid model for a) $\mu_0 \leq \mu_2$ in the $w_1 = 0$-plane, b) $\mu_0 \geq \mu_2$ in the $w_1 = 0$-plane, c) $\mu_0 \leq \mu_1$ in the $w_2 = 0$-plane.

4 Central Limit Theorem scaling for $\rho_1 = \rho_2$

In Section 3 we determined the optimal policy for the fluid model. When $\rho_1 \neq \rho_2$, the policy suggested by the fluid analysis approximates the optimal switching curve very well (Figure 4). However, when $\rho_1 = \rho_2$ the suggested policy might not even be stable: Consider for example the situation that $\mu_1, \mu_2 \geq \mu_0$. For the fluid model, the switching curves are both equal to zero, i.e. it is optimal to serve class 0 only if there is work of neither class 1 nor class 2. However, in the stochastic model, giving classes 1 and 2 preemptive priority, leads unnecessarily to an unstable system if $1 - \rho_i > \rho_0 > (1 - \rho_1)(1 - \rho_2)$, $i = 1, 2$. Evidently, this policy is then far from optimal. In the fluid model we have no instability since we can keep classes 1 and 2 simultaneously empty, while in the stochastic model there can be stochastic fluctuations that cause the instability effects. We can conclude that the straightforward translation of the optimal fluid policy does not give an asymptotically optimal policy for the stochastic model. In this section we will investigate the correct scaling to find the shape of the optimal switching curves when $\rho_1 = \rho_2$.

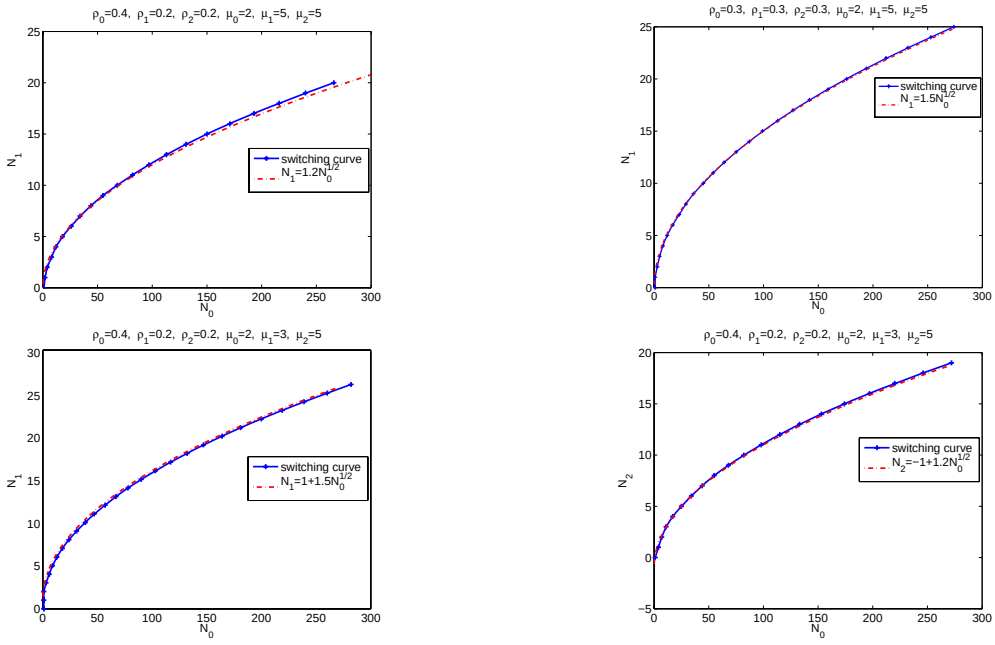


Figure 6: Optimal switching curves and square-root approximations for various parameters.

In Figure 6 we plotted the optimal switching curve for various parameters together with a function that provides a good approximation of the curve. The curves indicate that the switching curve has a sub-linear shape, and in fact is close to the square-root function.

To obtain the fluid model we scaled N_0 and N_1 identically. Due to its sub-linear shape, the switching curve collapses on the horizontal axis after taking the symmetric linear (fluid) scaling. Interpreting this as giving strict priority to class 1 can result in an unstable system. This illustrates the choice for a different scaling when $\rho_1 = \rho_2$: We need to scale the system such that the switching curve remains observable.

In the discussion below we do not consider the case where both $N_1(0)$ and $N_2(0)$ are positive. In that case we know it is optimal to serve both classes 1 and 2 until one of them empties. This is not different than in the fluid model. Without loss of generality we can therefore concentrate on initial points with $N_2(0) = 0$.

We generically denote the switching curves by $N_i = c_i f(N_0)$, for $i = 1, 2$. The function f is not specified for now (but we know that it will be close to the square-root function). Again we

consider the sequence of systems indexed by a superscript n, where the workload and number of users in the n-th system are denoted by $W_i^n(t)$ and $N_i^n(t)$ respectively. The initial queue lengths depend on n and are chosen in accordance with the above observations: $n_0(0) = a_0 > 0$, $n_1(0) = 0$, $\lim_{n\to\infty} \frac{N_1^n(0)}{\sqrt{n}} = b_1$ and $N_2^n(0) \equiv 0$. We are interested in the limit of the scaled processes: $\lim_{n\to\infty} \frac{N_i^n(nt)}{n} := n_i(t)$ and $\lim_{n\to\infty} \frac{W_i^n(nt)}{n} = w_i(t)$. We will see that $n_i(t) \equiv w_i(t) \equiv 0$ for $i = 1, 2$. Therefore, for classes $i = 1, 2$ we are also interested in the limit of the diffusion scaled processes:

$$\lim_{n\to\infty} \frac{N_i^n(nt) - n_i(t)n}{\sqrt{n}} \qquad \text{and} \qquad \lim_{n\to\infty} \frac{W_i^n(nt) - w_i(t)n}{\sqrt{n}},$$

see [9, 23]

Remark 4.1 *The workload and number of users present in the system under the fluid and diffusion scaling can be related in the following way:* $n_i(t) = \mu_i w_i(t)$ *and* $\lim_{n\to\infty} \frac{N_i^n(nt)}{\sqrt{n}} = \lim_{n\to\infty} \mu_i \frac{W_i^n(nt)}{\sqrt{n}}$*, respectively. This follows from the fact that we have exponentially distributed service requirements. Hence we may write* $W_i^n(nt) \stackrel{d}{=} N_i^n(nt) \frac{\sum_{k=1}^{N_i^n(nt)} Exp_k(\mu_i)}{N_i^n(nt)}$*, where* $Exp_k(\mu_i)$*,* $k = 1, 2, \ldots,$ *are i.i.d. exponential random variables with mean* $1/\mu_i$*. When* $\lim_{n\to\infty} N_i^n(nt) = \infty$*, we have that* $\lim_{n\to\infty} \frac{W_i^n(nt)}{N_i^n(nt)} = \frac{1}{\mu_i}$ *w.p. 1.*

So as to study the typical trajectories of the process above and below the switching curves, we will first describe the trajectories of the corresponding *free processes.*

4.1 Free process above the switching curve

Above the switching curve, class 1 is given preemptive priority. Hence, the free process that corresponds to the behavior above the switching curve is the process that gives class 1 priority, regardless of the number of class-0 users present. This means that, as in Section 3, the free process tends to move right and down, just as in Figure 5 a). Notice however that the initial point has an N_1-coordinate of the order $\sqrt{n}$, so that on the linear time scale, the process moves instantly in vertical direction downward until it hits the switching curve.

4.2 Free process below the switching curves

We now consider the free processes that correspond to the behavior of the stochastic process below the switching curve. In the free process class 0 receives priority and classes 1 and 2 are only served during (short) excursions when both of them are positive. The trajectories of the original model below the switching curve are therefore identical to those of the free process.
We reflect the fact that we look at the free process by adding the symbol $\sim$ to the notation. In the following proposition it is stated that the free process has two different types of components: The component corresponding to class 0 behaves as a deterministic fluid component, just as in Section 3, while classes 1 and 2 show random fluctuations of the order $\sqrt{n}$ in a time span n, i.e. their workloads remain of the order $\sqrt{n}$ with probability 1.

Proposition 4.2 *Consider the free process that gives class 0 priority, and classes 1 and 2 are served only when both of them are positive. When* $\rho_1 = \rho_2$*, we obtain the following limits:*

$$\begin{aligned} \tilde{w}_0(t) &= \tilde{w}_0(0) - (1 - \rho_0 - \rho_1)t, \\ \tilde{n}_0(t) &= \tilde{n}_0(0) - \mu_0(1 - \rho_0 - \rho_1)t, \end{aligned}$$

and $\tilde{w}_i(t) \equiv \tilde{n}_i(t) \equiv 0$ *for* $i = 1, 2$. *In addition,*

$$\lim_{n\to\infty} \frac{\tilde{W}_1^n(nt)}{\sqrt{n}} = \lim_{n\to\infty} \mathbf{1}_{(\tilde{W}_1^n(nt)\geq \tilde{W}_2^n(nt))} \frac{\tilde{W}_1^n(nt) - \tilde{W}_2^n(nt)}{\sqrt{n}} \stackrel{d}{=} \mathbf{1}_{(BM(t)+\frac{b_1}{\mu_1}\geq 0)} \Big(BM(t) + \frac{b_1}{\mu_1} \Big), \quad (4)$$

$$\lim_{n\to\infty} \frac{-\tilde{W}_2^n(nt)}{\sqrt{n}} = \lim_{n\to\infty} \mathbf{1}_{(\tilde{W}_2^n(nt)\geq \tilde{W}_1^n(nt))} \frac{\tilde{W}_1^n(nt) - \tilde{W}_2^n(nt)}{\sqrt{n}} \stackrel{d}{=} \mathbf{1}_{(BM(t)+\frac{b_1}{\mu_1}\leq 0)} \Big(BM(t) + \frac{b_1}{\mu_1} \Big), \quad (5)$$

and hence are well defined. Here $BM(t)$ *is a Brownian motion with variance* $\theta^2 := \lambda_1\theta_1^2 + \lambda_2\theta_2^2$ *and* $\theta_j^2 = Var(B_j)$.

Proof: Denote by $A_i(0,t)$ the amount of class-i work that arrived in the interval $(0,t)$ and by $\tilde{B}_i(0,t)$ the cumulative capacity that is given to class i in the interval $(0,t)$. We can write for $i = 1, 2$

$$\begin{aligned} \tilde{W}_i^n(t) &= \tilde{W}_i^n(0) + A_i(0,t) - \tilde{B}_i(0,t) & (6)\\ \tilde{B}_1(0,t) &= \tilde{B}_2(0,t), & (7) \end{aligned}$$

where the last equality holds since classes 1 and 2 are served only when both of them are positive. Using (6) and (7), we obtain

$$\tilde{W}_1^n(nt) - \tilde{W}_2^n(nt) = \tilde{W}_1^n(0) - \tilde{W}_2^n(0) + A_1(0,nt) - A_2(0,nt).$$

From Remark 4.1, the FCLT (Functional Central Limit Theorem) and the fact that we have Poisson arrivals, we conclude that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{\sqrt{n}} \Big(\tilde{W}_1^n(nt) - \tilde{W}_2^n(nt) \Big) &= \lim_{n\to\infty} \frac{1}{\sqrt{n}} \Big(A_1(0,nt) - A_2(0,nt) + \tilde{W}_1^n(0) - \tilde{W}_2^n(0) \Big) \\ &\stackrel{d}{=} BM(t) + \frac{b_1}{\mu_1}, & (8) \end{aligned}$$

where $BM(t)$ is a Brownian motion with variance θ^2. Now note that the smallest of the workloads of classes 1 and 2 is always served at rate 1 whenever it is positive:

$$W_{min}^n(t) := \min\{\tilde{W}_1^n(t), \tilde{W}_2^n(t)\} = \sup_{s\leq t}\{A_{\min}(s,t) - (t-s)\} \leq \sup_{s\leq t}\{\hat{A}(s,t) - (t-s)\}. \quad (9)$$

Here, $A_{\min}$ is the arrival process of the workload in the queue with the smallest workload, $\hat{A}(s,t) \equiv \max\{A_1(s,t), A_2(s,t)\}$ and in Appendix C it is proved that

$$A_{min}(s,t) \leq \hat{A}(s,t). \quad (10)$$

Since $\hat{A}(s,t)/(t-s) \to \rho_1 = \rho_2 < 1$ as $t - s \to \infty$, we may interpret the right hand side of (9) as the workload in a stable queue. Consequently, $\lim_{n\to\infty} W_{min}^n(nt)$, is bounded from above by a non-defective variable, which implies $\lim_{n\to\infty} \frac{W_{min}^n(nt)}{\sqrt{n}} = 0$. Together with (8) and

$$\tilde{W}_i^n(nt) = (\tilde{W}_i^n(nt) - \tilde{W}_{3-i}^n(nt) + W_{min}^n(nt))\mathbf{1}_{(\tilde{W}_i^n(nt)\geq \tilde{W}_{3-i}^n(nt))} + W_{min}^n(nt)\mathbf{1}_{(\tilde{W}_i^n(nt)< \tilde{W}_{3-i}^n(nt))},$$

we then obtain (4) and (5).

Let us now turn the attention to class 0 for the free process. As long as class 0 is not empty, both nodes are work-conserving, so that

$$\tilde{W}_0^n(nt) + \tilde{W}_1^n(nt) = \tilde{W}_0^n(0) + \tilde{W}_1^n(0) + A_0(0,nt) + A_1(0,nt) - nt.$$

Since $\lim_{n\to\infty} \frac{1}{n}\tilde{W}_1^n(nt) = 0$, this gives $\tilde{w}_0(t) = \tilde{w}_0(0) + (\rho_0 + \rho_1 - 1)t$. □

4.3 Discussion: shape of switching curve

From the above we can intuitively explain the square-root shape of the optimal switching curve. From the fluid scaling, we learned that optimality requires the process to stay close to the horizontal axis (the switching curve in fact coincides with the horizontal axis in the fluid scaling). Letting the switching curve be too close to the horizontal axis, however, poses the risk of significant capacity loss: Capacity is lost in node 2 if we are above the switching curve in the plane $N_2 = 0$ and, vice versa, capacity is lost in node 1 if we are above the switching curve in the plane $N_1 = 0$. The switching curve must therefore be high enough to make it sufficiently unlikely for the process to reach it from below. But it need not be impossible to reach the switching curve, because above the switching curve the departure rate is higher.
Since the free process has zero drift for the components $\tilde{N}_1$ and $\tilde{N}_2$ and fluctuations in linear time $O(n)$ are of the order $O(\sqrt{n})$, the CLT (Central Limit Theorem) indicates that a square-root switching curve is able to strike the right balance between short and long term optimality. For comparison: a linear switching curve would be impossible to reach, therefore the strategy would not profit from serving the fast class 1 or 2 even if there is a lot of work from it. On the other hand, a threshold strategy (a constant switching curve) can quickly give instability problems as at large states, it is too easy to move up to the switching curve, thus risking considerable capacity loss.
Hence we may conclude that switching curves are of the shape $N_i = c_i\sqrt{N_0}$. To find the best coefficient c_i is not straightforward and involves calculating the exact first passage probabilities for the switching curves. In the next section the impact of the c_i is further described.

4.4 Illustration

Determining the optimal coefficient c_i is not straightforward and involves calculating the exact first passage probabilities for the switching curve. In this section we numerically illustrate the impact the choice for c_i has. The parameters are chosen as $\rho_0 = 0.4, \rho_1 = \rho_2 = 0.2$ and $\mu_0 = 2, \mu_1 = \mu_2 = 5$.
In the first set of simulations we chose the switching curves $N_i = c_i\sqrt{N_0}$ with $c_i = 6/5$, for $i = 1, 2$. In Figure 7 we see that the number of class-0 users indeed decreases linearly in time (left graph), while the minimum of the number of class-1 and class-2 users is typically very small (middle graph). The most right graph shows the trajectory of the difference between the number of class-1 and class-2 users. Recall from Proposition 4.2 that in the limit the process $W_1 - W_2$ represents W_1 when it is positive, and $-W_2$ when it is negative. We see that as the number of class-0 users decreases, the trajectory stays mostly below the switching curves, making some excursions between the switching curves in both planes.

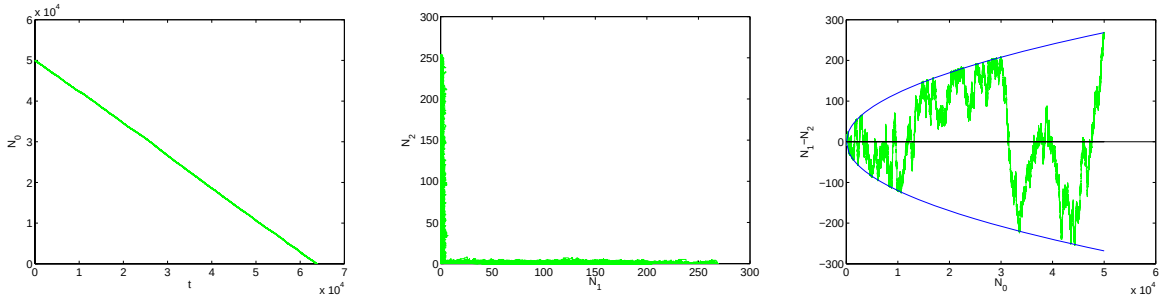


Figure 7: Trajectories of N_0, N_1, N_2 and $N_1 - N_2$ under a policy with switching curve $6/5\sqrt{N_0}$.

Taking c_i larger implies that for points that just lie under the switching curve in the $N_2 = 0$ plane, the probability of emptying the work in class 1 before hitting the switching curve again,

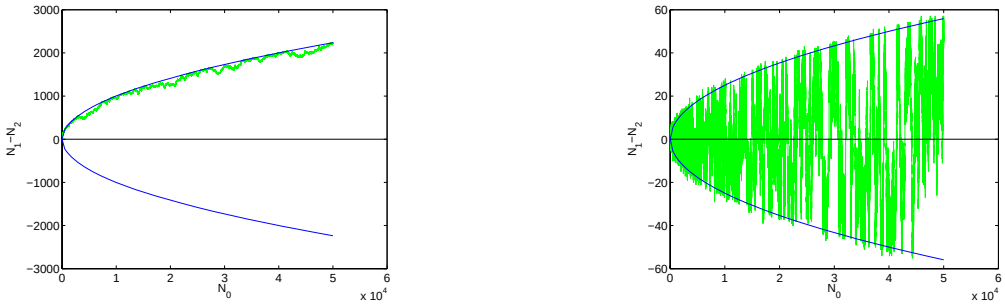


Figure 8: Trajectory of $N_1 - N_2$ under the policy with switching curve: a) $10\sqrt{N_0}$ and b) $1/4\sqrt{N_0}$.

becomes almost zero, and the number of class-2 users is zero or very small, see Figure 8 a) where $c_i = 10$. Therefore, if c_i is too large, the policy will probably focus too much on being work-conserving. Since $\mu_1 > \mu_0$, we could better serve more class-1 users when there are many of them. On the other hand, taking c_1 too small, we see that we switch too often between the two planes and we loose too much capacity, see Figure 8 b) where $c_i = 1/4$.

5 Numerical comparisons of α-fair allocations

In this section we numerically compare α-fair allocations with (asymptotically) optimal switching curve policies. Determining the true optimal policy is extremely time consuming (the typical computation time of a single scenario is in the order of a week, versus a few hours for the asymptotically optimal strategies) and could therefore not be reported in all cases.

5.1 Switching curve policies

We have conducted a large set of simulation experiments to assess the effectiveness of different switching curve policies. Class i has switching curve $N_i = c_i f(N_0)$, where $f(N_0)$ is either a square-root, linear or constant (threshold policy). The value of c_i is varied to assess its impact. We let $\mu_0 = 2$, $\mu_1 = \mu_2 = 5$. We simulate $3 \cdot 10^6$ busy periods and the obtained mean total numbers of users under the different policies are compared with the proportional fair policy (PF). This policy falls within the class of so-called α-fair allocations [19] (α=1), which are of practical interest as a convenient modeling approach for (distributed) allocation mechanisms. Other special cases are $\alpha = 0$ (maximizing total throughput), $\alpha = 2$ (mimics TCP) and $\alpha = \infty$ (max-min fairness). For proportional fairness the joint distribution of the numbers of users of the various classes in the linear network is known, see [18]. In particular, $\mathbb{E}(N^{PF}) = \frac{1}{1-\rho_0}\left(\rho_0 + \sum_{i=1}^{2} \frac{\rho_i}{1-\rho_0-\rho_i}\right)$. In Figures 9 and 10 we plot the relative improvement of switching policies over the mean number of users under proportional fair scheduling. On the horizontal axis we vary the value of the pre-constant.

In Figure 9 a) we considered $\rho_1 \neq \rho_2$ and chose $c_2 = 0$ and let c_1 vary. From the figure we observe that the linear policy indeed attains the value of the optimal policy given that the best coefficient c_1 is chosen. This is in accordance with Proposition 3.3 which stated that when $\mu_1, \mu_2 \geq \mu_0$ the optimal fluid policy has a linear switching curve for class 1 and gives preemptive priority to class 2. Suprisingly the square-root policy performs very well as well.

When c_1 grows large, the behavior of the system converges to that of policy π^{***}. We observe that policy π^{***} is already close to optimal. This is not surprising since policy π^{***} is optimal in heavy traffic: suppose that only node 2 is heavily loaded while node 1 is not ($\rho_0 + \rho_2 \approx 1$ and

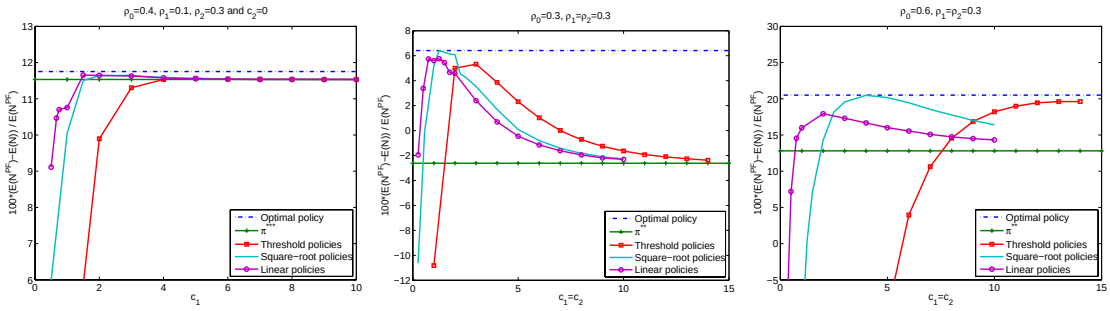


Figure 9: Relative improvement over proportional fair policy for the optimal policy and the switching curve policies: a) $\rho_0 = 0.4, \rho_1 = 0.1, \rho_2 = 0.3$, b) $\rho_0 = \rho_1 = \rho_2 = 0.3$, c) $\rho_0 = 0.6, \rho_1 = \rho_2 = 0.3$.

$\rho_0 + \rho_1 < 1$). To minimize the mean of the scaled total number of users present in the system, only what happens in node 2 matters. Since $\mu_2 > \mu_0$, in node 2 it is optimal to give preemptive priority to class 2 over class 0 and this is exactly what π^{***} does.
In Figures 9 b) and c) we considered $\rho_1 = \rho_2$ and chose $c_1 = c_2$, i.e. the switching curves for both classes are identical. We observe that for $\rho_1 = \rho_2$ the square-root policy attains the value of the optimal policy given that the best coefficient c_i is chosen. This coincides with the discussion of the shape of the switching curve in Section 4.3. Also note that Figure 9 b) corresponds to the graph in Figure 6. The approximation we found there for the switching curve was $1.5\sqrt{N_0}$ which indeed is close to optimal. When c_i grows large, the behavior of the system converges to that of policy π^{**} (defined in Proposition 2.2), which is work-conserving. The fastest convergence takes place for the linear policy.
In Figure 10 we test the effect the preconstant c_i has on the square-root, threshold and linear policies for several combinations of the loads with $\rho_1 = \rho_2$. We see that the square-root policy performs the best, given that the value of c_i is chosen optimally. The linear strategies perform surprisingly well, although they are less efficient than the square-root strategies. The square-root and linear policies are not that sensitive to the actual value of c_i, as long as its value is not too small. The threshold strategies are more sensitive to the choice of the pre-constant and run a higher risk of being unstable. Overall we observe that the best choice among these strategies only gives a modest improvement over proportional fair.

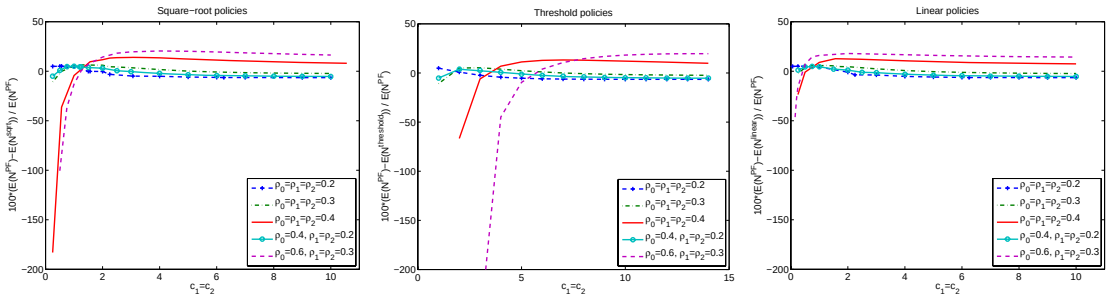


Figure 10: Relative improvement over proportional fair policy for a) square-root b) threshold and c) linear.

5.2 Weighted α-fair policies

For completeness, in Figure 11 (left) we test the efficiency of α-fair policies, against proportional fair and notice that α-fair policies seem to be rather insensitive to the value of α, as long as α

is not too small.
We next test the scope for improvement of α-fair allocations by adding weights to the various classes. With weighted α-fair policies [7], the capacity for class-0 users is given by

$$s_0 = \frac{(w_0 N_0(t)^\alpha)^{1/\alpha}}{(w_0 N_0(t)^\alpha)^{1/\alpha} + (w_1 N_1(t)^\alpha + w_2 N_2(t)^\alpha)^{1/\alpha}}.$$

The remaining capacity in node i is allocated to class i. (The standard α-fair allocation has unit weights.) There are very few results on how to set the weights in these algorithms. Recently, explicit results were obtained in [12] under an overload regime. In this section we will investigate what the effect is of changing the weights and whether we can approximate the optimal policy with a weighted α-fair strategy. Without loss of generality we fix $w_0 = 1$.
For $\mu_0 > \mu_1 + \mu_2$ the optimal policy is π^* (Proposition 2.1), i.e. preemptive priority to class 0. This policy can be approximated by the weighted α-fair policy by setting the weights w_1 and w_2 approximately equal to zero.
In the numerical examples we choose $\mu_0 = 2$ and $\mu_1 = \mu_2 = 5$. However, the observations hold for $\mu_0 < \mu_1, \mu_2$. In Figures 11 b) and c) ($\alpha = 1$, i.e. proportional fair) and Figures 12 a) and b) ($\alpha = 2$, corresponding to TCP) we compare weighted α-fair policies with the optimal policy.
From Figures 11 b) and 12 a) we observe that when $\rho_1 < \rho_2$, choosing $w_1 = 0$ and $w_2 = \infty$ approximates the optimal policy very well. In fact, when choosing these weights we obtain policy π^{***}, which, as observed in Section 5.1, is close to optimal.
From Figures 11 c) and 12 c) we observe that when $\rho_1 = \rho_2$, one of the two weights is ∞ and the other weight is strictly positive. For proportional fair the optimal weights are $w_i = 1/2$, $w_{3-i} = \infty$, $i = 1, 2$ and for TCP the optimal weights are $w_i = 1/8, w_{3-i} = \infty$, $i = 1, 2$. This can be explained as follows. From Proposition 2.3 we know that when both class-1 and class-2 users are present, the optimal allocation gives the full capacity to classes 1 and 2. Having one of the weights equal to ∞, say w_2, guarantees that the weighted α-fair policy does this as well. Now when there are no class-2 users present, there exists a switching curve that determines the optimal trade-off between serving class 0 or class 1, see Proposition 2.3. In the case of weighted α-fair policy, when there are no class-2 users present the allocated capacity to class 0 is $s_0 = \frac{N_0(t)}{N_0(t)+w_1^{1/\alpha} N_1(t)}$, which coincides with Discriminatory Processor Sharing. Here as well, there exists a $0 < w_1 < \infty$ that finds the best way to share the capacity between class 0 and class 1 (note that $w_1 = 0$ ($w_1 = \infty$) implies that class 0 (class 1) is given strict priority).
Similarly, the optimal weights for the remaining cases can be found. When $\mu_2 < \mu_0 < \mu_1$ and $\rho_2 < \rho_1$, the optimal weights will be $w_1 = \infty$ and $w_2 = 0$. This coincides with the optimal fluid policy π^{***} (Proposition 3.2). However, when $\rho_1 \leq \rho_2$, the optimal fluid policy has a switching curve in the $n_2 = 0$ plane (Proposition 3.3). So then the optimal w_1 is non-degenerate and $w_2 = 0$.
When $\mu_1, \mu_2 < \mu_0 < \mu_1 + \mu_2$, the weights w_1 and w_2 that approximate the optimal policy π^{**} will be strictly positive and small compared to $w_0 = 1$.

6 Conclusion and future work

Using appropriate scaling approaches, we determined accurate approximations to optimal allocation strategies in a linear bandwidth-sharing network. The (theoretical) asymptotically optimal policies obtained after scaling were shown to provide sensible benchmarks for assessing the performance of any allocation strategy. Doing so, α-fair allocations were shown to perform quite well in general, and are practically insensitive to the value of α, as long as this value is not too small. For $\alpha \downarrow 0$, the performance can either be very good or quite bad, depending on the specific choice of arrival rates, service requirements, etc. Specifically, we showed that

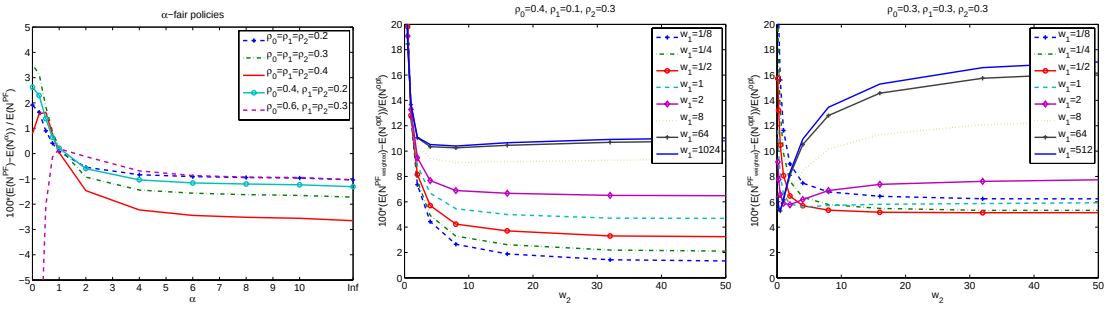


Figure 11: Left: Relative improvement over proportional fair policy for α-fair policies. Right: Comparison of the optimal policy with the Proportional Fair policy (α=1) for different choices of the weights: b) $\rho_0 = 0.4, \rho_1 = 0.1, \rho_2 = 0.3$, c) $\rho_0 = \rho_1 = \rho_2 = 0.3$.

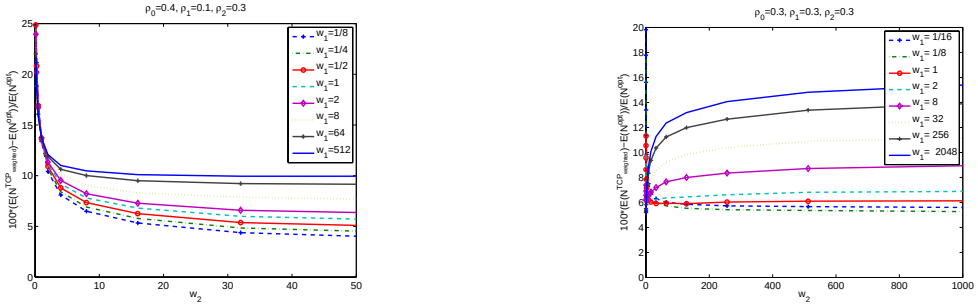


Figure 12: Comparison of the optimal policy with TCP (α=2) for different choices of the weights: a) $\rho_0 = 0.4, \rho_1 = 0.1, \rho_2 = 0.3$, b) $\rho_0 = \rho_1 = \rho_2 = 0.3$.

the (weighted) proportional fair allocation ($\alpha = 1$) performed well in all our experiments and is usually only a few percent away from the theoretical optimum.

We allowed the optimal allocations to use global information about the numbers of flows traversing all nodes in the network. In practice, such global information is usually not available. The fact that (weighted) α-fair strategies in general, and the proportional fair allocation in particular, achieve close-to-optimal behavior is therefore extremely encouraging since these can be implemented in a distributed fashion.

The main motivation for this work is the inappropriateness of size-based scheduling *across classes*, as discussed in the introduction. It is important to note that applying size-based scheduling *among flows within a class* (i.e., flows that traverse the same path through the network) is, in general, advantageous [1] and therefore offers potential for further improvement of any of the allocations discussed here. Assessing the scope of these mechanisms is an interesting avenue for further investigation. One practical complication in this respect could however be how to identify whether flows share (large parts of) the same path.

As a final remark, we note that the optimal multiplicative prefactor in the square-root rule (for the case $\rho_1 = \rho_2$), has so far been determined numerically. We saw in all our experiments, that the optimum can indeed be attained for a specific choice of the multiplier. The computation time of this procedure is virtually negligible compared to numerically determining the true optimal strategy. In order to *analytically* characterize the optimal value for the limit process, further investigation of the reflection of that process on the switching curve is required. In a different context, it is known that the reflection can have a decisive effect on the global behavior of the system, see [21]. Finding the appropriate constant requires calculation of first-entrance probabilities at the switching curve, as for example in [2, 6, 20]. Investigating these matters is

the subject of on-going research.

References

[1] Aalto, S., Ayesta, U. (2006). SRPT applied to bandwidth-sharing networks. In: *Proc. Euro-NGI workshop Stopera*, November 2006, Amsterdam.

[2] Abundo, M. (2002). Some conditional crossing results of Brownian motion over a piecewise-linear boundary. *Stat. & Prob. Letters* **58**, 131–145.

[3] Altman, E., Avrachenkov, K.E., Barakat, C. (2002). TCP network calculus: The case of large delay-bandwidth product. In: *Proc. IEEE Infocom 2002.*

[4] Altman, E., Avrachenkov, K.E., Kherani, A.A., Prabhu, B.J. (2005). Performance analysis and stochastic stability of congestion control protocols. In: *Proc. IEEE Infocom 2005.*

[5] Asmussen, S. (1987). *Applied Probability and Queues.* John Wiley & Sons.

[6] Beghin, L., Orsingher, E. (1999). On the maximum of the generalized Brownian bridge. *Lith. Math. Journ.* **39**, 157–167.

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

[8] Bonald, T., Massoulié, L., Proutière, A., Virtamo, J. (2006). A queueing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems* **53**, 65–84.

[9] Chen, H., Yao, D.D. (2001). *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization.* Springer-Verlag, New York.

[10] Cohen, J.W., Boxma, O.J. (1983). *Boundary Value Problems in Queueing System Analysis.* North-Holland, Amsterdam.

[11] Dai, J.G., Lin, W. (2005). Maximum pressure policies in stochastic processing networks. *Oper. Res.* **53**, 197–218.

[12] Egorova, R., Borst, S.C., Zwart, A.P. (2007). Bandwidth-sharing networks in overload. To appear in: *Proc. Performance 2007 Conf.*, Cologne, Germany.

[13] Fayolle, G., de la Fortelle, A., Lasgouttes, J.M., Massoulié, L., Roberts, J.W. (2001). Best-effort networks: modeling and performance analysis via large networks asymptotics. In: *Proc. IEEE Infocom 2001.*

[14] Kelly, F.P. (2003). Fairness and stability of end-to-end congestion control. *Eur. J. Control* **9**, 159–176.

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

[16] Kelly, F.P., Williams, R.J. (2004). Fluid model for a network operating under a fair bandwidth-sharing policy. *Ann. Appl. Prob.* **14**, 1055–1083.

[17] Lu, S.H., Kumar, P.R. (1991). Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Aut. Control* **36**, 1406–1416.

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

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

[20] Novikov, A., Frishling, V., Kordzakhia, N. (1999). Approximations of boundary crossing probabilities for a Brownian motion. *J. Appl. Prob.* **36**, 1019–1030.

[21] Puhalskii, A.A., Reiman, M.I. (1998). A critically loaded multirate link with trunk reservation. *Queueing Systems* **28**, 157–190.

[22] Righter, R., Shanthikumar, J.G. (1989). Scheduling multiclass single-server queueing systems to stochastically maximize the number of successful departures. *Prob. Eng. Inf. Sc.* **3**, 323–333.

[23] Robert, P. (2003). *Stochastic Networks and Queues.* Springer-Verlag, New York.

[24] Rybko, A.N., Stolyar, A.L. (1992). Ergodicity of stochastic processes describing the operation of open queuing networks. *Problems of Information Transmission* **28**, 199–220.

[25] Schrage, L.E. (1968). A proof of the optimality of the shortest remaining processing time discipline. *Oper. Res.* **16**, 687–690.

[26] Srikant, R. (2004). *The Mathematics of Internet Congestion Control.* Birkhauser, Boston.

[27] Takagi, H. (1991). *Queueing Analysis, Vol. I: Vacation and Priority Systems.* North-Holland, Amsterdam.

[28] Verloop, I.M. (2005). Efficient flow scheduling in resource-sharing networks. Master thesis, Utrecht University. http://www.cwi.nl/ maaike.

[29] Verloop, I.M., Borst, S.C. (2007). Heavy-traffic delay minimization in bandwidth-sharing networks. In: *Proc. IEEE Infocom 2007.*

[30] Verloop, I.M., Borst, S.C., Núñez-Queija, R. (2005). Stability of size-based scheduling disciplines in resource-sharing networks. In: *Proc. Performance 2005 Conf.*, Juan-les-Pins, France, 247–262.

[31] Verloop, I.M., Borst, S.C., Núñez-Queija, R. (2006). Delay optimization in bandwidth-sharing networks. In: *Proc. CISS 2006.*

Appendix A: Proof of Lemma 1

We couple the systems that arise under the two switching strategies by taking the same arrival and service requirement sequences. We will show that (1) and (2) hold on each sample path. Since the service requirements are exponentially distributed, the scheduling within classes does not influence the stochastic behavior of the system (recall that we restrict to size-oblivious strategies). For our coupling arguments it is convenient to assume that FCFS (First-Come First-Served) is applied within each class. As a consequence, (1) and (2) which hold in terms of workloads, immediately translate to the same inequalities in terms of the numbers of users. Let s be the first time instant that one of the three inequalities is violated. We will show that such an s does not exist.

First assume that at time s, equation (1) is violated, that is $W_0^g(s^+) > W_0^h(s^+)$ while $W_0^g(s) = W_0^h(s)$. From (2) we have $W_i^g(s) \leq W_i^h(s)$, $i = 1, 2$. To ensure that $W_0^g(s^+) > W_0^h(s^+)$, policy g must serve classes 1 and/or 2 while policy h serves class 0 at time s. Since $W_i^g(s) \leq W_i^h(s)$, $i = 1, 2$, serving classes 1 and/or 2 under policy g, implies that also under policy h classes 1 and/or 2 are served (since $h_i(n_0) \leq g_i(n_0)$ and $N_0^g(s) = N_0^h(s)$), which yields a contradiction.

Now assume equation (2) for $i = 2$ is the first to be violated at time s. Hence $W_0^g(s) + W_2^g(s) = W_0^h(s) + W_2^h(s)$, $W_0^g(s^+) + W_2^g(s^+) > W_0^h(s^+) + W_2^h(s^+)$, $W_0^g(t) \leq W_0^h(t)$, for $t \leq s^+$, and at time s, policy g serves class 1 and there is no work of class 2 present ($W_2^g(s) = 0$). We can conclude from the above that $W_2^g(s) \geq W_2^h(s)$. But $W_2^g(s) = 0$, so that $W_2^h(s) = 0$ as well. Since $W_0^g(s^+) \leq W_0^h(s^+)$, we now obtain that $W_0^g(s^+) + W_2^g(s^+) \leq W_0^h(s^+) + W_2^h(s^+)$, which contradicts the initial assumption. □

Appendix B: Proof of Lemma (2)

In node 2 the policy is work-conserving, hence class 0 and class 2 are stable if and only if $\rho_0 + \rho_2 < 1$.
Define $s_1 = \sup\{u \leq t : W_1(u) = 0\}$ and $s = \sup\{u \leq s_1 : W_0(u) + W_2(u) = 0\}$. Then

$$\begin{aligned}
W_0(t) + W_1(t) &= W_0(t) + A_1(s_1,t) - B_1(s_1,t) = W_0(t) + A_1(s_1,t) - (t - s_1) + B_0(s_1,t) \\
&= W_0(t) + A_1(s_1,t) - (t - s_1) + W_0(s_1) - W_0(t) + A_0(s_1,t) \\
&\leq A_1(s_1,t) - (t - s_1) + W_0(s_1) + W_2(s_1) + A_0(s_1,t) \\
&= A_1(s_1,t) - (t - s_1) + A_0(s_1,t) + A_0(s,s_1) + A_2(s,s_1) - (s_1 - s) \\
&= A_1(s_1,t) + A_0(s,t) + A_2(s,t) - A_2(s_1,t) - (t - s) \\
&= A_1(s_1,t) - (\rho_1 + \epsilon)(t - s_1) + A_0(s,t) - (\rho_0 + \epsilon)(t - s) + A_2(s,t) - (\rho_2 + \epsilon)(t - s) \\
&\quad + (\rho_2 - \epsilon)(t - s_1) - A_2(s_1,t) + R,
\end{aligned}$$

with $\epsilon = \frac{1 - \rho_0 - \max(\rho_1,\rho_2)}{4}$ and $R = (\rho_1 + \epsilon)(t - s_1) + (\rho_0 + \epsilon)(t - s) + (\rho_2 + \epsilon)(t - s) - (\rho_2 - \epsilon)(t - s_1) - (t - s)$. The fourth equation follows from the fact that node 2 is work-conserving, i.e. when node 2 is backlogged, the work is served at full rate.
For $\rho_2 \geq \rho_1$, we can bound R from above as follows:

$$\begin{aligned}
R &\leq (\rho_2 + \epsilon)(t - s_1) + (\rho_0 + \epsilon)(t - s) + (\rho_2 + \epsilon)(t - s) - (\rho_2 - \epsilon)(t - s_1) - (t - s) \\
&= (\rho_0 + \rho_2 - 1)(t - s) + \epsilon(4t - 2s_1 - 2s) \leq (\rho_0 + \rho_2 + 4\epsilon - 1)(t - s) = 0.
\end{aligned}$$

For $\rho_2 \leq \rho_1$, we have $\rho_1 - \rho_2 + 2\epsilon \geq 0$ and we bound R from above as follows:

$$\begin{aligned}
R &= t(\rho_0 + \rho_1 + 4\epsilon - 1) - s(\rho_0 + \rho_2 + 2\epsilon - 1) - s_2(\rho_1 - \rho_2 + 2\epsilon) \\
&\leq t(\rho_0 + \rho_1 + 4\epsilon - 1) - s(\rho_0 + \rho_2 + 2\epsilon - 1) - s(\rho_1 - \rho_2 + 2\epsilon) = (\rho_0 + \rho_1 + 4\epsilon - 1)(t - s) = 0.
\end{aligned}$$

Denote by $\hat{W}_i^c(t)$ the workload at time t in a reference system with class-i traffic only, service rate c, and with $\hat{W}_i^c(0) = 0$. Define $U_j^d(t) := \sup_{0 \leq s \leq t}\{d(t - s) - A_j(s,t)\}$. Since $R \leq 0$, we have

$$\begin{aligned}
W_0(t) + W_1(t) &\leq \sup_{0 \leq s \leq t}\{A_1(s,t) - (\rho_1 + \epsilon)(t - s)\} + \sup_{0 \leq s \leq t}\{A_0(s,t) - (\rho_0 + \epsilon)(t - s)\} \\
&\quad + \sup_{0 \leq s \leq t}\{A_2(s,t) - (\rho_2 + \epsilon)(t - s)\} + \sup_{0 \leq s \leq t}\{(\rho_2 - \epsilon)(t - s) - A_2(s,t)\} \\
&= \hat{W}_1^{\rho_1 + \epsilon}(t) + \hat{W}_0^{\rho_0 + \epsilon}(t) + \hat{W}_2^{\rho_2 + \epsilon}(t) + U_2^{\rho_2 - \epsilon}(t). \qquad (11)
\end{aligned}$$

The first three terms in (11) are now the workloads in stable queues since the service rate is larger than the offered loads. The last term can be replaced by the supremum of a random walk with drift $\rho_2 - \epsilon - \rho_2 < 0$. Since the drift is negative, $U_2^{\rho_2 - \epsilon}(t) < \infty$ almost surely, [5]. Hence the workload in node 1 can be bounded from above by four terms that are almost surely finite, which implies stability of classes 0 and 1. □

Appendix C: Proof of relation (10)

Consider the free process below the switching curve as described in Section 4.2. Define $W_{min}(t) = \min(\tilde{W}_1(t), \tilde{W}_2(t))$ and $W_{max}(t) = \max(\tilde{W}_1(t), \tilde{W}_2(t))$. In the proof of Proposition 4.2 we relied on equation (10), which follows from the following lemma.

Lemma 3 *For $i = 1, 2$ we have $A_{min}(s,t) = A_i(s,t) + (\tilde{W}_i(s) - \tilde{W}_{3-i}(s))^+ - (\tilde{W}_i(t) - \tilde{W}_{3-i}(t))^+$, hence $A_{min}(s,t) \leq \max(A_1(s,t), A_2(s,t)) = \hat{A}(s,t)$.*

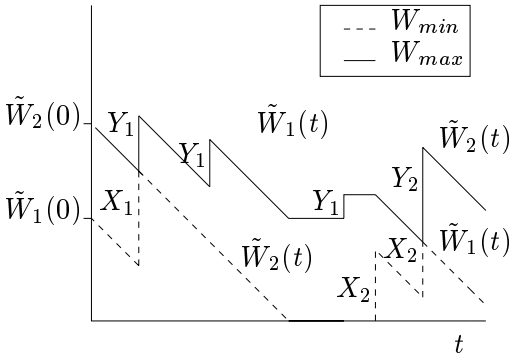
Proof: Let τ_{ik} be the k-th time in the interval (s,t), that a class-i user arrives in the system and let $B_{i,k}$ be the corresponding service requirement. Let $N_i(s,t)$ be the number of class-i users that arrive in the system during (s,t). We define for $i=1,2$ and $k=1,2,\ldots,N_i(s,t)$:

$$\begin{aligned} X_{i,k} &= \min((\tilde{W}_{3-i}(\tau_{i,k}^-) - \tilde{W}_i(\tau_{i,k}^-))^+, B_{i,k}), \\ Y_{i,k} &= B_{i,k} - X_{i,k}. \end{aligned}$$

Obviously

$$A_i(s,t) = \sum_{k=1}^{N_i(s,t)} (X_{i,k} + Y_{i,k}). \tag{12}$$

The processes $\tilde{W}_1$ and $\tilde{W}_2$ decrease either both at constant rate 1, or stay both constant (this happens when $\tilde{W}_i = 0$ for an i). $\tilde{W}_i$ increases with $B_{i,k} = X_{i,k} + Y_{i,k}$ when an arrival of class i occurs. From the figure on the right we see that the term $X_{i,k}$ corresponds to the arrival of work in the process $W_{min}(t)$. Hence



$$A_{min}(s,t) = \sum_{k=1}^{N_1(s,t)} X_{1,k} + \sum_{k=1}^{N_2(s,t)} X_{2,k}. \tag{13}$$

Assume for the moment that we can prove for $i=1,2$:

$$(\tilde{W}_i(s) - \tilde{W}_{3-i}(s))^+ + \sum_{k=1}^{N_i(s,t)} Y_{i,k} = \sum_{k=1}^{N_{3-i}(s,t)} X_{3-i,k} + (\tilde{W}_i(t) - \tilde{W}_{3-i}(t))^+. \tag{14}$$

Together with (12) and (13) we then obtain the desired result: $A_{min}(s,t) = A_i(s,t) + (\tilde{W}_i(s) - \tilde{W}_{3-i}(s))^+ - (\tilde{W}_i(t) - \tilde{W}_{3-i}(t))^+$.

We prove equation (14) by induction. Assume it holds at time u, that is for $i=1,2$:

$$(\tilde{W}_i(s) - \tilde{W}_{3-i}(s))^+ + \sum_{k=1}^{N_i(s,u)} Y_{i,k} = \sum_{k=1}^{N_{3-i}(s,u)} X_{3-i,k} + (\tilde{W}_i(u) - \tilde{W}_{3-i}(u))^+,$$

and the next arrival occurs at time v. Assume it is an arrival of class 1.

First assume that $\tilde{W}_1(u) \geq \tilde{W}_2(u)$. At time v, we still have $\tilde{W}_1(v) \geq \tilde{W}_2(v)$. Now $X_{1,N_1(s,v)} = 0$ and $Y_{1,N_1(s,v)} = B_{1,N_1(s,v)}$ and (14) immediately holds for $i=2$ at time v. Furthermore, $(\tilde{W}_1(v) - \tilde{W}_2(v))^+ - (\tilde{W}_1(u) - \tilde{W}_2(u))^+ = B_{1,N_1(s,v)} = Y_{1,N_1(s,v)}$, hence for $i=2$ it holds as well.

Now assume that $\tilde{W}_1(u) \leq \tilde{W}_2(u)$ and $B_{1,N_1(s,v)} \leq \tilde{W}_2(u) - \tilde{W}_1(u)$. At time v we still have $\tilde{W}_1(v) \leq \tilde{W}_2(v)$. Now $X_{1,N_1(s,v)} = B_{1,N_1(s,v)}$ and $Y_{1,N_1(s,v)} = 0$ and (14) immediately holds for $i=1$ at time v. Furthermore, $(\tilde{W}_2(v) - \tilde{W}_1(v))^+ - (\tilde{W}_2(u) - \tilde{W}_1(u))^+ = -B_{1,N_1(s,v)} = -X_{1,N_1(s,v)}$, hence for $i=2$ it holds as well.

Finally, assume that $\tilde{W}_1(u) \leq \tilde{W}_2(u)$ and $B_{1,N_1(s,v)} \geq \tilde{W}_2(u) - \tilde{W}_1(u)$. Hence at time v we have $\tilde{W}_1(v) \geq \tilde{W}_2(v)$. Now $X_{1,N_1(s,v)} = \tilde{W}_2(u) - \tilde{W}_1(u)$ and $Y_{1,N_1(s,v)} = B_{1,N_1(s,v)} - (\tilde{W}_2(u) - \tilde{W}_1(u))$. Then $(\tilde{W}_1(v) - \tilde{W}_2(v))^+ - (\tilde{W}_1(u) - \tilde{W}_2(u))^+ = \tilde{W}_1(v) - \tilde{W}_2(v) + 0 = \tilde{W}_1(u) - \tilde{W}_2(u) + B_{1,N_1(s,v)} = Y_{1,N_1(s,v)}$. Hence for $i=1$, equation (14) holds. Furthermore, $(\tilde{W}_2(v) - \tilde{W}_1(v))^+ - (\tilde{W}_2(u) - \tilde{W}_1(u))^+ = -\tilde{W}_2(u) - \tilde{W}_1(u) = -X_{1,N_1(s,v)}$, hence (14) holds for $i=2$ as well. □