Fluid queues with synchronized output

Operations Research Letters 47 (2019) 629–635

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Fluid queues with synchronized output

Onno Boxma [b], Offer Kella [a], Liron Ravner [b,c,*]

[a] Department of Statistics, The Hebrew University of Jerusalem, Israel
[b] Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands
[c] KDVI, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 10 May 2019
Received in revised form 10 September 2019
Accepted 15 October 2019
Available online 23 October 2019

Keywords:
Coupled queues
Capacity allocation
Lévy driven queues
Non-cooperative games

ABSTRACT

We present a model of parallel Lévy-driven queues that mix their output into a final product; whatever cannot be mixed is sold on the open market for a lower price. The queues incur holding and capacity costs and can choose their processing rates. We solve the ensuing centralized (system optimal) and decentralized (individual station optimal) profit optimization problems. In equilibrium the queues process work faster than desirable from a system point of view. Several model extensions are also discussed.

1. Introduction

In many production and assembly systems multiple components, produced in different locations, are combined to produce a final product. This is often termed as kitting process (see [18]), where kits are composed of different items produced by different machines. We study a system of two coupled fluid queues with independent Lévy input (see [7]) which continuously mix their output whenever possible. The mixed output is sold for a high price while the unmixed product from each station can also be sold, but for some lower price. We assume that the processing rates can be chosen with the goal of maximizing the profit in the presence of capacity and holding costs.

Multiqueue systems with some coupling between the queues are only rarely tractable. There are some two-queue exceptions; see, e.g., the pioneering paper of Fayolle and Iasnogorodski [9] on two coupled processors, the books [5,6,10] on boundary value problems for two-dimensional random walks and queues, and the PhD thesis of Blanc [2]. In the latter thesis, cf. also [3], Blanc studies a single server who serves two queues. If, upon completion of a service, both queues are non-empty, the server serves a pair of customers; if only one queue is non-empty, a customer of that queue is being served (if the server would wait for an arrival to the empty queue, the system would never reach steady state). He obtains the generating function of the two-dimensional queue length distribution by solving a Riemann–Hilbert boundary value problem.

For our problem, such an approach is not suitable because we aim at profit optimization by using explicit workload expressions. In view of the complexity of the above-mentioned multiqueue models, numerical methods are usually employed for their analysis, or some relaxing assumptions are made (e.g. [18,19] and [8]). Approximations and asymptotic techniques have also been used (e.g. [4]). Our fluid model can also be seen as a tractable approximation for a discrete system, whose performance analysis is otherwise intractable.

The economic analysis involves maximizing utility functions that include profit from throughput and costs incurred from capacity allocation and holding storage content. This is a classical problem in queueing analysis, see for example [16] and [20]. The decentralized analysis touches upon the issue of cooperation, or lack of, between servers, which is surveyed in Chapter 8.1.2 of [11].

The contribution of this paper is providing a tractable model that enables explicit performance analysis of a system of queues with coupled output. The model is still general in the sense that it assumes subordinator Lévy input, which includes the classical M/G/1 queue. For two stations with pure jump input processes we derive the centralized solution when both rates are chosen by a single system administrator, and the decentralized solution (Nash equilibrium) when stations choose their rates independently. We find that in equilibrium the stations over-invest in capacity resulting in more idle time than desired from a system point of view. Note that if only one of the queues is idle then the output of the busy queue is not mixed and potential higher gains are lost.

2. Model and preliminaries

In what follows, we denote $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, $a^+ = a \vee 0$, $a^- = a \wedge 0$.

* Corresponding author.
E-mail address: l.ravner@uva.nl (L. Ravner).

https://doi.org/10.1016/j.orl.2019.10.007
0167-6377/

For $i = 1, 2$ let $J_i(\cdot)$ be two independent subordinators (nondecreasing Lévy processes) starting from zero with Laplace exponents $-\eta_i(\alpha)$ satisfying for $\alpha \geq 0$,

$$\eta_i(\alpha) = c_i\alpha + \int_{(0,\infty)} \left(1 - e^{-\alpha x}\right) \nu_i(dx) = \alpha\left(c_i + \int_0^\infty e^{-\alpha x}\nu_i(x,\infty)dx\right),$$

with $c_i \geq 0$ and ν_i a (Lévy) measure necessarily satisfying $\int_{(0,\infty)}(x \wedge 1)\,\nu_i(dx) < \infty$. Let

$$\rho_i = \mathrm{E}J_i(1) = \eta_i'(0) = c_i + \int_0^\infty \nu_i(x,\infty)dx,$$

where $\eta_i'(0) \equiv \eta_i'(0+)$ is the limit from the right at $\alpha = 0$; we assume that $\rho_i < \infty$.

Also let $Z_1(0), Z_2(0)$ be independent nonnegative random variables which are also independent of the Lévy processes. For $r_i > \rho_i$ let

$$L_i(t) = -\inf_{0\leq s\leq t}(Z_i(0) + J_i(s) - r_i s)^-,$$
$$Z_i(t) = Z_i(0) + J_i(t) - r_i t + L_i(t).$$

Then, $Z_1(\cdot)$ and $Z_2(\cdot)$ are independent (Markov) processes and it is well known that they have a stationary/limiting/ergodic distribution so that if Z_i^* is a random variable having this distribution then for $\alpha \geq 0$ and $i = 1, 2$ (see, e.g., [15]),

$$\mathrm{E}e^{-\alpha Z_i^*} = \frac{(r_i - \rho_i)\alpha}{r_i\alpha - \eta_i(\alpha)} = \frac{1 - \frac{\rho_i}{r_i}}{1 - \frac{\rho_i}{r_i}\eta_{ei}(\alpha)},$$

where

$$\eta_{ei}(\alpha) = \frac{1}{\rho_i}\left(c_i + \int_0^\infty e^{-\alpha x}\nu_i(x,\infty)dx\right)$$

is the Laplace–Stieltjes transform (LST) of a $\left(\frac{c_i}{\rho_i}, 1 - \frac{c_i}{\rho_i}\right)$ mixture of zero and a distribution having the density $\nu_i(x,\infty)/\int_0^\infty \nu_i(y,\infty)\,dy$ (stationary residual jump sizes, see, e.g. [13]). In fact, it is clear from the formula and known (e.g. [14]) that Z_i^* has a compound geometric distribution. Notice that $Z_i(\cdot)$ can be interpreted as the workload in a queue/buffer with $J_i(\cdot)$ as input process and r_i as fixed service/outflow speed. A special case is the steady state distribution of the waiting time in an M/G/1 queue.

Now $0 < p_1 < 1$ units of station 1 need to be mixed with $p_2 = 1 - p_1$ units from station 2 in order to produce one unit of a new product. Station i will receive $K_i \geq 0$ for every unit which is mixed. Whatever output of station i that is not mixable is sold for a price of $k_i \geq 0$ per unit, and we assume $K_i > k_i$. It is clear that at every time unit only one of the two original products can be unmixable. Fig. 1 illustrates an example of the operation of such a system. We would like to compute the revenue rate from running such an operation and then optimize over r_1, r_2 under some additional cost structure.

In the above setup $L_i(t) = (r_i - c_i)\int_0^t 1_{\{Z_i(s)=0\}}ds$, and thus the cumulative (total) outflow is

$$O_i(t) = r_i t - L_i(t) = c_i\int_0^t 1_{\{Z_i(s)=0\}}ds + r_i\int_0^t 1_{\{Z_i(s)>0\}}ds.$$

This would be false if instead of a nondecreasing Lévy process minus a drift we would have a more general Lévy process with no negative jumps. That is, if either there is a Brownian component or $\int_{(0,1]} x\nu_i(dx) = \infty$, then it fails.

When the content levels of the two stations are positive, then the total output rate from station i is r_i, and after some sufficiently small amount of time $\epsilon > 0$, there has been a release of $r_1\epsilon$ and $r_2\epsilon$ from the two stations. Every x units of the mixed

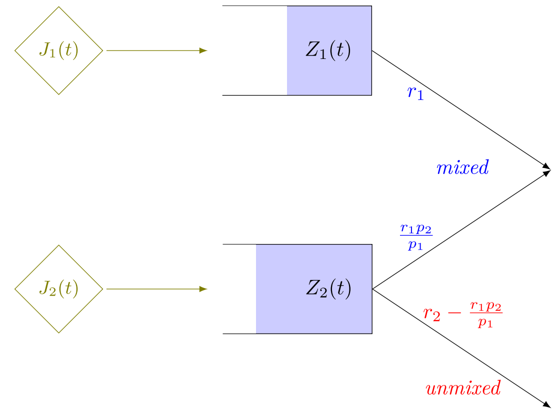

Fig. 1. The rates of mixed and unmixed output when both buffers are not empty, $J_1(t), J_2(t) > 0$, and $\frac{r_1}{p_1} < \frac{r_2}{p_2}$. All of the output of station 1 is mixed and station 2 mixes as much as it can and the remainder is unmixed.

product is combined of p_1x unit from station 1 and p_2x from station 2. Thus the total amount that can be mixed should satisfy $p_ix \leq r_i\epsilon$ for $i = 1, 2$ and thus the maximal amount is $x = \left(\frac{r_1}{p_1} \wedge \frac{r_2}{p_2}\right)\epsilon$ and thus the total output from station 1 that was mixed is $p_1x = r_1\epsilon\left(\frac{p_1r_2}{p_2r_1} \wedge 1\right)$ and similarly for station 2 it is $p_2x = r_2\epsilon\left(\frac{p_2r_1}{p_1r_2} \wedge 1\right)$. Note that in this case at least one of the stations is mixing its entire output while the other may not. Thus the mixed output rates from stations 1, 2 are either $(r_1, p_2r_1/p_1)$ or $(p_1r_2/p_2, r_2)$, depending on whether $\frac{r_1}{p_1} \leq \frac{r_2}{p_2}$ or $\frac{r_1}{p_1} \geq \frac{r_2}{p_2}$, respectively. In the first case, the unmixed product output from the second station is

$$r_2 - \frac{p_2r_1}{p_1} = p_2\left(\frac{r_2}{p_2} - \frac{r_1}{p_1}\right),$$

and, similarly, in the second case the unmixed product from the first station is

$$p_1\left(\frac{r_1}{p_1} - \frac{r_2}{p_2}\right).$$

If we do not want to break this into cases then the unmixed outputs from the two stations are

$$p_1\left(\frac{r_1}{p_1} - \frac{r_2}{p_2}\right)^+,\ p_2\left(\frac{r_2}{p_2} - \frac{r_1}{p_1}\right)^+,$$

respectively. The mixing dynamics are illustrated for an example in Fig. 1.

Thus, when the contents of both stations are positive, the revenue rate for station 1 is

$$g_1(r_1, r_2) = K_1\left(r_1 \wedge \frac{p_1r_2}{p_2}\right) + k_1p_1\left(\frac{r_1}{p_1} - \frac{r_2}{p_2}\right)^+, \tag{1}$$

and for station 2 it is

$$g_2(r_1, r_2) = K_2\left(\frac{p_2r_1}{p_1} \wedge r_2\right) + k_2p_2\left(\frac{r_2}{p_2} - \frac{r_1}{p_1}\right)^+. \tag{2}$$

Clearly, when the contents of both stations are zero, the total output rates are c_i (what continuously flows in immediately flows out) and thus the same computations lead to (1) and (2) where r_i are replaced by c_i. When the content of station 1(2) is positive and that of station 2(1) is zero then (1) and (2) hold with $c_2(c_1)$ replacing $r_2(r_1)$. The fraction of time station i is not empty is ρ_i/r_i.

Therefore, its total revenue rate is

$$f_i(r_1,r_2) = \frac{\rho_1}{r_1}\frac{\rho_2}{r_2}g_i(r_1,r_2) + \frac{\rho_1}{r_1}\left(1-\frac{\rho_2}{r_2}\right)g_i(r_1,c_2) + \left(1-\frac{\rho_1}{r_1}\right)\frac{\rho_2}{r_2}g_i(c_1,r_2) + \left(1-\frac{\rho_1}{r_1}\right)\left(1-\frac{\rho_2}{r_2}\right)g_i(c_1,c_2). \tag{3}$$

Now let us assume that for every unit of time, station i pays a_i for every unit of capacity. Also, let us assume that there is linear holding cost of h_i for station i. Denoting $\sigma_i^2 = -\eta_i''(0) = \int_0^\infty x^2\nu_i(dx)$, this results in a long run average cost of h_i multiplied by

$$EZ_i^* = \frac{\sigma_i^2}{2(r_i-\rho_i)},$$

where, from here on, we assume that $\sigma_i^2 < \infty$ for $i = 1, 2$. Thus, with $b_i = h_i\sigma_i^2/2$, the profit for station i is $f_i(r_1,r_2) - a_ir_i - \frac{b_i}{r_i-\rho_i}$. Note that when $c_1 = c_2 = 0$ we have that $g_1(r_1,0) = k_1r_1$, $g_1(0,r_2) = g_2(r_1,0) = 0$ and $g_2(0,r_2) = k_2r_2$, and in particular $g_1(0,0) = g_2(0,0) = 0$. Moreover, if in addition we assume, without loss of generality, that $\frac{r_1}{p_1} \le \frac{r_2}{p_2}$, then

$$g_1(r_1,r_2) = K_1r_1, \quad g_2(r_1,r_2) = K_2\frac{r_1p_2}{p_1} + k_2\left(r_2 - \frac{r_1p_2}{p_1}\right).$$

We can now insert these values in (3) and obtain the desired expressions for this case which become a simpler expression.

3. System optimization

In this section we characterize the optimal processing rates that maximize the profit of the coupled systems for the special case of no external linear input; $c_1 = c_2 = 0$. We first of all do this for a centralized system in which both queues are controlled by a single entity. This is followed by a decentralized analysis where each queue can independently choose a rate that maximizes its own profit.

The stationary revenue per unit of time for system 1, for $r_1 > \rho_1$ and $r_2 > \rho_2$, is obtained by applying Eq. (3),

$$f_1(r_1,r_2) = \frac{\rho_1}{r_1}\frac{\rho_2}{r_2}g_1(r_1,r_2) + \frac{\rho_1}{r_1}\left(1-\frac{\rho_2}{r_2}\right)g_1(r_1,0),$$

and the stationary utility, which is the net profit after deduction of costs from the revenue, per unit of time is

$$u_1(r_1,r_2) := f_1(r_1,r_2) - a_1r_1 - \frac{b_1}{r_1-\rho_1}.$$

Similarly, we have that the stationary revenue per unit of time for system 2 is

$$f_2(r_1,r_2) = \frac{\rho_1}{r_1}\frac{\rho_2}{r_2}g_2(r_1,r_2) + \frac{\rho_2}{r_2}\left(1-\frac{\rho_1}{r_1}\right)g_1(0,r_2),$$

and the stationary profit per unit of time is

$$u_2(r_1,r_2) := f_2(r_1,r_2) - a_2r_2 - \frac{b_2}{r_2-\rho_2}.$$

Note that due to the breakpoints in the functions g_i, as defined in (1) and (2), u_i are piecewise-smooth nonlinear functions of the processing rates r_1 and r_2.

Suppose both queues have a single controller, then clearly an optimal solution will satisfy $r_1 > \rho_1$ and $r_2 > \rho_2$. The system optimal rates (r_1^*, r_2^*) are given by the non-linear program,

$$\begin{aligned} &\max\{u_1(r_1,r_2)+u_2(r_1,r_2)\}, \\ &\text{s.t.} \quad r_1>\rho_1,\ r_2>\rho_2. \end{aligned} \tag{4}$$

The decentralized solution we are interested in is a Nash equilibrium: a pair (r_1^e, r_2^e) such that

$$r_1^e \in \underset{r_1>\rho_1}{\operatorname{argmax}}\left\{u_1(r_1,r_2^e)\right\}, \quad r_2^e \in \underset{r_2>\rho_2}{\operatorname{argmax}}\left\{u_2(r_1^e,r_2)\right\}. \tag{5}$$

We first state a lemma that will be useful in characterizing both the optimal and equilibrium rates. As will be detailed in the next sections, the utilities u_i generally have the form $B - \frac{A}{r_i}$ with respect to r_i, where A and B are constants given by the specific parameters in each case.

Lemma 1. *For any positive constants (a, b, c, d, e, f), if we consider the three equations*

$$-a + \frac{b}{(x-d)^2} = 0, \tag{6}$$

$$-\frac{c}{x^2} - a + \frac{b}{(x-d)^2} = 0, \tag{7}$$

$$-\frac{c}{x^2} - a + \frac{b}{(x-d)^2} + \frac{e}{(x-f)^2} = 0, \tag{8}$$

then (6) has a unique solution $x_1 = d + \sqrt{\frac{b}{a}}$ in the interval (d,∞), (7) has a unique solution x_2 in the interval (d,∞), (8) has a unique solution x_3 in the interval $(d \vee f, \infty)$, and $x_2 < x_1$.

Proof. Straightforward computation yields that (6) has a unique solution $x_1 = d + \sqrt{\frac{b}{a}}$. We will first show the uniqueness of the solution to (8), which further implies a unique solution to (7) by taking $e \to 0$. If $d < f$ then we rewrite (8) as

$$b + e\left(\frac{x-d}{x-f}\right)^2 = a(x-d)^2 + c\left(1-\frac{d}{x}\right)^2.$$

The right hand side is increasing with x for $x > d$ and as $d < f$ the left hand side is decreasing. Moreover, as the left hand side is also unbounded as $x \to f$, there exists exactly one solution. If $d \ge f$ then, similarly, we rewrite (8) as

$$b\left(\frac{x-f}{x-d}\right)^2 + e = a(x-f)^2 + c\left(1-\frac{f}{x}\right)^2,$$

and applying the same argument yields the existence and uniqueness. If x_1 is the solution of (6) then $-a + \frac{b}{(x-d)^2} \le 0$ for all $x \ge x_1$, hence $-\frac{c}{x^2} - a + \frac{b}{(x-d)^2} < 0$ for all $x \ge x_1$, and therefore $x_2 < x_1$. □

3.1. *Centralized optimization ($c_1 = c_2 = 0$)*

From now we assume that $r_1 > \rho_1$ and $r_2 > \rho_2$, as these are the feasible solutions of (4). Applying (1) and (2) yields $u_1(r_1,r_2) =$

$$\begin{cases} \rho_1\left(\frac{\rho_2}{r_2}(K_1-k_1)+k_1\right) - a_1r_1 - \frac{b_1}{r_1-\rho_1}, & r_1 \le \frac{r_2p_1}{p_2}, \\ \rho_1\left(\frac{\rho_2p_1}{r_1p_2}(K_1-k_1)+k_1\right) - a_1r_1 - \frac{b_1}{r_1-\rho_1}, & r_1 > \frac{r_2p_1}{p_2}, \end{cases} \tag{9}$$

and $u_2(r_1,r_2) =$

$$\begin{cases} \rho_2\left(\frac{\rho_1p_2}{r_2p_1}(K_2-k_2)+k_2\right) - a_2r_2 - \frac{b_2}{r_2-\rho_2}, & r_2 \ge \frac{r_1p_2}{p_1}, \\ \rho_2\left(\frac{\rho_1}{r_1}(K_2-k_2)+k_2\right) - a_2r_2 - \frac{b_2}{r_2-\rho_2}, & r_2 < \frac{r_1p_2}{p_1}. \end{cases} \tag{10}$$

Let $U(r_1,r_2) := u_1(r_1,r_2) + u_2(r_1,r_2)$, then for $r_1 \le \frac{r_2p_1}{p_2}$,

$$U(r_1,r_2) = \rho_1\left(\frac{\rho_2}{r_2}(K_1-k_1)+k_1\right) + \rho_2\left(\frac{\rho_1p_2}{r_2p_1}(K_2-k_2)+k_2\right) - \sum_{i=1}^{2}\left[a_ir_i + \frac{b_i}{r_i-\rho_i}\right],$$

and for $r_1 > \frac{r_2 p_1}{p_2}$,

$$U(r_1,r_2) = \rho_1\left(\frac{\rho_2 p_1}{r_1 p_2}(K_1-k_1)+k_1\right)+ \rho_2\left(\frac{\rho_1}{r_1}(K_2-k_2)+k_2\right)-\sum_{i=1}^{2}\left[a_i r_i+\frac{b_i}{r_i-\rho_i}\right].$$

By taking derivatives,

$$\frac{d}{dr_1}U(r_1,r_2) = \begin{cases} -a_1+\frac{b_1}{(r_1-\rho_1)^2}, & r_1 \le \frac{r_2 p_1}{p_2}, \\ -\frac{A_1}{r_1^2}-a_1+\frac{b_1}{(r_1-\rho_1)^2}, & r_1 > \frac{r_2 p_1}{p_2}, \end{cases} \tag{11}$$

where $A_1 := \rho_1\rho_2\left(\frac{p_1}{p_2}(K_1-k_1)+(K_2-k_2)\right)$. Similarly,

$$\frac{d}{dr_2}U(r_1,r_2) = \begin{cases} -\frac{A_2}{r_2^2}-a_2+\frac{b_2}{(r_2-\rho_2)^2}, & r_2 \ge \frac{r_1 p_2}{p_1}, \\ -a_2+\frac{b_2}{(r_2-\rho_2)^2}, & r_2 < \frac{r_1 p_2}{p_1}, \end{cases} \tag{12}$$

where $A_2 := \rho_1\rho_2\left((K_1-k_1)+\frac{p_2}{p_1}(K_2-k_2)\right)$.

Note that $\frac{d}{dr_i}U(r_1,r_2) = 0$ has the form of either Eq. (6) or (7) in Lemma 1 which implies that it has a unique solution on (ρ_i,∞). Furthermore, $u_i(r_1,r_2)\to-\infty$ as $r_i\to\rho_i$ or $r_i\to\infty$ and therefore the solution must be a local maximum and we can solve (4) without the constraints. There are two possible types of optimal solutions (r_1^*,r_2^*) for this piecewise nonlinear optimization problem:

1. An interior solution such that either $\frac{r_1^*}{p_1} > \frac{r_2^*}{p_2}$ or $\frac{r_1^*}{p_1} < \frac{r_2^*}{p_2}$, and both rates are local minima that solve the first order conditions in the respective ranges.
2. A boundary solution such that $\frac{r_1^*}{p_1} = \frac{r_2^*}{p_2}$. In this case the partial derivatives of the objective function at (r_1^*,r_2^*) are strictly negative for both coordinates.

Observe that in the first case of (9) (second case of (10)) the revenue term of system 1 (2) is not a function of r_1 (r_2), hence if the optimal rate is in this range it is given by $\frac{d}{dr_i}U(r_1,r_2) = 0$ or equivalently $a_i = \frac{b_i}{(r_i-\rho_i)^2}$, which yields $\bar{r}_i := \rho_i+\sqrt{\frac{b_i}{a_i}}$. Unsurprisingly, $\rho+\sqrt{\frac{b}{a}}$ is the optimal rate that balances costly service rate and holding costs in an M/G/1 queue (see p.329 of [16] for a proof for the M/M/1 case which extends trivially to M/G/1).

For the second case of (9) (first case of (10)) the revenue term of system 1 (2) is a function of its own rate. If the optimal rate r_i is in this range then it is given by a solution to the first order condition $\frac{d}{dr_i}U(r_1,r_2) = 0$. Let $\tilde{r}_i$ denote this solution and by Lemma 1 we have that it is unique and that $\tilde{r}_i\in(\rho_i,\bar{r}_i)$.

The optimal solution may also be on the boundary, $\frac{r_1^*}{p_1} = \frac{r_2^*}{p_2}$. In this case the optimal solution solves the one-dimensional problem $\max_{r>\left(\rho_1\wedge\frac{\rho_2 p_1}{p_2}\right)}\left\{U\left(r,\frac{rp_2}{p_1}\right)\right\}$. The objective function is now $U(r) := U\left(r,\frac{rp_2}{p_1}\right)$, hence

$$U(r) = \rho_1 k_1+\rho_2 k_2+\frac{A_1}{r}-\left(a_1+\frac{a_2 p_2}{p_1}\right)r-\frac{b_1}{r-\rho_1}-\frac{\frac{b_2 p_1}{p_2}}{r-\frac{\rho_2 p_1}{p_2}}.$$

As $U(r)\to-\infty$ as $r\to\left(\rho_1\wedge\frac{\rho_2 p_1}{p_2}\right)$, if there is a unique solution to $\frac{d}{dr}U(r) = 0$, then it is a local and global maximum. The condition can be written as

$$\frac{b_1}{(r-\rho_1)^2}+\frac{\frac{b_2 p_1}{p_2}}{\left(r-\frac{\rho_2 p_1}{p_2}\right)^2} = a_1+\frac{a_2 p_2}{p_1}+\frac{A_1}{r^2}, \tag{13}$$

and by Lemma 1 there is a unique solution that we denote by $\hat{r}$.

In conclusion, we can compute the optimal rates by computing all roots $(\bar{r}_1,\bar{r}_2,\tilde{r}_1,\tilde{r}_2,\hat{r})$ and

$$(r_1^*,r_2^*) = \underset{(r_1,r_2)\in\left\{(\bar{r}_1,\tilde{r}_2),(\tilde{r}_1,\bar{r}_2),\left(\hat{r},\frac{\hat{r}p_2}{p_1}\right)\right\}}{\operatorname{argmax}} U(r_1,r_2).$$

3.2. Decentralized optimization ($c_1 = c_2 = 0$)

We now consider a competitive setting where each station is controlled by a different player that can set his rate aiming to maximize the expected profit of his own station. The goal is to derive all Nash equilibria satisfying (5).

Given r_2 the best response for player 1 is in exactly one of the two possible cases of (11). As before, the revenue term of queue 1 is not a function of r_1 in the first case, hence the optimal rate is $\bar{r}_1 = \rho_1+\sqrt{\frac{b_1}{a_1}}$. The second case is also similar to the system optimization with the only difference that the effect on the other queue's revenue is not taken into account. In particular, if $r_1 \ge \frac{r_2 p_1}{p_2}$,

$$\frac{d}{dr_1}u_1(r_1,r_2) = -\frac{\rho_1\rho_2\frac{p_1}{p_2}(K_1-k_1)}{r_1^2}-a_1+\frac{b_1}{(r_1-\rho_1)^2},$$

and similarly by (12), if $r_2 \ge \frac{r_1 p_2}{p_1}$,

$$\frac{d}{dr_2}u_2(r_1,r_2) = -\frac{\rho_1\rho_2\frac{p_2}{p_1}(K_2-k_2)}{r_2^2}-a_2+\frac{b_2}{(r_2-\rho_2)^2}.$$

Observe that for $i = 1,2$, $\frac{d}{dr_i}u_i(r_1,r_2) = 0$ has the same form as (7) in Lemma 1, and therefore has a single real root in (ρ_i,∞) and we denote this root by $\check{r}_i$, for $i = 1,2$.

Both pairs $(\bar{r}_1,\check{r}_2)$ and $(\check{r}_1,\bar{r}_2)$ are candidates for equilibrium. There may, however, be other types of equilibria on the boundary set $\mathcal{R} := \{(r_1,r_2) : \frac{r_2}{p_2} = \frac{r_1}{p_1}\}$. If

$$\lim_{r_1\uparrow r}\frac{d}{dr_1}u_1\left(r_1,\frac{rp_2}{p_1}\right)\ge 0,\quad \lim_{r_1\downarrow r}\frac{d}{dr_1}u_1\left(r_1,\frac{rp_2}{p_1}\right)\le 0, \tag{14}$$

then there is no incentive for system 1 to deviate. Let $\mathcal{R}_1$ denote the set of rates satisfying condition (14). By applying (11), we have that for every $r\in\mathcal{R}_1$,

$$0\le -a_1+\frac{b_1}{(r-\rho_1)^2}\le\frac{\rho_1\rho_2\frac{p_1}{p_2}(K_1-k_1)}{r^2}, \tag{15}$$

and hence $\mathcal{R}_1 = [\check{r}_1,\bar{r}_1]$, by Lemma 1. If, in addition,

$$\lim_{r_2\uparrow\frac{rp_2}{p_1}}\frac{d}{dr_2}u_2(r,r_2)\ge 0,\quad \lim_{r_2\downarrow\frac{rp_2}{p_1}}\frac{d}{dr_2}u_2(r,r_2)\le 0, \tag{16}$$

then the pair $\left(r,\frac{rp_2}{p_1}\right)$ is a Nash equilibrium. By applying (12) the condition (16) yields

$$0\le -a_2+\frac{b_2}{(\frac{rp_2}{p_1}-\rho_1)^2}\le\frac{\rho_1\rho_2\frac{p_2}{p_1}(K_2-k_2)}{\left(\frac{rp_2}{p_1}\right)^2},$$

which is equivalent to $\mathcal{R}_2 = \{r : \frac{rp_2}{p_1}\in[\check{r}_2,\bar{r}_2]\}$. Therefore, the set of all Nash equilibria on the boundary is given by $\mathcal{R}_e := \left\{\left(r,\frac{rp_2}{p_1}\right) : r\in\mathcal{R}_1\cap\mathcal{R}_2\right\}$. Hence a Nash equilibrium may have a similar form as the optimal solution of the previous section, but there are also cases when there is a continuum of Nash equilibria on the boundary. We summarize our analysis in Proposition 2 that provides a characterization of the Nash equilibria in terms of the roots $(\bar{r}_1,\bar{r}_2,\check{r}_1,\check{r}_2)$ and also a simple recipe for their computation.

Proposition 2. *The Nash equilibrium rates* (r_1^e, r_2^e) *satisfy:*

1. *If* $\frac{p_1}{p_2}\bar{r}_2 < \check{r}_1$ *then* $\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$ *and* $(r_1^e, r_2^e) = (\check{r}_1, \bar{r}_2)$ *is the unique Nash equilibrium.*
2. *If* $\frac{p_1}{p_2}\check{r}_2 > \bar{r}_1$ *then* $\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$ *and* $(r_1^e, r_2^e) = (\bar{r}_1, \check{r}_2)$ *is the unique Nash equilibrium.*
3. *If* $\frac{p_1}{p_2}\bar{r}_2 \geq \check{r}_1$ *and* $\frac{p_1}{p_2}\check{r}_2 \leq \bar{r}_1$ *then* $\mathcal{R}_1 \cap \mathcal{R}_2 \neq \emptyset$ *and* $\mathcal{R}_e$ *is the set of all Nash equilibria.*

Proof. 1. If $\frac{p_1}{p_2}\bar{r}_2 < \check{r}_1$ then $(\check{r}_1, \bar{r}_2)$ is a Nash equilibrium because both stations are using the optimal rate in the correct range. Moreover, for any $r \in \mathcal{R}_2$ we have that $\frac{rp_2}{p_1} \leq \bar{r}_2$, hence $r \leq \frac{p_1}{p_2}\bar{r}_2 < \check{r}_1$ and therefore $r \notin \mathcal{R}_1$. We conclude that $\mathcal{R}_e = \mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$.

2. Similarly, if $\frac{p_1}{p_2}\check{r}_2 > \bar{r}_1$ then $(\check{r}_1, \bar{r}_2)$ is a Nash equilibrium. If $r \in \mathcal{R}_1$ then $r \leq \bar{r}_1 < \frac{p_1}{p_2}\check{r}_2$, hence $r \notin \mathcal{R}_2$ and $\mathcal{R}_e = \mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$.

3. If $\frac{p_1}{p_2}\bar{r}_2 \geq \check{r}_1$ and $\frac{p_1}{p_2}\check{r}_2 \leq \bar{r}_1$ then for every $r \in \mathcal{R}_2$,

$$\check{r}_1 \leq \frac{p_1}{p_2}\bar{r}_2 \leq r \leq \frac{p_1}{p_2}\check{r}_2 \leq \bar{r}_1,$$

and thus $r \in \mathcal{R}_1$. This implies that $\mathcal{R}_2 \subseteq \mathcal{R}_1$ and $\mathcal{R}_2 \neq \emptyset$ because $\check{r}_2 \leq \bar{r}_2$ by Lemma 1. □

A closer look at the first order conditions enables a comparison of the optimal and the interior equilibrium solutions. In particular, for the case of $r_1 > \frac{r_2 p_1}{p_2}$, the term $\frac{\rho_1\rho_2}{r_1}(K_2 - k_2)$ in $U(r_1, r_2)$ corresponds to a positive externality that queue 1 imposes on the profit of queue 2 which is not taken into account in the individual optimization. Proposition 3 asserts that this externality always leads to higher rates in equilibrium than in the system optimal solution. This means that in equilibrium the faster server over-invests in capacity which results in longer idle periods and loss of profit for the slow server, which outweighs the gains made by the fast server. For the sake of the explanation assume that 1 is the faster server in terms of mixing, i.e., $r_1 \geq r_2\frac{p_1}{p_2}$, then in equilibrium server 1 has a strictly higher processing rate than the optimal rate. This means that in equilibrium both servers extract more revenue when they are both working but the idle periods of server 1 are longer on average. Therefore, for server 2 there are longer periods such that it is busy but cannot mix because server 1 is idle. However, for server 1 the proportion of its busy time such that server 2 is idle stays the same and so the higher rate yields a better balance between the capacity cost and revenue stream. By definition the equilibrium is not system optimal and so the total "gain" of the system from the individual choice of server 1 is negative, and in particular the negative externality imposed on server 2 due to more idle time is bigger than the positive externality due to a higher revenue rate when both are busy.

Proposition 3. *The Nash equilibrium processing rates are at least as high as the system optimal rates:* $r_i^e \geq r_i^*$ *for* $i = 1, 2$. *Furthermore, there always exists an equilibrium such that* $r_i^e > r_i^*$ *for at least one* $i = 1, 2$.

Proof. First (i) we show that $\tilde{r}_i < \check{r}_i$ for $i = 1, 2$. This implies the result for interior optimal and equilibrium rates. Next (ii) we will show that if the optimal solution is in the interior then so is the Nash equilibrium and $\tilde{r}_i < \check{r}_i$ for at least one $i = 1, 2$. Finally (iii) we will show that an optimal solution on the boundary yields lower rates than all possible equilibria.

(i) By (11), $\frac{d}{dr_1}u_1(\check{r}_1, r_2) = 0$ is equivalent to

$$-\frac{\rho_1\rho_2\frac{p_1}{p_2}(K_1 - k_1)}{\check{r}_1^2} - a_1 + \frac{b_1}{(\check{r}_1 - \rho_1)^2} = 0,$$

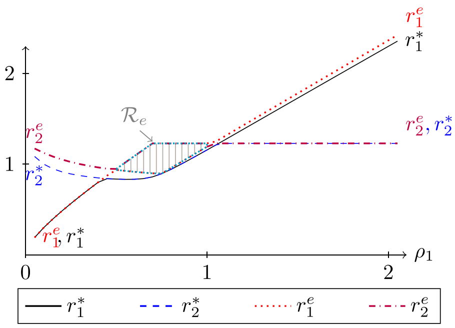

Fig. 2. The optimal rates (r_1^*, r_2^*) and equilibrium rates (r_1^e, r_2^e) as a function of ρ_1 when all other parameters are fixed: $K_1 = 3$, $K_2 = 2$, $k_1 = k_2 = 1$, $p_1 = p_2 = 0.5$, $\rho_2 = 0.7$, $b_1 = b_2 = 0.08$, $c_1 = c_2 = 0$, $a_1 = a_2 = 0.1$. Every vertical line in the shaded region $\mathcal{R}_e$ corresponds to a continuum of equilibrium points (r, r).

hence by (12)

$$\frac{d}{dr_1}U(\check{r}_1, r_2) = -\frac{\rho_1\rho_2(K_2 - k_2)}{\check{r}_1^2} < 0.$$

Lemma 1 implies that $\tilde{r}_1$ is the unique local (and global) maximum of $U(r_1, r_2)$ with respect to $r_1 \in (\rho_1, \infty)$, and therefore $\tilde{r}_1 < \check{r}_1$. The same argument yields that $\tilde{r}_2 < \check{r}_2$.

(ii) If $(\tilde{r}_1, \bar{r}_2)$ is the optimal solution then $\check{r}_1 > \tilde{r}_1 \geq \frac{p_2}{p_1}\bar{r}_2$, hence $(\check{r}_1, \bar{r}_2)$ is individually optimal for both stations, and thus a Nash equilibrium. Furthermore, by Proposition 2, $\check{r}_1 \geq \frac{p_2}{p_1}\bar{r}_2$ implies that $\mathcal{R}_2$ is either an empty set when the inequality is strict, or the singleton $\check{r}_1$ when there is equality. Therefore, $(\check{r}_1, \bar{r}_2)$ is the unique Nash equilibrium and it satisfies $r_1^* > r_1^e$ and $r_2^* = r_2^e$. The same argument yields the result for $(r_1^*, r_2^*) = (\bar{r}_1, \tilde{r}_2)$ and $(r_1^e, r_2^e) = (\bar{r}_1, \check{r}_2)$.

(iii) If (r_1^*, r_2^*) is on the boundary, i.e., $r_1^* = r$ and $r_2^* = \frac{p_2}{p_1}r$ for some r, then by (13),

$$-a_1 - \frac{A_1}{r^2} + \frac{b_1}{(r - \rho_1)^2} = -\frac{p_2}{p_1}(-a_2 + \frac{b_2}{(\frac{p_2}{p_1}r - \rho_2)^2}). \tag{17}$$

The left hand side of (17) equals $\frac{d}{dr_1}U(r, r_2)$ in the case of $r \leq \frac{p_1}{p_2}r_2$ and optimality implies that it is positive, otherwise lowering r would increase the total profit. Therefore, both sides of (17) are positive. Lemma 1 implies that $r \leq \tilde{r}_1 < \check{r}_1 = \min\{s : s \in \mathcal{R}_1\}$ and $r \leq \frac{p_1}{p_2}\check{r}_2 = \min\{s : s \in \mathcal{R}_2\}$. Therefore, the optimal rates are lower than those in all possible equilibria, with strict inequality for all rates in the interior. □

In Fig. 2 the optimal and equilibrium rates are illustrated for an example with varying ρ_1 and all other parameters fixed. For low levels of ρ_1 the optimal and equilibrium rate of the first station is $\bar{r}_1$ (i.e., $r_1^e = r_1^* = \bar{r}_1$), and the equilibrium rate in station 2 is higher than the optimal rate, as discussed previously. For moderate levels of ρ_1 we observe boundary solutions where $r_1^* = r_2^*$, and a continuum of equilibria such that $r_1^e = r_2^e$ for an interval $\mathcal{R}_e$. For higher ρ_1—levels the optimal and equilibrium rates of station 2 are lower and both equal $\bar{r}_2$, while the optimal rate at station 1 is a bit lower than the equilibrium rate.

4. Extensions

There are several natural extensions to consider for the above discussed model and the associated system optimization problem. First of all, the network structure may be more elaborate with additional stations for storage of the mixed and unmixed outputs. Moreover, one may be interested in a system of $n > 2$ stations that can potentially mix their output. If there are multiple stations then there are multiple ways to define the mixing process, e.g., all have to be working to mix, any two stations can mix, or some k out of the n stations are required for mixing. In Section 3 we assumed that there is no linear input to the stations ($c_1 = c_2 = 0$), but if this is not the case then the optimization analysis is more involved.

In this section we provide some details on three important extensions: (1) considering the joint distribution of the storage processes in the two stations with an additional storage station of the mixed product, (2) allowing the external inputs to be general subordinators with positive linear inputs $c_1, c_2 > 0$, and (3) $n > 2$ stations that potentially mix their products for a higher valued product.

4.1. Steady state distribution of the mixed product storage process

In this subsection we assume that the mixed product is transferred (instantly) to a third station that has its own processing rate. The goal is to compute, under some further restriction, the joint LST of steady state contents in the stations.

In particular, here we assume that $c_1 > 0$, $\nu_1 = 0$ and $c_2 = 0$: the input to station 1 is a positive linear flow and the input to station 2 is a pure jump subordinator. In this case, $\rho_1 = c_1$ and $\rho_2 = \int_{(0,\infty)} x\nu_2(dx)$. We further assume that the output of the mixed product flows into a third station which releases fluid at rate r_3. Since the first station becomes empty and remains so after some finite time, for the purpose of steady state behavior we may assume that $Z_1(0) = 0$ and thus $Z_1(t) = 0$ for all $t \geq 0$. Hence, the output from station 1 is linear at rate c_1. Since the input to station 2 is a pure jump subordinator, the output rate from station 2 is r_2 when it is not empty and zero when it is. Therefore, the flow into station 3 is $c_3 := \frac{c_1}{p_1} \wedge \frac{r_2}{p_2}$ when station 2 is not empty and zero when it is. The fraction of time station 2 is not empty is ρ_2/r_2. Thus the long run average input rate into station 3 is $\frac{\rho_2}{r_2}c_3$. Thus, for stability we need to assume that r_3 is larger than this quantity. Now, if $r_3 \geq c_3$, then eventually station 3 will become empty and remain empty from that time on. This is a trivial case. Hence, we assume that $\frac{\rho_2}{r_2}c_3 < r_3 < c_3$.

Since in this case $dL_2(t) = r_2 1_{\{Z_2(t)=0\}}dt$, we now observe that the third station behaves as follows:

$$Z_3(t) = Z_3(0) + c_3 \int_0^t 1_{\{Z_2(s)>0\}} ds - r_3 t + L_3(t)$$
$$= Z_3(0) + c_3(t - r_2^{-1}L_2(t)) - r_3 t + L_3(t),$$

so that we have

$$Z_2(t) = Z_2(0) + J_2(t) - r_2 t + L_2(t),$$
$$Z_3(t) = Z_3(0) + (c_3 - r_3)\, t - c_3 r_2^{-1} L_2(t) + L_3(t),$$

where $L_3(t) = -\inf_{0\leq s\leq t} \left(Z_3(0) + (c_3 - r_3)s - c_3 r_2^{-1}L_2(s)\right)^-$. Letting $\tilde{Z}_3(t) := \frac{Z_3(t)}{c_3 r_2^{-1}}$, $\tilde{L}_3(t) := \frac{L_3(t)}{c_3 r_2^{-1}}$ and $\tilde{c}_3 := \frac{c_3 - r_3}{c_3 r_2^{-1}} = r_2\left(1 - \frac{r_3}{c_3}\right) > 0$ gives that

$$\tilde{Z}_3(t) = \tilde{Z}_3(0) + \tilde{c}_3 t - L_2(t) + \tilde{L}_3(t).$$

Therefore, the joint structure of $(Z_2(\cdot), \tilde{Z}_3(\cdot))$ is a special case of the model considered in Section 4 of [12], where the joint steady state LST and the covariance structure were computed explicitly. Since this only requires substitution in the formulas given there, we omit the details.

4.2. Positive linear inputs: $c_1, c_2 > 0$

The stationary revenue per unit of time for system 1, for $r_1 > \rho_1$ and $r_2 > \rho_2$, is obtained by applying Eqs. (1)–(3). There are six distinct cases:

1. If $c_1 < r_1 \leq \frac{c_2 p_1}{p_2} < \frac{r_2 p_1}{p_2}$, then

$$f_1(r_1, r_2) = K_1\left(\rho_1 + \left(1 - \frac{\rho_1}{r_1}\right)c_1\right),$$

2. if $c_1 < \frac{c_2 p_1}{p_2} \leq r_1 < \frac{r_2 p_1}{p_2}$, then

$$f_1(r_1, r_2) = \rho_1\left(\frac{\rho_2}{r_2}(K_1 - k_1) + k_1\right) + K_1 c_1 - \frac{\rho_1}{r_1}\left(K_1 c_1 - \left(1 - \frac{\rho_2}{r_2}\right)(K_1 - k_1)\frac{p_1}{p_2}c_2\right),$$

3. if $\frac{c_2 p_1}{p_2} \leq c_1 < r_1 < \frac{r_2 p_1}{p_2}$, then

$$f_1(r_1, r_2) = \rho_1\left(\frac{\rho_2}{r_2}(K_1 - k_1) + k_1\right) + \frac{\rho_2}{r_2}K_1 c_1 + \left(1 - \frac{\rho_2}{r_2}\right)\left(k_1 c_1 + (K_1 - k_1)\frac{p_1}{p_2}c_2\right) - \frac{\rho_1 c_1}{r_1}\left(\frac{\rho_2}{r_2}K_1 + \left(1 - \frac{\rho_2}{r_2}\right)k_1\right),$$

4. if $c_1 < \frac{c_2 p_1}{p_2} < \frac{r_2 p_1}{p_2} \leq r_1$, then

$$f_1(r_1, r_2) = \rho_1 k_1 + K_1 c_1 + \frac{\rho_1}{r_1}\left((K_1 - k_1)\frac{p_1}{p_2}\left(\rho_2 + \left(1 - \frac{\rho_2}{r_2}\right)c_2\right) - K_1 c_1\right),$$

5. if $\frac{c_2 p_1}{p_2} \leq c_1 < \frac{r_2 p_1}{p_2} \leq r_1$, then

$$f_1(r_1, r_2) = \rho_1 k_1 + (K_1 - k_1)\frac{p_1}{p_2}\left(\frac{\rho_1 \rho_2}{r_1} + \left(1 - \frac{\rho_2}{r_2}\right)c_2\right) + \left(1 - \frac{\rho_1}{r_1}\right)\left(\frac{\rho_2}{r_2}K_1 + \left(1 - \frac{\rho_2}{r_2}\right)k_1\right)c_1,$$

6. if $\frac{c_2 p_1}{p_2} < \frac{r_2 p_1}{p_2} \leq c_1 < r_1$, then

$$f_1(r_1, r_2) = \left(\rho_1 + \left(1 - \frac{\rho_1}{r_1}\right)c_1\right)k_1 + (K_1 - k_1)\frac{p_1}{p_2}\left(\rho_2 + \left(1 - \frac{\rho_2}{r_2}\right)c_2\right).$$

Again, the centralized optimization problem (4) is thus a two-dimensional piecewise-smooth (non-linear) maximization program, but now there are six intervals to search in for a solution. A brute-force method for finding the globally optimal rates is by solving the two-dimensional constrained non-linear optimization problem for each of the six possible ranges and then picking the solution with the highest objective value.

4.3. Multiple stations: $n > 2$

Suppose that there are $n > 2$ parallel queues with independent subordinator inputs $J_i(t)$ for $i = 1, \dots, n$. For $n = 2$ we saw that f_i is a piecewise non-linear function with six different segments and the optimal rates can be in any of these segments. For a system of n queues the number of segments is as the number of order permutations of $(c_1, \dots, c_n, r_1, \dots, r_n)$ such that $r_i > c_i$ for all $i = 1, \dots, n$. The number of such permutations is $\frac{(2n)!}{2^n}$. Solving the optimization problem therefore becomes unfeasible from a computational perspective and approximations or a heuristic approach are required, as done in [17] for a scheduling problem with a similar structure.

A special case that may be tractable is a system of homogeneous queues that all have to mix in order to produce the higher valued product. Specifically, consider the completely symmetric case with n queues such that $p_i = \frac{1}{n}$ and $(\rho, c, a, b, K, k)_i = (\rho, c, a, b, K, k)$ for $i = 1, \ldots, n$. Let $G_i(r, c)$ denote the total revenue rate when i stations are busy, then $G_i(r, c) = nKc + ik(r - c)$ for $1 \le i < n$ and $G_n(r, c) = nKr$. If the optimal solution is symmetric, all use rate $r > \rho$, a claim that also needs to be verified, then the mean revenue rate $f(r) := \frac{1}{n}\sum_{i=1}^{n} f_i(r, \ldots, r)$ equals

$$f(r) = \sum_{i=1}^{n} \binom{n}{i} \left(\frac{\rho_i}{r_i}\right)^i \left(1 - \frac{\rho_i}{r_i}\right)^{n-i} G_i(r, c)$$
$$= Kc + (K - k)(r - c)(\frac{\rho}{r})^n + k(r - c)\frac{\rho}{r}.$$

Hence the profit is given by (the underscore c indicates the dependence on c):

$$H_c(r) := Kc + (K - k)(r - c)(\frac{\rho}{r})^n + k(r - c)\frac{\rho}{r} - ar - \frac{b}{r - \rho},$$

and its derivative w.r.t. r by

$$D_c(r) := (K - k)\left[(\frac{\rho}{r})^n - n(r - c)\frac{\rho^n}{r^{n+1}}\right] + kc\frac{\rho}{r^2} - a + \frac{b}{(r - \rho)^2}.$$

There are examples where $D_c(r)$ has multiple roots that correspond to multiple local maxima or minima. An exhaustive method for finding the optimal solution involves computing all roots, potentially $n + 1$ of them, and their respective objective values and then choosing the best. This is computationally tractable but it is hard to make any qualitative claims. Computing the decentralized symmetric solutions can be achieved in similar fashion. However, additional individual optimality constraints, similar to (14) and (16), have to be applied to ensure that the solution is indeed an equilibrium. Furthermore, the equilibrium need not be unique as we saw for $n = 2$ in Proposition 2.

4.4. Other extensions

A challenging open problem is what can be said about the joint distribution of the associated storage process. For the special case of deterministic linear input to one station and a pure-jump subordinator to the second, the joint distribution of the storage processes with the mixed output storage is obtainable (Section 4.1) by establishing an equivalence to a fluid queue studied in [12]. However, when also considering the storage processes for unmixed outputs it is not clear if the joint distribution of the storage processes can be obtained, even for the special case.

From an economic point of view an additional interesting issue to consider in future work is that the price of the higher valued product may be endogenous. This occurs for example if the total value per unit is some $K > 0$ and the stations have to decide how to split the value between them. A joint decision on the rate and how to split the value may be formulated as a Nash bargaining problem or as cooperative game, for example as done in [1] and [21] for service rate pooling in a queueing network.

Acknowledgments

Offer Kella was supported in part by grant 1647/17 from the Israel Science Foundation and the Vigevani Chair in Statistics, Israel; Onno Boxma and Liron Ravner were supported by the NWO, Netherlands Gravitation Project NETWORKS, Grant Number 024.002.003.

References

[1] H. Anily, M. Haviv, Cooperation in service systems, Oper. Res. 58 (2010) 660–673.
[2] J.P.C. Blanc, Application of the theory of boundary value problems in the analysis of a queueing model with paired services (Ph.D. thesis), University of Utrecht, 1982.
[3] J.P.C. Blanc, Asymptotic analysis of a queueing system with a two-dimensional state space, J. Appl. Probab 21 (1984) 870–886.
[4] S. Borst, O. Boxma, M. van Uitert, The asymptotic workload behavior of two coupled queues, Queueing Syst. 43 (2003) 81–102.
[5] J.W. Cohen, Analysis of Random Walks, IOS Press, 1992.
[6] J.W. Cohen, O.J. Boxma, Boundary Value Problems in Queueing System Analysis, North-Holland, 1983.
[7] K. Dębicki, M. Mandjes, Queues and Lévy Fluctuation Theory, Springer, 2015.
[8] E. Evdokimova, K. De Turck, D. Fiems, Coupled queues with customer impatience, Perform. Eval. 118 (2018) 33–47.
[9] G. Fayolle, R. Iasnogorodski, Two coupled processors: the reduction to a Riemann-Hilbert problem, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 47 (1979) 325–351.
[10] G. Fayolle, R. Iasnogorodski, V. Malyshev, Random Walks in the Quarter Plane, second ed., Springer, 2017.
[11] R. Hassin, Rational Queueing, CRC Press, 2016.
[12] O. Kella, Parallel and tandem fluid networks with dependent Lévy inputs, Ann. Appl. Probab. 3 (1993) 682–695.
[13] O. Kella, An exhaustive Lévy storage process with intermittent output, Stoch. Models 14 (1998) 979–992.
[14] O. Kella, The class of distributions associated with the generalized Pollaczek-Khinchine formula, J. Appl. Probab. 49 (2012) 883–887.
[15] O. Kella, W. Whitt, Useful martingales for stochastic storage processes with Lévy input, J. Appl. Probab. 29 (1992) 396–403.
[16] L. Kleinrock, Queueing Systems, Vol. II: Computer Applications, Wiley, 1976.
[17] L. Ravner, Y. Nazarathy, Scheduling for a processor sharing system with linear slowdown, Math. Methods Oper. Res. 86 (2017) 71–102.
[18] P. Som, W.E. Wilhelm, R.L. Disney, Kitting process in a stochastic assembly system, Queueing Syst. 17 (1994) 471–490.
[19] M. Takahashi, Hideo Ōsawa, Takehisa Fujisawa, On a synchronization queue with two finite buffers, Queueing Syst. 36 (2000) 107–123.
[20] R. Weber, S. Stidham, Optimal control of service rates in networks of queues, Adv. Appl. Probab. 19 (1987) 202–218.
[21] Y. Yu, S. Benjaafar, Y. Gerchak, Capacity sharing and cost allocation among independent firms with congestion, Prod. Oper. Manage. 24 (2015) 1285–1310.