Queueing Systems 37, 99–139, 2001
© 2001 Kluwer Academic Publishers. Manufactured in The Netherlands.

Joint Distributions for Interacting Fluid Queues

DIRK P. KROESE kroese@maths.ug.edu.au
Department of Mathematica, University of Queensland, Brisbane 4072, Australia

WERNER R.W. SCHEINHARDT werner@win.tue.nl
Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract. Motivated by recent traffic control models in ATM systems, we analyse three closely related systems of fluid queues, each consisting of two consecutive reservoirs, in which the first reservoir is fed by a two-state (on and off) Markov source. The first system is an ordinary two-node fluid tandem queue. Hence the output of the first reservoir forms the input to the second one. The second system is dual to the first one, in the sense that the second reservoir accumulates fluid when the first reservoir is *empty*, and releases fluid otherwise. In these models both reservoirs have infinite capacities. The third model is similar to the second one, however the second reservoir is now finite. Furthermore, a feedback mechanism is active, such that the rates at which the first reservoir fills or depletes depend on the state (empty or nonempty) of the second reservoir.

The models are analysed by means of Markov processes and regenerative processes in combination with truncation, level crossing and other techniques. The extensive calculations were facilitated by the use of computer algebra. This approach leads to closed-form solutions to the steady-state joint distribution of the content of the two reservoirs in each of the models.

Keywords: fluid queue, tandem queue, stationary distribution, joint distribution, feedback, traffic shaper

AMS subject classification: 60K25, 90B22

1. Introduction

Fluid queues have been widely accepted as convenient and sound models for various modern telecommunication and manufacturing systems. However, the analysis of *networks* of fluid queues – which is the subject of this paper – has thus far obtained little attention when compared to the vast amount of literature on networks of ordinary queueing systems. This may be explained by the difficulty in finding exact expressions. In particular, these models generally do not have product-form solutions. Proofs of this unfortunate fact may be found in [18,19] for fluid networks with deterministic linear internal flows and external nondecreasing Lévy input. Nonetheless, for some networks of this type, progress has been made in determining the steady-state behaviour (apart from structural results as they appear in [18,19] and references mentioned therein). In [21] an n-node tandem fluid queue with nondecreasing Lévy input into the first reservoir has been analyzed, while in [17] a generalization is studied where (Lévy) input into other nodes of the tandem is allowed as well. Both models were analyzed using a convenient martingale, leading to explicit expressions for (the Laplace-transform of) the stationary joint distribution of the contents of two reservoirs.

However, not much work has been done for networks with external fluid input(s). It seems that determining the steady-state behaviour is more difficult in this case. To our knowledge, explicit solutions have thus far been found only for a Markov-modulated two-buffer model with priorities which was considered in [10,27]. The latter reference contains an explicit solution for the steady-state distribution, Laplace-transformed in one variable. Finally we mention the techniques presented in [20] for fluid reservoirs in a random environment, which can be fruitfully applied to particular parts of fluid networks, typically leading to marginal distributions, see [2] and the marginal distribution results in the current paper.

In this paper we consider three closely related fluid systems, each consisting of two fluid reservoirs regulated by a two-state (on and off) continuous time Markov process, (M_t) say. In all models the first buffer is filled up (depleted) whenever (M_t) is in the on state (off state), so that the differences between the systems are mainly in the different behaviour of the second buffer.

In the first system the content of the second reservoir increases at times when the first reservoir is nonempty, while it decreases otherwise (unless also the second reservoir itself is empty). We will naturally refer to this fluid model as the *tandem model*. The second model will be referred to as the *dual model*. It may be regarded as "dual" to the tandem model, in the sense that the content of the second reservoir behaves opposite to that in the tandem model. Specifically, it increases when the first reservoir is empty, and decreases otherwise. Notice that both the tandem model and its dual fit into the context of Markov-modulated fluid models: the second fluid reservoir is driven by a Markov process, (M_t, D_t), where D_t is the content of the first reservoir at time t. In the third model, which we will call the *feedback model*, this is no longer the case. The second reservoir is regulated by the process (M_t, D_t) in the same way as in the dual system. However, an additional "feedback" mechanism, as introduced in [25], is in force such that the rates at which the first reservoir fills up or is depleted depend on whether the second reservoir is empty or not. A second difference between this model and the other two is that the second reservoir has a finite size K. Thus, whenever this reservoir is filling up (due to the first reservoir being empty, as in the dual model) it will do so at most until the level K is reached, after which it will remain at that level (until the first reservoir starts filling up again).

For all three systems, we are interested in the joint steady-state distribution of the content of the reservoirs and the state of the regulating Markov process. In each case, this joint distribution can be viewed as the stationary distribution of some multi-dimensional Markov process. For the derivation of the three distributions we use a variety of techniques from Markov process theory, renewal theory, Laplace transformation, stochastic integration and standard queueing theory. Due to the complicated nature of the generator equations of the multi-dimensional Markov processes mentioned above and the vast amount of algebra involved, we found this approach to be convenient. However, various (sub)results in the analysis can undoubtedly be obtained via other approaches as well, such as employing rate conservation principles, see [24], or applying martingale results, see remark 2.5.

Our motivation for studying the various models developed historically as follows. First, the tandem model was an obvious candidate for analysis since it is likely the most simple non-trivial fluid *system* with obvious applications. It was known that the analysis of the tandem system was closely related to the analysis of the waiting time in an M/G/1 queue. Drawing an analogy with the M/G/1 versus the G/M/1 queue, it was expected that the dual system would have a much more simple solution than the tandem model. Parallel investigations in [3] (see also [13]) supported this view. In addition, these investigations suggested that the dual model could be used to model so-called *two-level traffic shapers* in ATM models, to control the burstiness of traffic that is presented to an ATM communication network, see [3] and references mentioned there. However, under typical circumstances the dual system would be unstable. For these traffic shaping applications it is essential that the second reservoir be finite. The effort involved in finding explicit solutions for finite buffer models led to a new set of techniques, which eventually led to the feedback model. It was realized that the feedback mechanism could be incorporated into the model without complicating the analysis too much. Moreover, when the feedback mechanism is turned off, the model may be seen as a generalization of both the tandem and dual model.

Since the present paper will remain on a theoretical level, we will not elaborate on the relation between feedback models and traffic shaping. For more on this, we refer to [3], where another feedback model was introduced, that may be considered as a special case of the current one. Another valuable paper on more practical aspects is [5], where the same model as our current feedback model is considered. The (discretization) method employed there works fast and finds close approximations for various performance measures.

The rest of the paper is organized as follows. In section 2 the tandem model is analysed. First, we give some preliminary results for the behaviour of the first reservoir and we derive the stability conditions for the system. Then we present a stochastic decomposition result for the second reservoir. In particular this leads to the limiting distribution of the content of this reservoir *given* that the first reservoir is empty. We then use this information to derive the stationary joint distribution of the process (M_t, D_t, C_t) for the tandem model. The solution is found by solving a Laplace-transformed version of the stationary forward equations, and is given in the form of several densities in terms of integrals of modified Bessel functions of the first kind. We illustrate that, despite the complexity of these expressions, it is not hard to employ them for numerical computations. For the dual model we follow a similar approach in section 3, leading to the earlier mentioned simple solution. In section 4 the feedback model is analysed, using the relation between the feedback model and the tandem model and various additional arguments. Again we illustrate that numerical results can be obtained, although with more computational effort. Finally, we sketch some special cases and generalizations of the feedback model in section 5, most notably one that describes a fluid tandem queue as in the first part of the paper, but with finite reservoirs.

Notation and terminology

In the context of traffic shaping in ATM networks, the content of the first reservoir is called data, while the second reservoir contains an entity called credit. In the feedback model of section 4 we will therefore refer to the first reservoir as the *data buffer* and to the second one as the *credit buffer*. This explains our convention, used throughout the paper, to use the letters d and c for quantities referring to the first and second buffer, respectively. From now on we will use the word *buffer* rather than reservoir or (fluid) queue.

2. Tandem model

Consider a fluid system consisting of two infinitely large buffers, with contents D_t and C_t at time t respectively, and a continuous-time Markov process (M_t), which is characterized by its state space $\{0, 1\}$ and its Q-matrix,

$$Q = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}. \tag{2.1}$$

The first buffer is driven by (M_t) in the following manner. When (M_t) is in state 1, the content of the first buffer increases at constant rate d_+, otherwise it decreases at rate d_-, provided that it is not empty. The second buffer is driven by the first one, in such a way that its content increases at rate c_+ when the first buffer is not empty, and else decreases at rate c_-, provided that the second buffer is not empty. We note that c_+, c_-, d_+ and d_- are *positive* numbers.

A schematic overview of the behaviour of the interaction between the processes (M_t), (D_t) and (C_t) is given in figure 1, while a realization of the processes (D_t) and (C_t) is given in figure 2. The parameter values used here and in other figures pertaining to the tandem model are $a = 1$, $b = 2$, $d_+ = 2$, $d_- = 6$, $c_+ = 3$ and $c_- = 2.5$.

For simplicity, we assume from now on that $M_0 = 1$ and $D_0 = C_0 = 0$. Observe that the stochastic process (M_t, D_t, C_t) is a Markov process with state space $\{0, 1\} \times S$, where

$$S = \big\{(x, y) \in \mathbb{R} \mid x \geqslant 0,\ y \geqslant xc_+/d_+\big\}.$$

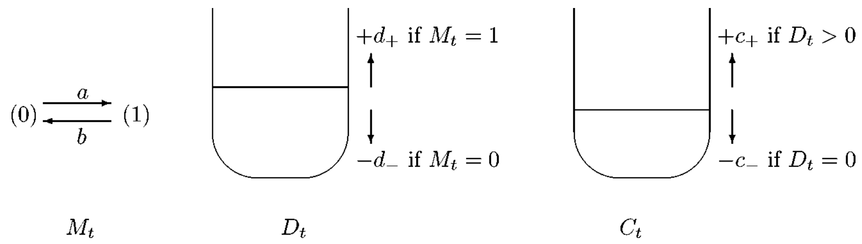

Figure 1. Interaction between the subsystems of the tandem system.

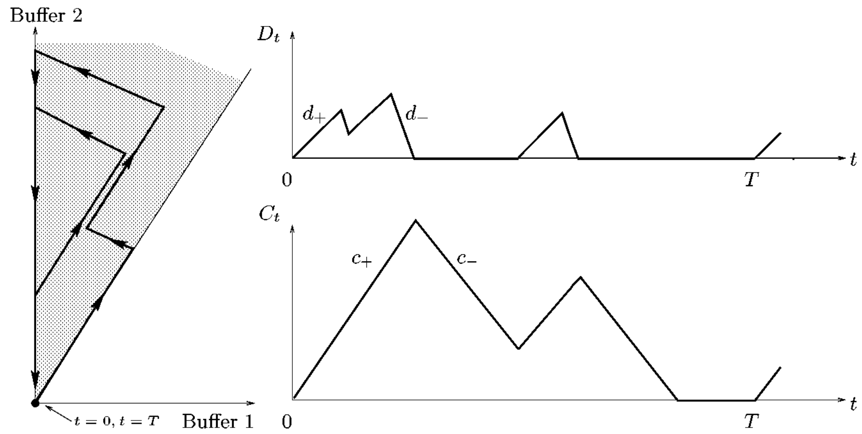

Figure 2. Realization of the buffer content processes for the tandem model.

The model may be used to describe a fluid version of the classical tandem model: two fluid buffers with constant release rates are placed in series, the first buffer is fed by an exponential on-off source while the second one is fed by the output of the first. In this case $d_- = c_+ + c_-$; notice however, that our model can handle slightly more general scenarios.

As an aside we mention that this model is related to that of [26], see also [4] and [14], where a fluid reservoir is driven by an M/M/1 queue. In fact, when we let b and d_+ grow to infinity such that their quotient remains constant and identify parameters appropriately, the second buffer here corresponds to the buffer in [26], while the content of the first buffer is the amount of work in the M/M/1 queue.

Our aim is to derive the joint stationary distribution of the Markov process (M_t, D_t, C_t). In order to do this, we first give some preliminaries, namely some known results on the stationary behaviour of the first buffer, a theorem regarding stability issues and a stochastic decomposition result for the second buffer.

Behaviour of the first buffer

It is well known (see, e.g., [6]) that when

$$bd_- - ad_+ > 0, \tag{2.2}$$

the stationary distribution of the process (D_t) exists and is given by

$$\mathbb{P}[D \leqslant x] = 1 - \rho_d\, \mathrm{e}^{-\alpha x}, \quad x \geqslant 0. \tag{2.3}$$

Here α is called the *decay rate* and is given by

$$\alpha = \frac{b}{d_+} - \frac{a}{d_-}, \tag{2.4}$$

while the *utilization* ρ_d is given by

$$\rho_d = \frac{a}{a+b} \frac{d_- + d_+}{d_-}. \tag{2.5}$$

Furthermore it is clear that the idle periods of the first buffer have an exponential distribution with parameter a. Also it is not difficult to derive, e.g., using example 3.1 in [1], that the Laplace transform L_B of the generic busy period B, say, is given by

$$L_B(s) = 1 + \frac{s + d_- \lambda_1(s)}{a}, \quad s \geqslant 0, \tag{2.6}$$

where

$$\lambda_1(s) = \frac{\eta(s) - \sqrt{\xi(s)}}{2d_- d_+}, \tag{2.7}$$

with

$$\eta(s) = bd_- - ad_+ + s(d_- - d_+),$$
$$\xi(s) = (bd_- - ad_+)^2 + 2s(d_- + d_+)(bd_- + ad_+) + s^2(d_- + d_+)^2.$$

Notice that $\lambda_1(s) \leqslant 0$ for $s \geqslant 0$. It follows that when (2.2) holds,

$$\mathbb{E}B = \frac{d_- + d_+}{bd_- - ad_+}. \tag{2.8}$$

When (2.2) does not hold, the expected length of a busy cycle is infinite.

Stability

Clearly, the process (M_t, D_t, C_t) is regenerative. As regeneration epochs we may, and henceforth will, choose the times when (M_t, D_t, C_t) is in state $(1, 0, 0)$, including time 0. Let T denote the first strictly positive regeneration epoch, see figure 2. For stability, the point at issue is under which condition the expectation of T is finite. This makes the Markov process (M_t, D_t, C_t) positive recurrent. The limiting distribution of the regenerative process (M_t, D_t, C_t) is then the same as the stationary distribution of (M_t, D_t, C_t).

The question whether a stationary distribution of the process (M_t, D_t, C_t) indeed exists can be answered using the fact that the second buffer can be viewed as a fluid queue in a "two-state random environment", as described in [20]. In such a model, the buffer content is driven by an i.i.d. sequence $\{(D_i, U_i)\}$ of down- and up-times, such that the content increases at down-times and decreases at up-times, see also [9]. In our case, the second buffer is driven by the two-state environment with down- and up-times $\{(B_i, I_i)\}$ of busy and idle periods of the first buffer. The proof of the following theorem relates the behaviour of an embedded process to that of the waiting time in a G/G/1 queue, along the lines of [20].

Theorem 2.1. The process (M_t, D_t, C_t) converges in distribution to a proper random vector (M, D, C), as $t \to \infty$, if and only if

$$\frac{bd_-}{c_+d_- + c_-d_+ + c_+d_+} - \frac{a}{c_-} > 0. \tag{2.9}$$

Proof. Let $\{(B_i, I_i)\}$ denote the sequence of busy and idle periods of the first buffer, forming the two-state random environment that drives the second buffer, as described above. Let Z_i be the content of the second buffer at the beginning of the ith busy period of the first buffer, $i = 0, 1, 2, \ldots$. Obviously, the process $\{Z_i\}$ is regenerative. Analogous to the proof of theorem 3 of [20] the expected regeneration time is finite if and only if

$$c_-\mathbb{E}I > c_+\mathbb{E}B, \tag{2.10}$$

where I and B are generic idle and busy periods of the first buffer, respectively. In view of (2.8) this is equivalent with (2.9). Notice that (2.2) is implied by (2.9).

The proof is concluded by applying Wald's lemma to show that the expected length of a regeneration epoch of the regenerative process (M_t, D_t, C_t) is finite as well. $\square$

We will henceforth assume (2.9) to be satisfied and interpret (M, D, C) as the state of the system in "steady-state". Its distribution will be denoted by $\mathbf{F} = (F_0(\mathrm{d}x, \mathrm{d}y), F_1(\mathrm{d}x, \mathrm{d}y))$, where

$$\begin{aligned} F_i(\mathrm{d}x, \mathrm{d}y) &= \mathbb{P}[M = i, D \in \mathrm{d}x,\ C \in \mathrm{d}y] \\ &= \lim_{t\to\infty} \mathbb{P}[M_t = i, D_t \in \mathrm{d}x,\ C_t \in \mathrm{d}y], \quad i \in \{0, 1\}. \end{aligned} \tag{2.11}$$

Stochastic decomposition

Next, we describe another consequence of the fact that the second buffer may be viewed as a fluid queue in a random environment, as described above.

Let Z_i, as before, be the content of the second buffer at the beginning of the ith busy period of the first buffer, $i = 0, 1, 2, \ldots$. From theorem 3 of [20] Z_i converges in distribution to a random variable Z, as $i \to \infty$. This Z is distributed as the steady-state waiting time in an M/G/1 queue with interarrival times which are distributed as c_- times the idle period of the first buffer and service times which are distributed as c_+ times the busy period of the first buffer. In particular, we have, with B and I defined as before,

$$Z \stackrel{d}{=} [Z + c_+B - c_-I]^+, \tag{2.12}$$

where Z, B and I are mutually independent, and where $[x]^+$ denotes the maximum of x and 0.

Thus, using the Pollaczek–Khintchine formula and (2.6) and (2.8), the Laplace transform of Z (L_Z say) is given by

$$L_Z(s) = \frac{bc_-d_- - a(c_+d_- + c_-d_+ + c_+d_+)}{bd_- - ad_+} \frac{s}{(c_- + c_+)s + d_-\lambda_1(c_+s)}, \tag{2.13}$$

where the function λ_1 is given in (2.7).

By theorem 4 and theorem 5 of [20] the distribution of C has the following *stochastic decomposition*,

$$C \stackrel{d}{=} \begin{cases} [Z + c_+B - c_-I^*]^+, & \text{w.p. } 1 - \rho_d, \\ Z + c_+B^*, & \text{w.p. } \rho_d, \end{cases} \tag{2.14}$$

where B, B^*, I^* and Z are mutually independent, and B^* and I^* are distributed as the residual lifetimes of B and I, respectively.

From (2.12) and the fact that $I^* \stackrel{d}{=} I$ we conclude that the conditional distribution of $(C|D = 0)$ is the same as the distribution of Z, and hence given by (2.13). We will use this information in the following section to reach our final goal. As an aside we mention that the marginal distribution of C can now be found, either by inversion of the Laplace transform L_C of C, which is clearly given by

$$\begin{aligned} L_C(s) &= (1 - \rho_d)L_Z(s) + \rho_d L_Z(s)\big[1 - L_B(c_+s)\big]/[\mathbb{E}Bc_+s] \\ &= L_Z(s)\frac{bd_- - ad_+}{a + b}\frac{-\lambda_1(c_+s)}{c_+s}, \end{aligned} \tag{2.15}$$

or by inversion of (2.13) and using

$$\Pr[C > y] = (1 - \rho_d)\frac{c_- + c_+}{c_+}\Pr[Z > y], \quad y \geqslant 0, \tag{2.16}$$

which follows from corollary 3 in [20]. The result is given in (2.34)–(2.35).

Joint stationary distribution

We are now ready to derive the joint distribution $\mathbf{F}$ of the random vector (M, D, C). The form of the distribution is easily established (see also figure 2). As a consequence of theorem 2.1, the state $(0, 0, 0)$ is a positive recurrent state of the Markov process (M_t, D_t, C_t). This state is entered via the set $\{(0, 0, y) \mid y \geqslant 0\}$ and left via the set $\{(1, x, y) \mid x \geqslant 0, y = xc_+/d_+\}$. Moreover, the set $\{0, 1\} \times \{(x, y) \mid y < xc_+/d_+\}$ is never entered. These considerations suggest that $\mathbf{F}$ be of the following form,

$$F_0\big(\{0, 0\}\big) = 1 - \rho_c, \tag{2.17}$$

$$F_0\big(\{0\}, \mathrm{d}y\big) = \sigma_0(y)\,\mathrm{d}y, \quad y > 0, \tag{2.18}$$

$$F_1(\mathrm{d}x, c_+/d_+\mathrm{d}x) = \sigma_1(x)\,\mathrm{d}x, \quad x > 0, \tag{2.19}$$

$$F_i(\mathrm{d}x, \mathrm{d}y) = f_i(x, y)\,\mathrm{d}x\,\mathrm{d}y, \quad x > 0,\ y > xc_+/d_+,\ i = 0, 1, \tag{2.20}$$

for some constant ρ_c and certain densities σ_0, σ_1, f_0 and f_1.

Since $F_0(\{0,0\}) = \mathbb{P}[C = 0]$, it follows that ρ_c is the utilization of the second buffer, and can be found from a flow balance equation,

$$(c_+ + c_-)(1-\rho_d) = c_-(1-\rho_c).$$

Using (2.5) this immediately leads to

$$\rho_c = \frac{a}{a+b}\frac{c_- + c_+}{c_-}\frac{d_- + d_+}{d_-}. \tag{2.21}$$

The following theorem gives explicit expressions for the densities.

Theorem 2.2. For the tandem model, the stationary joint distribution $\mathbf{F}$ of the process (M_t, D_t, C_t) is of the form (2.17)–(2.20), where

$$\sigma_0(y) = (1-\rho_c)\mathrm{e}^{-\beta y}\left(\frac{a}{c_-} - \frac{c_+\nu\omega}{2}\int_0^y \mathrm{e}^{-(\theta-\beta)u} H_0(0,u)\,\mathrm{d}u\right), \tag{2.22}$$

$$\sigma_1(x) = (1-\rho_c)\frac{a}{d_+}\mathrm{e}^{-bx/d_+}, \tag{2.23}$$

$$\begin{aligned} f_0(x,y) = (1-\rho_c)\frac{\nu b c_-}{d_- + d_+}\mathrm{e}^{-(b/d_+)x}\Bigg(&\frac{d_+\gamma\omega}{b}\mathrm{e}^{-\theta(y-(c_+/d_+)x)}H_1\left(x, y - \frac{c_+}{d_+}x\right) \\ &+ \frac{a}{c_-}\mathrm{e}^{-\beta(y-(c_+/d_+)x)}\left\{1 + x\omega\gamma\int_0^{y-(c_+/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_0(x,u)\,\mathrm{d}u\right\} \\ &- \frac{c_+\nu\omega}{2}\mathrm{e}^{-\beta(y-(c_+/d_+)x)}\int_0^{y-(c_+/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_1(x,u)\,\mathrm{d}u\Bigg), \end{aligned} \tag{2.24}$$

$$\begin{aligned} f_1(x,y) = (1-\rho_c)\frac{a}{d_+}\mathrm{e}^{-(b/d_+)x}\Bigg(&\omega\gamma x\mathrm{e}^{-\theta(y-(c_+/d_+)x)}H_0\left(x, y - \frac{c_+}{d_+}x\right) \\ &+ \frac{a}{c_-}\mathrm{e}^{-\beta(y-(c_+/d_+)x)}\left\{1 + x\omega\gamma\int_0^{y-(c_+/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_0(x,u)\,\mathrm{d}u\right\} \\ &- \frac{c_+\nu\omega}{2}\mathrm{e}^{-\beta(y-(c_+/d_+)x)}\int_0^{y-(c_+/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_1(x,u)\,\mathrm{d}u\Bigg). \end{aligned} \tag{2.25}$$

Here, the functions H_0 and H_1 are given by

$$H_0(x,y) = \frac{I_1\big(\sqrt{\omega(y^2+2xy\gamma)}\big)}{\sqrt{\omega(y^2+2xy\gamma)}}, \tag{2.26}$$

$$\begin{aligned} H_1(x,y) = &\frac{y^2+xy\gamma}{y^2+2xy\gamma}H_0(x,y) \\ &+ \frac{xy\gamma}{y^2+2xy\gamma}\frac{I_0\big(\sqrt{\omega(y^2+2xy\gamma)}\big) + I_2\big(\sqrt{\omega(y^2+2xy\gamma)}\big)}{2}, \end{aligned} \tag{2.27}$$

where I_i is the modified Bessel function of the first kind of order i, i.e.,

$$I_i(z) = \left(\frac{z}{2}\right)^i \sum_{k=0}^{\infty} \frac{(z/2)^{2k}}{k!(k+i)!}. \tag{2.28}$$

Furthermore, ρ_c is given in (2.21), and

$$\beta = \frac{bd_-}{c_+d_- + c_-d_+ + c_+d_+} - \frac{a}{c_-}, \tag{2.29}$$

$$\theta = \frac{bd_- + ad_+}{c_+(d_- + d_+)}, \tag{2.30}$$

$$\nu = \frac{d_- + d_+}{c_+d_- + c_-d_+ + c_+d_+}, \tag{2.31}$$

$$\omega = \frac{4abd_-d_+}{c_+^2(d_- + d_+)^2}, \tag{2.32}$$

$$\gamma = \frac{c_+(d_- + d_+)}{2d_-d_+}. \tag{2.33}$$

Clearly, it is not difficult to obtain numerical results from theorem 2.2. In figures 3 and 4 the various densities are shown for the parameter values given at the beginning of this section. Notice that β, the decay rate of the second buffer is rather small in this case, $\beta \approx 0.014$.

For completeness we mention that the distribution of the process C is given by

$$\mathbb{P}[C = 0] = 1 - \rho_c, \tag{2.34}$$

$$\mathbb{P}[C \in \mathrm{d}y] = \frac{c_- + c_+}{c_+}\sigma_0(y)\,\mathrm{d}y, \qquad y > 0. \tag{2.35}$$

Expression (2.35) can be found either by using (2.15) or (2.16) as indicated before, or more easily from theorem 2.2 by using level crossing arguments.

For the proof of theorem 2.2 it remains to prove the actual form of the various densities. We do so in three consecutive steps.

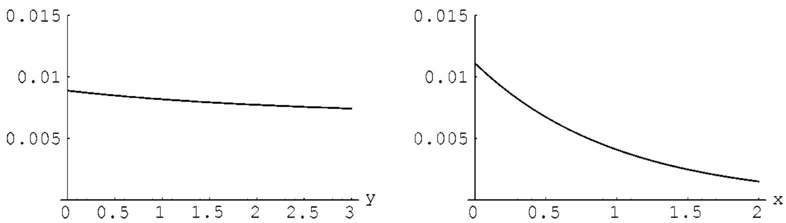

Figure 3. The densities σ_0 and σ_1 as functions of y and x, respectively.

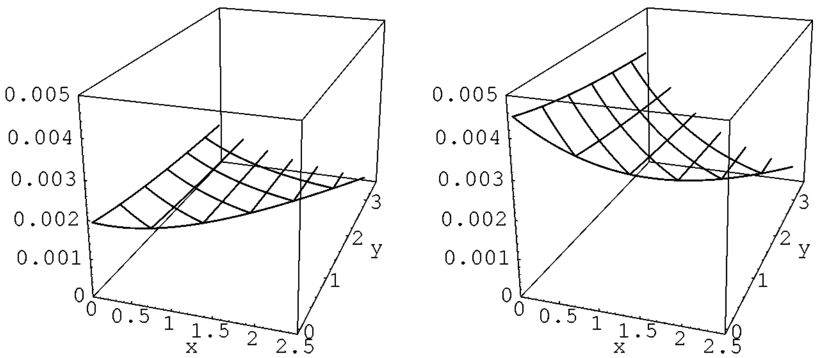

Figure 4. The densities f_0 and f_1 as functions of x and y.

Density σ_0
The conditional distribution of $(C \mid D = 0)$ has Laplace–Stieltjes transform L_Z given in (2.13). Thus,

$$\mathbb{E}\mathrm{e}^{-sC}\mathbf{1}_{\{D=0\}} = (1-\rho_d)L_Z(s). \tag{2.36}$$

Applying the shift $s \mapsto s-\theta$, with θ as in (2.30), yields after some algebra

$$\mathbb{E}\mathrm{e}^{-(s-\theta)C}\mathbf{1}_{\{D=0\}} = (1-\rho_c)\left(1+\frac{a}{c_-(s-(\theta-\beta))} - \frac{s-\sqrt{s^2-\omega}}{s-(\theta-\beta)}\frac{c_+\nu}{2}\right),$$

with β, ω and ν given in (2.29), (2.32) and (2.31), respectively. We now find σ_0 by inverting the above expression, using the fact that the inverse Laplace transform of the function $s \mapsto s-\sqrt{s^2-\omega}$ is the function

$$y \mapsto \omega\frac{I_1(y\sqrt{\omega})}{y\sqrt{\omega}}, \tag{2.37}$$

see for example [15, (28) on p. 235].

Density σ_1
The expected sojourn time of the process (D_t, C_t) in the set $\{(\widehat{x},\widehat{y}) \mid c_+\widehat{x} = d_+\widehat{y},\ \widehat{x} \leqslant x\}$ during the first regeneration period $[0, T]$ can be found by conditioning on the time the process stays on the line $\{(x, y) \mid x \geqslant 0,\ y = xc_+/d_+\}$ after $t = 0$. Since this time is exponentially distributed with parameter b, it follows after some calculations and applying the theory of regenerative processes, that

$$\mathbb{P}[c_+D = d_+C,\ D \leqslant x] = \frac{1-\mathrm{e}^{-bx/d_+}}{b\mathbb{E}T}.$$

Since we also have that

$$1-\rho_c = \mathbb{P}[C=0] = \frac{1}{\mathbb{E}T}\frac{1}{a},$$

we obtain

$$\mathbb{P}[c_+ D = d_+ C,\ D \leqslant x] = \frac{a(1-\rho_c)}{b}\big(1 - \mathrm{e}^{-bx/d_+}\big). \tag{2.38}$$

Finally, by differentiating with respect to x, we find (2.24).

Densities f_0 and f_1

This last step is the most difficult one. Our approach is to determine the densities f_0 and f_1 via a Laplace-transformed version of the stationary Kolmogorov forward equations for the Markov process (M_t, D_t, C_t). Thereto, we define the joint Laplace transforms q_i by

$$q_i(p,s) = \mathbb{E}\mathbf{1}_{\{M=i\}}\mathrm{e}^{-pD-sC}, \quad i \in \{0,1\},\ p,s \geqslant 0. \tag{2.39}$$

We will write $\mathbf{q}(p,s)$ for the column vector with entries $q_0(p,s)$ and $q_1(p,s)$.

Lemma 2.3. The vector $\mathbf{q}(p,s)$ satisfies:

$$A(p,s)\mathbf{q}(p,s) = B(p,s)\begin{pmatrix} q_0(\infty,s) \\ q_0(\infty,\infty)\end{pmatrix}, \tag{2.40}$$

where

$$A(p,s) = \begin{pmatrix} -a + d_- p - c_+ s & b \\ a & -d_+ p - c_+ s - b\end{pmatrix},$$

and

$$B(p,s) = \begin{pmatrix} d_- p - c_+ s - c_- s & c_- s \\ 0 & 0\end{pmatrix}.$$

Proof. We only prove the first row of the matrix equation, the second row can be proved in a similar manner.

Consider the stochastic processes $(X_i(t), t \geqslant 0)$, $i \in \{0,1\}$, defined by

$$X_i(t) = \mathrm{e}^{-pD_t - sC_t}\mathbf{1}_{\{M_t = i\}}.$$

Notice that both these processes are of bounded variation. We denote the continuous part of $(X_i(t))$ by $(X_i^c(t))$. In particular, we have for $t > 0$,

$$X_0(t) = X_0(0) + X_0^c(t) + \sum_{0<u\leqslant t}\big[X_0(u) - X_0(u-)\big]. \tag{2.41}$$

We now concentrate on $(X_0(t))$. The derivative of $(X_0^c(t))$ is easily found,

$$\frac{\mathrm{d}}{\mathrm{d}t}X_0^c(t) = X_0(t)\big(d_- p\mathbf{1}_{\{D_t>0\}} - c_+ s\mathbf{1}_{\{D_t>0\}} + c_- s\mathbf{1}_{\{D_t=0, C_t>0\}}\big), \quad t > 0. \tag{2.42}$$

Moreover, the pure jump part of $(X_0(t))$ can be written in stochastic integral form,

$$\sum_{0<u\leqslant t}\big[X_0(u)-X_0(u-)\big]=-\int_0^t X_0(u-)\,\mathrm{d}A_u+\int_0^t X_1(u-)\,\mathrm{d}B_u, \tag{2.43}$$

where (A_t) and (B_t) denote the counting processes that count the number of jumps of (M_t) from state 0 to 1 and from 1 to 0, respectively. The stochastic intensities at time t of (A_t) and (B_t) are given by $a\mathbf{1}_{\{M_t=0\}}$ and $b\mathbf{1}_{\{M_t=1\}}$, respectively. Because $(X_0(u-))$ is a left-continuous adapted process, we have by the theory of stochastic integration, (see, e.g., [23]) that

$$\mathbb{E}\int_0^t X_0(u-)\,\mathrm{d}A_u=\mathbb{E}\int_0^t X_0(u-)a\mathbf{1}_{\{M_u=0\}}\,\mathrm{d}u, \tag{2.44}$$

and a similar result holds for the other integral in (2.43). If we now take expectations in (2.41) and use (2.42)–(2.44), we arrive at

$$\begin{aligned}\mathbb{E}X_0(t)=\mathbb{E}X_0(0)&+d_-p\int_0^t\mathbb{E}X_0(u)\mathbf{1}_{\{D_u>0\}}\,\mathrm{d}u-c_+s\int_0^t\mathbb{E}X_0(u)\mathbf{1}_{\{D_u>0\}}\,\mathrm{d}u\\&+c_-s\int_0^t\mathbb{E}X_0(u)\mathbf{1}_{\{D_u=0,C_u>0\}}\,\mathrm{d}u-a\int_0^t\mathbb{E}X_0(u)\,\mathrm{d}u+b\int_0^t\mathbb{E}X_1(u)\,\mathrm{d}u.\end{aligned}$$

Now differentiate both sides of the previous equation with respect to t and let $t\to\infty$. By the continuity of Laplace transforms, we obtain

$$\begin{aligned}0=d_-p\big(q_0(p,s)-q_0(\infty,s)\big)&-c_+s\big(q_0(p,s)-q_0(\infty,s)\big)\\+c_-s\big(q_0(\infty,s)-q_0(\infty,\infty)\big)&-aq_0(p,s)+bq_1(p,s).\end{aligned}$$

The first row of (2.40) now follows. □

Notice that the quantities in the right-hand side of (2.40) are known. In particular, using (2.36) and (2.13), we have

$$q_0(\infty,s)=(1-\rho_d)\frac{c_-s-ac_+s\mathbb{E}B}{c_-s-a+aL_B(c_+s)}, \tag{2.45}$$

where $\mathbb{E}B$ is given in (2.8) and L_B in (2.6). Furthermore, we find $q_0(\infty,\infty)=\mathbb{P}[M=0,D=0,C=0]=\mathbb{P}[C=0]=1-\rho_c$ with ρ_c given in (2.21).

Solving $\mathbf{q}(p,s)$ from equation (2.40) yields for all $p,s\geqslant 0$,

$$q_1(p,s)=a\frac{(-d_-p+c_-s+c_+s)q_0(\infty,s)-c_-s(1-\rho_c)}{\det A(p,s)} \tag{2.46}$$

and

$$q_0(p,s)=\frac{b+d_+p+c_+s}{a}q_1(p,s),$$

which, after some algebra, reduces to

$$q_0(p,s) = \frac{b + d_+ p + c_+ s}{d_+(p + \lambda_2(c_+ s))} q_0(\infty, s) \tag{2.47}$$

and

$$q_1(p,s) = \frac{a}{d_+(p + \lambda_2(c_+ s))} q_0(\infty, s), \tag{2.48}$$

where

$$\lambda_2(s) = \frac{\eta(s) + \sqrt{\xi(s)}}{2d_- d_+}, \tag{2.49}$$

with $\eta(s)$ and $\xi(s)$ as in (2.7), and where $q_0(\infty, s)$ is given in (2.45).

It remains to be shown how $f_0(x,y)$ and $f_1(x,y)$ can be found from $q_0(p,s)$ and $q_1(p,s)$. First, inverse transformation of $q_0(p,s)$ and $q_1(p,s)$ with respect to p yields the functions

$$\begin{aligned} g_0(s) &= \left(\delta_0(x) + \frac{b - d_+\lambda_2(c_+ s) + c_+ s}{d_+} \mathrm{e}^{-\lambda_2(c_+ s)x}\right) q_0(\infty, s), \\ g_1(s) &= \frac{a}{d_+} \mathrm{e}^{-\lambda_2(c_+ s)x} q_0(\infty, s), \end{aligned}$$

where δ_0 denotes Dirac's delta function at 0. Since the distribution F_1 only has mass on S, we know that for fixed $x \geqslant 0$, g_1 must be the Laplace transform of a (generalized) function on the interval $[xc_+/d_+, \infty)$. Therefore, by multiplying $g_1(s)$ with $\exp(sxc_+/d_+)$ we obtain the Laplace transform $\tilde{h}_1(s) = \exp(sxc_+/d_+) g_1(s)$ of a function h_1 on $[0, \infty)$. After some calculations we find,

$$\begin{aligned} \tilde{h}_1(s - \theta) = (1 - \rho_c) &\frac{a}{d_+} \mathrm{e}^{-(b/d_+)x} \mathrm{e}^{x\gamma(s - \sqrt{s^2 - \omega})} \\ &\times \left(1 + \frac{a}{c_-(s - (\theta - \beta))} - \frac{c_+ \nu}{2} \frac{s - \sqrt{s^2 - \omega}}{s - (\theta - \beta)}\right). \end{aligned}$$

We can invert $\tilde{h}_1(s - \theta)$ straightforwardly (still for fixed $x \geqslant 0$) by using the following two facts. First, the function

$$y \mapsto H_0(x,y) x\omega\gamma,$$

is the inverse Laplace transform of

$$s \mapsto \exp\bigl(x\gamma\bigl(s - \sqrt{s^2 - \omega}\bigr)\bigr) - 1,$$

see, e.g., [15, p. 250, (41)]. Second, by differentiating H_0 with respect to x we see that

$$y \mapsto \omega H_1(x,y),$$

is the inverse Laplace transform of

$$s \mapsto \bigl(s - \sqrt{s^2 - \omega}\bigr) \exp\bigl(x\gamma\bigl(s - \sqrt{s^2 - \omega}\bigr)\bigr).$$

It follows that $h_1(y) = \delta_0(y)\sigma_1(x) + f_1(x, y + xc_+/d_+)$, with σ_1 and f_1 as in (2.24) and (2.25).

Similarly, for fixed $x > 0$, let $\tilde{h}_0(s) = \exp(sxc_+/d_+)g_0(s)$. We find

$$\tilde{h}_0(s-\theta) = (1-\rho_c)\nu e^{-(b/d_+)x} e^{x\gamma(s-\sqrt{s^2-\omega})}\left(\frac{ab}{(d_-+d_+)(s-(\theta-\beta))} + \frac{c_-c_+}{2d_-}\big(s-\sqrt{s^2-\omega}\big) - \frac{bc_-c_+\nu}{2(d_-+d_+)}\frac{s-\sqrt{s^2-\omega}}{s-(\theta-\beta)}\right).$$

Notice that the term $\delta_0(x)q_0(\infty, s)$ in $g_0(s)$ does not play a role, since we assume x to be strictly positive. Inversion of $\tilde{h}_0$ finally yields $h_0(y) = f_0(x, y + xc_+/d_+)$. This completes the proof of theorem 2.2.

Remark 2.4. It is interesting to note that $q_0(\infty, s)$ can be derived directly from lemma 2.3 using a "boundedness" argument. For this, write

$$\det A(p, s) = -d_-d_+\big(p + \lambda_1(c_+s)\big)\big(p + \lambda_2(c_+s)\big),$$

where $\lambda_1(s)$ and $\lambda_2(s)$ are given in (2.7) and (2.49); recall that $\lambda_1(s) \leqslant 0 \leqslant \lambda_2(s)$ for $s \geqslant 0$. Since for all $p, s \geqslant 0$, $\mathbf{q}(p, s)$ must remain *bounded*, in particular for $p = -\lambda_1(c_+s)$, the numerator in (2.46) must be zero on the set $\{(p, s)|s \geqslant 0, p = -\lambda_1(c_+s)\}$. This gives a linear equation in $q_0(\infty, s)$, from which (2.45) follows.

Remark 2.5. We mention that the main result for the tandem model can also be found by first conditioning on the on-off source being off, so that we obtain a model with two-dimensional Lévy input, after which we can proceed along the lines of [21]. Alternatively, we can find lemma 2.3 using a two-dimensional martingale as in [8].

3. Dual model

In this model we also consider a fluid system consisting of two infinitely large buffers. The first buffer is regulated by (M_t) in the same way as in the tandem model; the transition intensities of (M_t) are again given by a (from 0 to 1) and b (from 1 to 0). The only difference with the tandem model is that the content of the second buffer increases at rate c_+ when the first buffer is *empty*, and decreases at rate c_- otherwise, provided that it is not empty.

A schematic overview of the behaviour of the three subsystems is given in figure 5, while a realization of the processes (D_t) and (C_t) is given in figure 6. This time we assume that $(M_0, D_0, C_0) = (0, 0, 0)$.

As for the tandem model, the stochastic process (M_t, D_t, C_t) is a Markov process. Its state space is simply given by $\{0, 1\}\times\mathbb{R}_+\times\mathbb{R}_+$. Obviously it is a regenerative process. As regeneration epochs we choose the times (including 0) at which $(M_t, D_t, C_t) = (0, 0, 0)$.

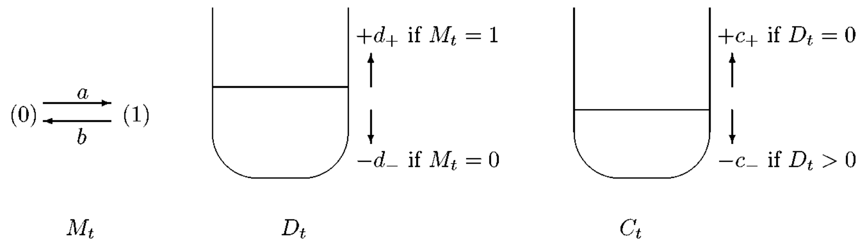

Figure 5. Interaction between the subsystems of the dual system.

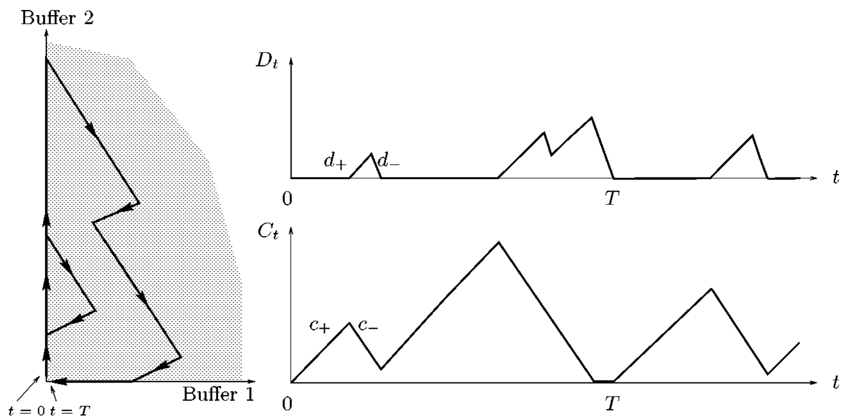

Figure 6. Realization of the buffer content processes for the dual model.

Stability

Next, in analogy to theorem 2.1, we establish the conditions under which the limiting distribution of the process (M_t, D_t, C_t) exists.

Consider the embedded process $\{Z_i\}$ describing the content of the second buffer at the beginning of the *idle* periods of the first buffer. While for the tandem case, the embedded process is related to the actual waiting time in an M/G/1-queue (with interarrival times distributed as $c_- I$ and service times distributed as $c_+ B$), we now have an embedded process that is related to the waiting time in a G/M/1-queue (with interarrival times distributed as $c_- B$ and service times distributed as $c_+ I$).

Theorem 3.1. The process (M_t, D_t, C_t) converges in distribution to a proper random vector (M, D, C), as $t \to \infty$, if and only if

$$\frac{b}{d_+} - \frac{a}{d_-} > 0, \tag{3.1}$$

and

$$\frac{a}{c_+} - \frac{bd_-}{c_- d_- + c_- d_+ + c_+ d_+} > 0. \tag{3.2}$$

Proof. The proof can be copied from the proof of theorem 2.1, apart from (2.10), which is replaced by

$$c_-\mathbb{E}B > c_+\mathbb{E}I. \tag{3.3}$$

Note also that now (3.1) is not implied by (3.2), so that we need two conditions for stability. □

We will henceforth assume conditions (3.1) and (3.2) to be satisfied. The interpretation of (M, D, C) and the definition of the limiting distribution $\mathbf{F}$ are the same as for the tandem model case, see (2.11).

Stochastic decomposition

Repeating the arguments used in the tandem model, the embedded process $\{Z_i\}$ converges in distribution to a proper random variable Z which is distributed as the waiting time in a G/M/1 queue. Specifically, by theorems IX.1.2(b) and IX.1.3 of [7], we have

$$\mathbb{P}[Z \leqslant z] = 1 - (1 - \beta c_+/a)\mathrm{e}^{-\beta z}.$$

Here β is the unique strictly positive solution of the equation $1 = \mathbb{E}\mathrm{e}^{\beta U}$, where U is distributed as $c_+I - c_-B$ and I and B are generic idle and busy periods of the first buffer respectively. It follows that β satisfies

$$1 = \frac{a}{a - \beta c_+}\frac{b}{\beta c_- + b - \lambda_1(\beta c_-)d_+},$$

which is readily solved to give

$$\beta = \frac{a}{c_+} - \frac{bd_-}{c_-d_- + c_-d_+ + c_+d_+}. \tag{3.4}$$

Moreover, the distribution of C has the following *stochastic decomposition*

$$C \stackrel{d}{=} \begin{cases} [Z + c_+I - c_-B^*]^+, & \text{w.p. } \rho_d, \\ Z + c_+I^*, & \text{w.p. } 1 - \rho_d, \end{cases} \tag{3.5}$$

similar to (2.14). Here, ρ_d is the same as for the tandem model, see (2.5). Since, I^* and I have the same distribution, we have

$$\mathbb{E}\big(\mathrm{e}^{-sC} \mid D = 0\big) = \left(\frac{\beta c_+}{a} + \left(1 - \frac{\beta c_+}{a}\right)\frac{\beta}{\beta + s}\right)\frac{a}{a + c_+s} = \frac{\beta}{\beta + s}. \tag{3.6}$$

In other words, the conditional distribution of $(C \mid D = 0)$ is exponential with intensity β. Notice that $\beta > 0$.

Joint stationary distribution

We now derive the limiting distribution $\mathbf{F}$ of the process (M_t, D_t, C_t).

Theorem 3.2. For the dual model, the stationary joint distribution $\mathbf{F}$ of the process (M_t, D_t, C_t) is of the form

$$\begin{aligned} F_0\big(\{0\}, \mathrm{d}y\big) &= \sigma_0(y)\,\mathrm{d}y, && y > 0, \\ F_i\big(\mathrm{d}x, \{0\}\big) &= \mu_i(x)\,\mathrm{d}x, && x > 0, i \in \{0, 1\}, \\ F_i(\mathrm{d}x, \mathrm{d}y) &= f_i(x, y)\,\mathrm{d}x\,\mathrm{d}y, && x, y > 0,\ i \in \{0, 1\}, \end{aligned} \tag{3.7}$$

where the densities σ_0, μ_i and f_i, $i \in \{0, 1\}$, are given by

$$\sigma_0(y) = (1 - \rho_d)\beta \mathrm{e}^{-\beta y}, \tag{3.8}$$

$$\mu_0(x) = (1 - \rho_d)\left(\frac{a}{d_-}\mathrm{e}^{-\alpha x} - \frac{bc_+}{c_-d_- + c_-d_+ + c_+d_+}\mathrm{e}^{-\zeta x}\right), \tag{3.9}$$

$$\mu_1(x) = (1 - \rho_d)\frac{a}{d_+}(\mathrm{e}^{-\alpha x} - \mathrm{e}^{-\zeta x}), \tag{3.10}$$

$$f_0(x, y) = (1 - \rho_d)\frac{bc_+\beta}{c_-d_- + c_-d_+ + c_+d_+}\mathrm{e}^{-\zeta x - \beta y}, \tag{3.11}$$

$$f_1(x, y) = (1 - \rho_d)\frac{a\beta}{d_+}\mathrm{e}^{-\zeta x - \beta y}, \tag{3.12}$$

and the constants ρ_d, α and β are given in (2.5), (2.4) and (3.4) respectively, and

$$\zeta = \alpha + \beta\frac{c_-d_- + c_+d_+}{d_-d_+} = \frac{ac_-}{c_+d_+} + \frac{bc_-}{c_-d_- + c_-d_+ + c_+d_+}. \tag{3.13}$$

Proof. The proof is similar to that of the tandem model, except that in this case much more (computer) algebra is involved. The basic structure of the proof is that we first derive a set of algebraic equations for the Laplace transform of $\mathbf{F}$, as in lemma 2.3, and then use (3.6) and a "boundedness" argument, as in remark 2.4, to solve these equations.

Let $\mathbf{q}(p, s)$ be the vector with components $q_0(p, s)$ and $q_1(p, s)$, given by

$$q_i(p, s) = \mathbb{E}\mathbf{1}_{\{M=i\}}\mathrm{e}^{-pD-sC}, \quad i \in \{0, 1\},\ p, s \geqslant 0. \tag{3.14}$$

Similar to the proof of lemma 2.3 we can show that $\mathbf{q}(p, s)$ satisfies:

$$A(p, s)\mathbf{q}(p, s) = B(p, s)\begin{pmatrix} q_0(\infty, s) \\ q_0(p, \infty) \\ q_1(p, \infty) \end{pmatrix}, \tag{3.15}$$

with

$$A(p, s) = \begin{pmatrix} -a + d_-p + c_-s & b \\ a & -b - d_+p + c_-s \end{pmatrix}, \tag{3.16}$$

and

$$B(p, s) = \begin{pmatrix} d_-p + c_+s + c_-s & c_-s & 0 \\ 0 & 0 & c_-s \end{pmatrix}. \tag{3.17}$$

(Note that $q_0(\infty, \infty) = 0$.) Consequently, for all $p, s \geqslant 0$,

$$\mathbf{q}(p,s) = \frac{H(p,s)}{\det A(p,s)} \begin{pmatrix} q_0(\infty,s) \\ q_0(p,\infty) \\ q_1(p,\infty) \end{pmatrix}, \tag{3.18}$$

where

$$H(p,s) = \begin{pmatrix} -b - d_+ p + c_- s & -b \\ -a & -a + d_- p + c_- s \end{pmatrix} B(p,s).$$

Next, we use (3.6), by which we have

$$q_0(\infty,s) = (1-\rho_d)\frac{\beta}{s+\beta}. \tag{3.19}$$

It remains to determine $q_i(p,\infty)$, $i \in \{0,1\}$, which we will do via an argument that is similar to the argument in remark 2.4. Let $s_1(p)$ and $s_2(p)$ denote the two roots of the quadratic equation $\det A(p,s) = 0$, see figure 7. We note that both roots are real and that for the smallest, s_1 say, we have $s_1(-\alpha) = s_1(0) = 0$, where α is given in (2.4). By writing out (3.18) we find that $q_0(p,s)$ is of the form

$$q_0(p,s) = \frac{c_3(p)s^3 + c_2(p)s^2 + c_1(p)s + c_0(p)}{(s - s_1(p))(s - s_2(p))(s+\beta)}, \tag{3.20}$$

where the c_i are unknown but analytic functions of p, at least for $p > -\alpha$ because $q_i(p,\infty) < \mathbb{E}\mathrm{e}^{-pD}$ and α is the decay rate of the first buffer. We now fix p such that $-\alpha < p < 0$. Because for $s > 0$ we have that $q_0(p,s) < \mathbb{E}\mathrm{e}^{-pD}$ we can conclude that $q_0(p,s)$ must be bounded for $s > 0$. Moreover, since it is not difficult to show that $s_1(p) > 0$ and $s_2(p) > 0$ (see figure 7), it follows that the numerator in (3.20) must be zero for $s = s_1(p)$ and for $s = s_2(p)$. This provides us with two linearly independent equations for $q_0(p,\infty)$ and $q_1(p,\infty)$. As an aside we note that taking $q_1(p,s)$ instead of $q_0(p,s)$ in the reasoning above leads to an equivalent set of equations. After quite a bit of algebra, the solution can be written as

$$q_0(p,\infty) = (1-\rho_d)\frac{bc_+ + ac_- + c_+ d_+ p}{c_- d_- + c_+ d_+}\,\frac{\zeta - \alpha}{(p+\alpha)(p+\zeta)}, \tag{3.21}$$

$$q_1(p,\infty) = (1-\rho_d)\frac{a}{d_+}\,\frac{\zeta - \alpha}{(p+\alpha)(p+\zeta)}, \tag{3.22}$$

with ζ given in (3.13).

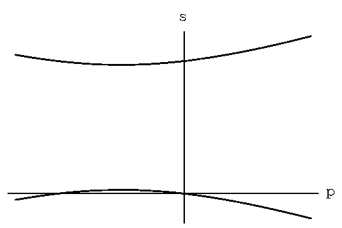

Figure 7. The roots s_1 and s_2 as functions of p.

The Laplace transforms q_0 and q_1 now follow from (3.18), (3.19), (3.21) and (3.22) and take, after some strenuous rewriting, the form

$$q_0(p,s) = (1-\rho_d)\beta \frac{(p+\zeta)(p+b/d_+) + s(ac_- + bc_+ + c_+d_+p)/(d_+d_-)}{(p+\alpha)(p+\zeta)(s+\beta)}, \quad (3.23)$$

$$q_1(p,s) = (1-\rho_d)\beta \frac{a}{d_+} \frac{p+\zeta + s(c_+d_+ + c_-d_-)/(d_+d_-)}{(p+\alpha)(p+\zeta)(s+\beta)}. \quad (3.24)$$

Equation (3.6) gives (3.8), and inverse Laplace transformation of (3.21) and (3.22) yields (3.9) and (3.10). In order to obtain the densities f_i, we first rewrite $q_i(p,s)$ to a form in which we can recognize (the transforms of) the densities we just found. The result is given by

$$\begin{aligned} q_0(p,s) = (1-\rho_d)\beta \Bigg\{ & \frac{ac_- + bc_+ + c_+d_+p}{d_+d_-(p+\alpha)(p+\zeta)} \\ & + \left(1 + \frac{bc_+}{c_-d_- + c_-d_+ + c_+d_+} \frac{1}{p+\zeta}\right) \frac{1}{s+\beta} \Bigg\}, \end{aligned} \quad (3.25)$$

$$q_1(p,s) = (1-\rho_d)\frac{a}{d_+}\left\{ \frac{\zeta - \alpha}{(p+\alpha)(p+\zeta)} + \frac{\beta}{(p+\zeta)(s+\beta)} \right\}. \quad (3.26)$$

By inversion of these expressions, we now easily find (3.11) and (3.12). □

Remark 3.3. The reason that the dual model has such a remarkably simple solution when compared to the tandem model, is that there is only one state in the regulating process (M_t, D_t) for which the content of the second buffer increases, namely (0,0). As a consequence, the solution depends on y via one exponential term, namely $e^{-\beta y}$. In [22] another solution procedure is applied to solve the dual model, illustrating this phenomenon.

4. Feedback model

Our last model is related to both the tandem and the dual model but has two essentially different characteristics: a finite (second) buffer and a feedback mechanism.

The system consists of two buffers: an infinitely large *data buffer* and a finite *credit buffer* of size K. Again, the whole system is regulated by a continuous-time Markov process (M_t), with state space $\{0, 1\}$ and transition intensities a (from 0 to 1) and b (from 1 to 0). When the credit buffer is *not empty*, the content of the data buffer increases at rate d_+ when (M_t) is in state 1 and decreases at rate d_- when (M_t) is in state 0, provided that the data buffer is not empty. However, when the credit buffer *is* empty, the up and down rates are d_+^0 and d_-^0, instead of d_+ and d_-, respectively.

Furthermore, the content of the credit buffer increases at rate c_+ when the data buffer is empty (provided that the credit buffer is not completely filled), and decreases at rate c_- otherwise (provided that the credit buffer is not empty). Notice that

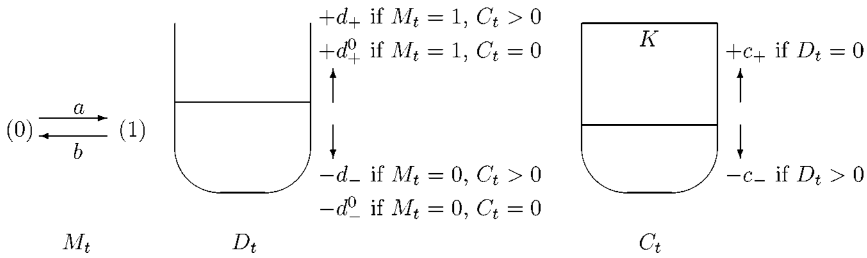

Figure 8. Interaction between the processes (M_t), (D_t) and (C_t).

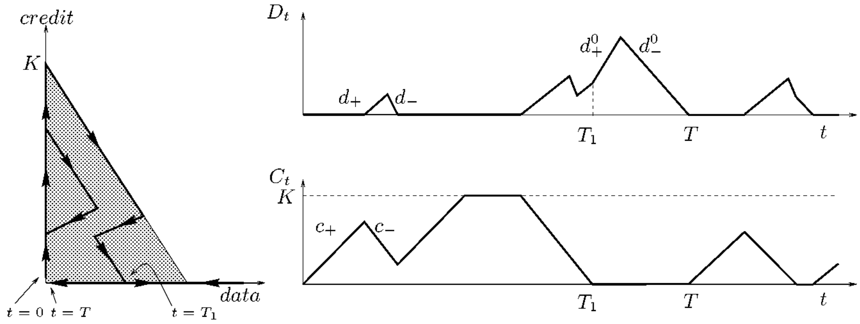

Figure 9. Realization of the buffer content processes.

$d_+, d_-, d_+^0, d_-^0, c_+$ and c_- are positive numbers, as in the other models and that the meaning of the symbols is again reflected in the notation (d for data, c for credit).

We let D_t and C_t denote the content of the data and credit buffer at time t, respectively, and observe that the stochastic process (M_t, D_t, C_t) is a Markov process, despite the presence of feedback. A schematic overview of the interaction between (M_t), (D_t) and (C_t) is given in figure 8.

As for the dual model we assume that $(M_0, D_0, C_0) = (0, 0, 0)$. A realization of the process (D_t, C_t) is given in figure 9. The parameter values used here and in other figures pertaining to this model are $a = 1$, $b = 2$, $d_+ = 2$, $d_- = 6$, $d_+^0 = 4$, $d_-^0 = 3$, $c_+ = 2.5$, $c_- = 3$ and $K = 3$.

Inspection of the behaviour of the system, see figure 9, shows that the state space of (M_t, D_t, C_t) is given by $\{0, 1\} \times S$ with

$$S = S_1 \cup S_2, \tag{4.1}$$

$$S_1 = \big\{(x, y) \mid 0 < y \leqslant K, 0 \leqslant x \leqslant (K - y)d_+/c_-\big\}, \tag{4.2}$$

$$S_2 = \{(x, y) \mid y = 0, x \geqslant 0\}. \tag{4.3}$$

Stability

It is clear that (M_t, D_t, C_t) is a regenerative process; as regeneration epochs we choose the times t when simultaneously $M_t = 0$, $D_t = 0$ and $C_t = 0$. Hence, $t = 0$ is a regeneration epoch and we denote the next one by T, i.e.,

$$T = \min\{t > 0 \mid M_t = 0, D_t = 0, C_t = 0\}. \tag{4.4}$$

We also define

$$T_1 = \min\{t > 0 \mid C_t = 0\}. \tag{4.5}$$

(See figure 9 for a visualization.)

Establishing a sufficient and necessary condition for stability of the feedback model (or the finiteness of $\mathbb{E}T$) is not much more difficult than for the tandem and dual model.

Theorem 4.1. The process (M_t, D_t, C_t) converges in distribution to a proper random vector (M, D, C), as $t \to \infty$, if and only if

$$\alpha = \frac{b}{d_+^0} - \frac{a}{d_-^0} > 0. \tag{4.6}$$

Proof. It can be shown by Wald's lemma that

$$ET_1 \leqslant \mathbb{E}N\left(\frac{1}{K} + \frac{K}{c_-}\right) < \infty,$$

where N is the number of times that the process (D_t, C_t) visits the positive y-axis during $[0, T_1]$. Furthermore $\mathbb{E}[T - T_1]$ is finite if and only if (4.6) holds. For details see [25]. $\square$

We will henceforth assume condition (4.6) to be satisfied. As in the previous models we will interpret (M, D, C) as the state of the system in stationarity. Its distribution $\mathbf{F}$ is given by $\mathbf{F}(\mathrm{d}x, \mathrm{d}y) = (F_0(\mathrm{d}x, \mathrm{d}y), F_1(\mathrm{d}x, \mathrm{d}y))$ with

$$\begin{aligned} F_i(\mathrm{d}x, \mathrm{d}y) &= \mathbb{P}\big[M = i, D \in \mathrm{d}x, C \in \mathrm{d}y\big] \\ &= \lim_{t\to\infty} \mathbb{P}\big[M_t = i, D_t \in \mathrm{d}x, C_t \in \mathrm{d}y\big], \quad i \in \{0, 1\}. \end{aligned} \tag{4.7}$$

Our primary interest is in finding this distribution.

Joint stationary distribution

In principle it should be possible to carry out the analysis of the Markov process (M_t, D_t, C_t) in a similar manner as for the tandem and dual system. That is, we derive an algebraic expression for the Laplace transforms $q_0(p, s)$ and $q_1(p, s)$ of the stationary distribution, and try to resolve any unknown function by finding an embedded process related to the waiting time in a G/G/1 queue, or by using boundedness arguments as in remark 2.4. However, due to the presence of feedback, we may no longer view the second buffer as an ordinary fluid queue in a two-state random environment. In this section, we take a completely different approach, using truncation and level crossing arguments. The (known) stationary distribution for the tandem queue will be a starting point in the

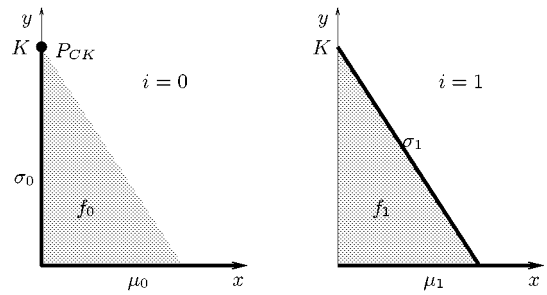

Figure 10. The stationary distribution.

analysis. However, the methodology of sections 2 and 3 will not be completely useless for the present model. In fact, in section 5 we will derive an explicit expression for the distribution of the marginal stationary distribution of the credit buffer, using this methodology.

When we let $\overset{\circ}{S}$ denote the interior of S, we expect $\mathbf{F}$ to be of the following form,

$$F_0\big(\{0\},\{K\}\big) = P_{CK}, \tag{4.8}$$

$$F_i(\mathrm{d}x,\mathrm{d}y) = f_i(x,y)\mathrm{d}x\mathrm{d}y, \quad (x,y)\in\overset{\circ}{S},\ i=0,1, \tag{4.9}$$

$$F_0\big(\{0\},\mathrm{d}y\big) = \sigma_0(y)\,\mathrm{d}y, \quad y\in[0,K], \tag{4.10}$$

$$F_1(\mathrm{d}x, K-c_-/d_+\mathrm{d}x) = \sigma_1(x)\,\mathrm{d}x, \quad x\in[0,Kd_+/c_-], \tag{4.11}$$

$$F_i\big(\mathrm{d}x,\{0\}\big) = \mu_i(x)\,\mathrm{d}x, \quad x\in[0,\infty),\ i=0,1. \tag{4.12}$$

Observe that the notation P_{CK} for the probability mass in $(0,0,K)$ is an abbreviation for $\mathbb{P}[C=K]$. In figure 10 the distribution $\mathbf{F}$ is rendered graphically.

The following theorem states that the form above is correct and gives explicit expressions for the densities.

Theorem 4.2. For the feedback model, the stationary joint distribution $\mathbf{F}$ of the process (M_t, D_t, C_t) is of the form (4.8)–(4.12), where the various densities are given as follows.

$$\sigma_0(y) = P_{CK}\mathrm{e}^{-\beta(K-y)}\left(\frac{a}{c_+} - \frac{c_-\nu\omega}{2}\int_0^{K-y}\mathrm{e}^{-(\theta-\beta)u}H_0(0,u)\,\mathrm{d}u\right), \tag{4.13}$$

$$\sigma_1(x) = P_{CK}\frac{a}{d_+}\mathrm{e}^{-(b/d_+)x}, \tag{4.14}$$

$$\begin{aligned} f_0(x,y) = P_{CK}\frac{\nu b c_+}{d_- + d_+}\mathrm{e}^{-(b/d_+)x}\Bigg(&\frac{d_+\gamma\omega}{b}\mathrm{e}^{-\theta(K-y-(c_-/d_+)x)}H_1\left(x, K-y-\frac{c_-}{d_+}x\right)\\ &+\frac{a}{c_+}\mathrm{e}^{-\beta(K-y-(c_-/d_+)x)}\left\{1 + x\omega\gamma\int_0^{K-y-(c_-/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_0(x,u)\,\mathrm{d}u\right\}\\ &-\frac{c_-\nu\omega}{2}\mathrm{e}^{-\beta(K-y-(c_-.d_+)x)}\int_0^{K-y-(c_-/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_1(x,u)\,\mathrm{d}u\Bigg), \end{aligned} \tag{4.15}$$

$$\begin{aligned}
f_1(x,y) = P_{CK}\frac{a}{d_+}\mathrm{e}^{-(b/d_+)x}\Bigg(&\omega\gamma x\mathrm{e}^{-\theta(K-y-(c_-/d_+)x)}H_0\left(x, K-y-\frac{c_-}{d_+}x\right)\\
&+\frac{a}{c_+}\mathrm{e}^{-\beta(K-y-(c_-/d_+)x)}\left\{1+x\omega\gamma\int_0^{K-y-(c_-/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_0(x,u)\,\mathrm{d}u\right\}\\
&-\frac{c_-\nu\omega}{2}\mathrm{e}^{-\beta(K-y-(c_-/d_+)x)}\int_0^{K-y-(c_-/d_+)x}\mathrm{e}^{-(\theta-\beta)u}H_1(x,u)\,\mathrm{d}u\Bigg), \qquad (4.16)
\end{aligned}$$

$$\mu_0(x) = \frac{\mathrm{e}^{-\alpha x}}{d_-^0}\left\{J_1(x\wedge Kd_+/c_-) + \eta_1(x\wedge Kd_+/c_-)J_2(x\wedge Kd_+/c_-)\right\}, \qquad (4.17)$$

$$\mu_1(x) = \frac{d_-^0}{d_+^0}\mu_0(x) - \frac{\mathbf{1}_{\{x<Kd_+/c_-\}}}{d_+^0}J_2(x). \qquad (4.18)$$

Here, the constant P_{CK} may be obtained by normalization and the functions H_0 and H_1 are given by

$$H_0(x,y) = \frac{I_1\left(\sqrt{\omega(y^2+2xy\gamma)}\right)}{\sqrt{\omega(y^2+2xy\gamma)}}, \qquad (4.19)$$

$$\begin{aligned}
H_1(x,y) = {} & \frac{y^2+xy\gamma}{y^2+2xy\gamma}\,H_0(x,y)\\
&+\frac{xy\gamma}{y^2+2xy\gamma}\,\frac{I_0\left(\sqrt{\omega(y^2+2xy\gamma)}\right)+I_2\left(\sqrt{\omega(y^2+2xy\gamma)}\right)}{2}, \qquad (4.20)
\end{aligned}$$

where I_i is the modified Bessel function of the first kind of order i as before. Furthermore, $x\wedge Kd_+/c_- \equiv \min(x, Kd_+/c_-)$,

$$\eta_0(u) = \frac{a(\mathrm{e}^{\alpha u}-1)}{d_-^0\alpha}, \qquad (4.21)$$

$$\eta_1(u) = \eta_0(u) + \mathrm{e}^{\alpha u}, \qquad (4.22)$$

$$J_1(x) = c_-\int_{u=0}^{x}\left\{\eta_0(u)f_0(u,0)+\eta_1(u)f_1(u,0)\right\}\mathrm{d}u, \qquad (4.23)$$

$$J_2(x) = c_-\int_{u=x}^{Kd_+/c_-}\left\{f_0(u,0)+f_1(u,0)\right\}\mathrm{d}u + \sigma_1(Kd_+/c_-), \qquad (4.24)$$

and finally,

$$\alpha = \frac{b}{d_+^0} - \frac{a}{d_-^0}, \qquad (4.25)$$

$$\beta = \frac{bd_-}{c_-d_-+c_-d_++c_+d_+} - \frac{a}{c_+}, \qquad (4.26)$$

$$\theta = \frac{bd_-+ad_+}{c_-(d_-+d_+)}, \qquad (4.27)$$

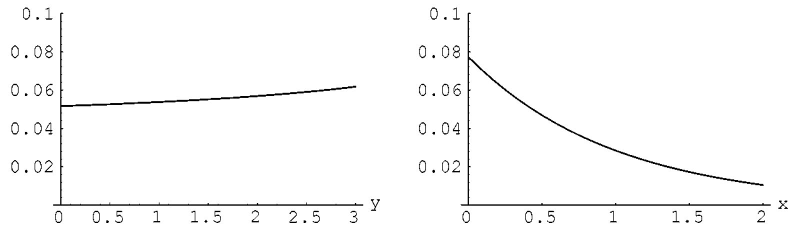

Figure 11. The densities σ_0 and σ_1 as functions of y and x, respectively.

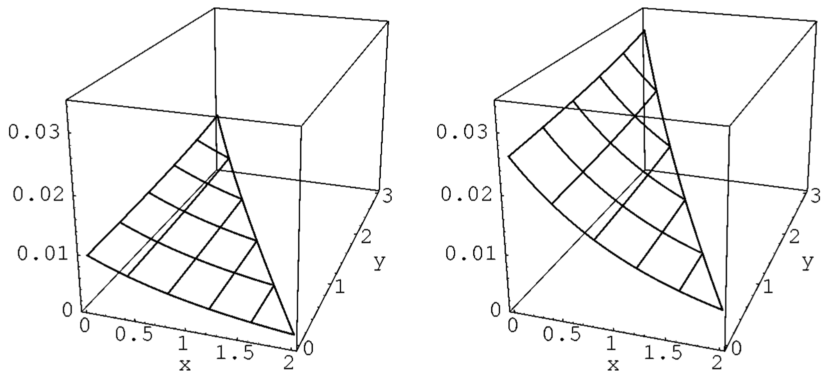

Figure 12. The densities f_0 and f_1 as functions of x and y.

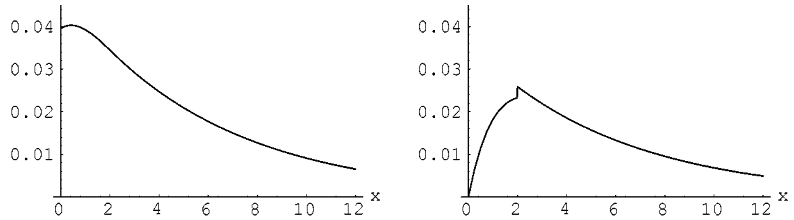

Figure 13. The densities μ_0 and μ_1 as functions of x.

$$\omega = \frac{4abd_-d_+}{c_-^2(d_- + d_+)^2}, \tag{4.28}$$

$$\nu = \frac{d_- + d_+}{c_-d_- + c_-d_+ + c_+d_+}, \tag{4.29}$$

$$\gamma = \frac{c_-(d_- + d_+)}{2d_-d_+}. \tag{4.30}$$

To illustrate that calculation of the densities in theorem 4.2 is numerically feasible, some graphs are shown in figures 11–13, where the parameter values are the same as in figure 9. The most difficult part of the numerical calculations is the normalization. For figures 11–13 we used the explicit expression for P_{CK} in (5.15).

It is interesting to note that the result in theorem 4.2 simplifies considerably when we let $K \to \infty$. In fact it takes the form of that in theorem 3.2, the only difference being the particular form of the constant coefficients of the exponential terms. Clearly, this difference vanishes when we remove the feedback by taking $d_+^0 = d_+$ and $d_-^0 = d_-$.

The proof of theorem 4.2 requires that we split the state space $\{0, 1\} \times S$ of the Markov process in two parts, namely $\{0, 1\} \times S_1$ and $\{0, 1\} \times S_2$, where S_1 and S_2 are defined in (4.2) and (4.3), see also figure 14(a). The proof is presented in three steps. In the first step we will find $\mathbf{F}$ on the set $\{0, 1\} \times S_1$ for the case $\beta > 0$ by relating it to the stationary distribution of a tandem fluid queue. In the second step, we find $\mathbf{F}$ on the set $\{0, 1\} \times S_2$. Finally, in the third step we show that the results are also valid for parameter values for which $\beta \leqslant 0$.

Densities σ_0, σ_1, f_0 and f_1

In this step we will establish a close relation between the model under consideration and the tandem model. Hereto, let $(M_t, D_t, \widehat{C}_t)$ be the stochastic process that corresponds to the tandem model with the following parameters. We identify the parameters a, b, d_+ and d_- with the parameters of the same name in the current model. Furthermore we will choose the parameters c_+ and c_- to be equal to the parameters c_- and c_+, respectively, of the current model, in other words the symbols are interchanged. In this and the following subsection we will assume that the stability condition for this tandem model holds; since this does not cover all parameter values for which the current model is stable, we will lift this restriction in the last step. The condition can be found from (2.9) by interchanging the symbols c_+ and c_- and is given by

$$\frac{bd_-}{c_-d_- + c_-d_+ + c_+d_+} - \frac{a}{c_+} > 0, \tag{4.31}$$

or, equivalently, $\beta > 0$, where β is given in (4.26). Theorem 2.1 now tells us that a stationary distribution for the process $(M_t, D_t, \widehat{C}_t)$ exists. We will denote this distribution by $\widehat{\mathbf{F}} = (\widehat{F}_0(\mathrm{d}x, \mathrm{d}y), \widehat{F}_1(\mathrm{d}x, \mathrm{d}y))$, where

$$\begin{aligned} \widehat{F}_i(\mathrm{d}x, \mathrm{d}y) &= \mathbb{P}\big[M = i, D \in \mathrm{d}x,\ \widehat{C} \in \mathrm{d}y\big] \\ &= \lim_{t\to\infty} \mathbb{P}\big[M_t = i, D_t \in \mathrm{d}x,\ \widehat{C}_t \in \mathrm{d}y\big], \quad i \in \{0, 1\}. \end{aligned} \tag{4.32}$$

Clearly, $\widehat{\mathbf{F}}$ can be found from theorem 2.2, again by interchanging c_+ and c_-.

To find the announced relation between the processes (M_t, D_t, C_t) and $(M_t, D_t, \widehat{C}_t)$, we consider yet another stochastic process $(\overline{C}_t)$, where $\overline{C}_t$ is the amount of free space in the credit buffer at time t. Hence, $\overline{C}_t = K - C_t$. In figure 14 the respective state spaces of the processes (D_t, C_t), $(D_t, \overline{C}_t)$ and $(D_t, \widehat{C}_t)$ are given.

We will now compare two processes. On the one hand we have the process $(M_t, D_t, \overline{C}_t)$, with state space $\{0, 1\} \times (\overline{S}_1 \cup \overline{S}_2)$, where $\overline{S}_i \equiv \{(x, y) \mid (x, K - y) \in S_i\}$. On the other hand we have the process $(M_t, D_t, \widehat{C}_t)$ with state space $\{0, 1\} \times \widehat{S}$ where $\widehat{S} \equiv \{(x, y) \mid y \geqslant 0, 0 \leqslant x \leqslant yd_+/c_-\}$. It is clear that $\widehat{S}$ can be written as $\widehat{S} = \overline{S}_1 \cup \widehat{S}_2$, with $\widehat{S}_2 = \{(x, y) \mid y \geqslant K, 0 \leqslant x \leqslant yd_+/c_-\}$. Moreover, the behaviour of the two

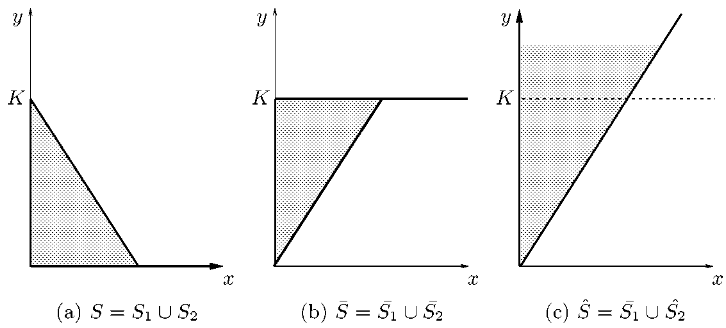

Figure 14. The sets S, $\overline{S}$ and $\widehat{S}$.

processes on $\{0, 1\} \times \overline{S}_1$ is identical, and both processes enter this set in the same way if $\alpha > 0$ (namely via state $(0, 0, K)$ with probability one). It is therefore possible to express the distribution of $(M, D, \overline{C})$ on $\{0, 1\} \times \overline{S}_1$ (and hence that of (M, D, C) on $\{0, 1\} \times S_1$) in terms of $\widehat{F}$, the stationary distribution of $(M_t, D_t, \widehat{C}_t)$. This is done in the following proposition.

Proposition 4.3. If $\alpha > 0$ and $\beta > 0$, the stationary joint distribution $\mathbf{F}$ of the process (M_t, D_t, C_t) on the set $\{0, 1\} \times S_1$ is given by

$$F_i(\mathrm{d}x, \mathrm{d}y) = k\widehat{F}_i(\mathrm{d}x, K - \mathrm{d}y), \quad (x, y) \in S_1,\ i = 0, 1. \tag{4.33}$$

The constant k is given by

$$k = \frac{\mathbb{P}[\overline{C} < K]}{\mathbb{P}[\widehat{C} < K]} = \frac{\mathbb{E}\widehat{T}}{\mathbb{E}T}, \tag{4.34}$$

where T $(\widehat{T})$ is the length of a generic regeneration period of the process $(M_t, D_t, \overline{C}_t)$ (the process $(M_t, D_t, \widehat{C}_t)$) if we choose state $(0, 0, K)$ as regeneration state.

Proof. We assume $\alpha, \beta > 0$ and consider figures 14(b) and 14(c). The choice of $(0, 0, K)$ as regeneration state for the process $(M_t, D_t, \overline{C}_t)$ entails that during any regeneration period this process first sojourns in $\{0, 1\} \times \overline{S}_1$, for a time period that is distributed as T_1 (which was defined in (4.5)), while during the remainder of such a regeneration period it stays in $\{0, 1\} \times \overline{S}_2$, with sojourn time distributed as $T - T_1$. A similar observation can be made for the process $(M_t, D_t, \widehat{C}_t)$: first it resides in $\{0, 1\} \times \overline{S}_1$, with sojourn time distributed as $\widehat{T}_1$, say, after which it remains in $\{0, 1\} \times \widehat{S}_2$, for a time period distributed as $\widehat{T} - \widehat{T}_1$. Moreover, the pathwise behaviour of both processes in the time interval $(0, T_1)$ on $\{0, 1\} \times \overline{S}_1$ is identical. Hence, we have for any $A \subset \{0, 1\} \times \overline{S}_1$,

$$\mathbb{P}\big[\big(M, D, \overline{C}\big) \in A \mid \big(D, \overline{C}\big) \in \overline{S}_1\big] = \mathbb{P}\big[\big(M, D, \widehat{C}\big) \in A \mid \big(D, \widehat{C}\big) \in \overline{S}_1\big],$$

or

$$\mathbb{P}\big[\big(M, D, \overline{C}\big) \in A\big] = \frac{\mathbb{P}[(D, \overline{C}) \in \overline{S}_1]}{\mathbb{P}[(D, \widehat{C}) \in \overline{S}_1]}\mathbb{P}\big[\big(M, D, \widehat{C}\big) \in A\big]$$

$$= \frac{\mathbb{E}T_1/\mathbb{E}T}{\mathbb{E}\widehat{T}_1/\mathbb{E}\widehat{T}} \mathbb{P}\big[\big(M, D, \widehat{C}\,\big) \in A\big] = k\mathbb{P}\big[\big(M, D, \widehat{C}\,\big) \in A\big].$$

Finally, since

$$F_i(\mathrm{d}x, \mathrm{d}y) = \mathbb{P}\big[M = i, D \in \mathrm{d}x, \overline{C} \in K - \mathrm{d}y\big], \quad i = 0, 1,$$

we easily find the stated results. □

It is now a matter of combining proposition 4.3 and theorem 2.2 (with the symbols c_+ and c_- interchanged), to find (4.8)–(4.11) and (4.13)–(4.16), when we take $P_{CK} \equiv F_0(\{0\}, \{K\}) = k\widehat{F}_0(\{0\}, \{0\})$.

Densities μ_0 and μ_1

Having found the distribution of (M_t, D_t, C_t) on $\{0, 1\} \times S_1$ (apart from normalization) in the previous subsection, we proceed to derive the densities μ_0 and μ_1 in (4.12). To do so, we first need to prove two lemmas. The first one gives us the entrance distribution G of the process (M_t, D_t, C_t) into the set $\{0, 1\} \times S_2$, that is,

$$G_i(\mathrm{d}x) = \mathbb{P}[M_{T_1} = i, D_{T_1} \in \mathrm{d}x], \quad 0 \leqslant x \leqslant Kd_+/c_-, \ i = 0, 1,$$

with T_1 as in (4.5).

Lemma 4.4. The joint distribution G of the stochastic variable (M_{T_1}, D_{T_1}) is given by

$$G_0(\mathrm{d}x) = \mathbb{E}T c_- f_0(x, 0)\,\mathrm{d}x, \tag{4.35}$$

$$G_1(\mathrm{d}x) = \mathbb{E}T\big\{c_- f_1(x, 0) + \delta_{Kd_+/c_-}(x)\sigma_1(Kd_+/c_-)\big\}\,\mathrm{d}x, \tag{4.36}$$

where δ_{Kd_+/c_-} denotes the Dirac measure at Kd_+/c_-, and σ_1, f_0 and f_1 are given in (4.14)–(4.16).

Proof. We consider the set $\{i\} \times (0, x] \times (0, \varepsilon)$. The sojourn time $V_i(x, \varepsilon)$ of (M_t, D_t, C_t) in this set during the interval $[0, T]$ is equal to $\varepsilon/c_- + \mathrm{o}(\varepsilon)$ if the event $\{M_{T_1} = i, D_{T_1} \leqslant x\}$ occurs, and is $\mathrm{o}(\varepsilon)$ otherwise. In other words, we have

$$V_i(x, \varepsilon) = \frac{\varepsilon}{c_-}\mathbf{1}_{\{M_{T_1}=i, D_{T_1}\leqslant x\}} + \mathrm{o}(\varepsilon).$$

If we take expectations, divide by $\mathbb{E}T$ and apply the Key Renewal theorem, we obtain

$$\mathbb{P}[M = i,\ D \leqslant x,\ 0 < C < \varepsilon] = \frac{\varepsilon}{c_-\mathbb{E}T}\,\mathbb{P}[M_{T_1} = i,\ D_{T_1} \leqslant x] + \mathrm{o}(\varepsilon).$$

We now find for $x < Kd_+/c_-$,

$$G_i\big((0, x]\big) = c_-\mathbb{E}T \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^\varepsilon \int_0^x f_i(u, v)\,\mathrm{d}u\,\mathrm{d}v = c_-\mathbb{E}T \int_0^x f_i(u, 0)\,\mathrm{d}u,$$

while an extra term $\mathbb{E}T\sigma_1(Kd_+/c_-)$ appears if $i = 1$ and $x = Kd_+/c_-$. The result is now immediate. □

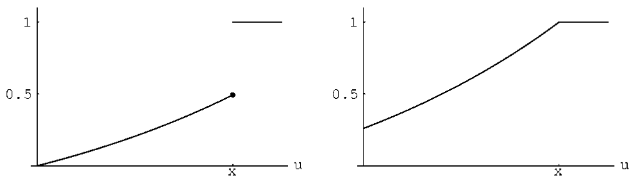

Figure 15. The probabilities $p_0(u, x)$ and $p_1(u, x)$ for fixed x.

For the second lemma, we define $N_i(x)$ as the number of times that the process (M_t, D_t, C_t) visits $(i, x, 0)$ before it reaches $(0, 0, 0)$ during the first regeneration period. Also, for $u \geqslant 0$ and $j = 0, 1$, we let

$$\mathbb{P}_{j,u}[\cdot] \equiv \mathbb{P}[\cdot \mid M_{T_1} = j, D_{T_1} = u],$$

and

$$\mathbb{E}_{j,u}[\cdot] \equiv \mathbb{E}[\cdot \mid M_{T_1} = j, D_{T_1} = u].$$

Lemma 4.5. The conditional expectations $\mathbb{E}_{j,u}N_i(x)$ are given by

$$\mathbb{E}_{j,u}N_0(x) = \begin{cases} \mathbb{E}_{j,u}N_1(x) = \mathrm{e}^{-\alpha x}\eta_j(u), & u \leqslant x, \ j = 0, 1, \\ \mathbb{E}_{j,u}N_0(x) = \mathrm{e}^{-\alpha x}\eta_1(x), & u > x, \ j = 0, 1, \\ \mathbb{E}_{j,u}N_1(x) = \mathrm{e}^{-\alpha x}\eta_0(x), & u > x, \ j = 0, 1, \end{cases} \tag{4.37}$$

where

$$\eta_0(u) = \frac{a(\mathrm{e}^{\alpha u} - 1)}{d_-^0 \alpha}, \tag{4.38}$$

$$\eta_1(u) = \frac{bd_-^0 \mathrm{e}^{\alpha u} - ad_+^0}{d_-^0 d_+^0 \alpha} = \eta_0(u) + \mathrm{e}^{\alpha u}, \tag{4.39}$$

and $\alpha = b/d_+^0 - a/d_-^0$.

Proof. First, we define

$$p_j(u, x) = \mathbb{P}_{j,u}\big[D_t = x \text{ for some } t \in (T_1, T]\big].$$

By conditioning on the first transition epoch of the process (M_t), we obtain the following relations for $u \leqslant x$,

$$p_0(u, x) = \int_0^{u/d_-^0} p_1(u - d_-^0 v, x) a\mathrm{e}^{-av}\,\mathrm{d}v,$$

$$p_1(u, x) = \int_0^{(x-u)/d_+^0} p_0(u + d_+^0 v, x) b\mathrm{e}^{-bv}\,\mathrm{d}v + \mathrm{e}^{-b(x-u)/d_+^0},$$

while for $u > x$ we have $p_j(u, x) = 1$.

Using the transformations $v \mapsto u - d_-^0 v$ and $v \mapsto u + d_+^0 v$, respectively, and differentiating with respect to u gives the following differential equation for the vector $\mathbf{p}(u, x) = (p_0(u, x), p_1(u, x))^{\mathrm{T}}$ in u,

$$\frac{\partial}{\partial u}\mathbf{p}(u, x) = \begin{pmatrix} -a/d_-^0 & a/d_-^0 \\ -b/d_+^0 & b/d_+^0 \end{pmatrix} \mathbf{p}(u, x), \quad 0 \leqslant u < x,$$

with boundary conditions $p_0(0, x) = 0$ and $p_1(x, x) = 1$. It follows that the probabilities $p_j(u, x)$ are given by

$$p_j(u, x) = 1, \qquad u > x,\ j = 0, 1, \tag{4.40}$$
$$p_j(u, x) = \eta_j(u)/\eta_1(x), \quad u \leqslant x,\ j = 0, 1, \tag{4.41}$$

see figure 15. Since the conditional distribution of $N_0(x)$ is given by

$$\mathbb{P}_{j,u}\big[N_0(x) = 0\big] = 1 - p_j(u, x),$$
$$\mathbb{P}_{j,u}\big[N_0(x) = k\big] = p_j(u, x)\big(1 - p_0(x, x)\big)\big(p_0(x, x)\big)^{k-1}, \quad k = 1, 2, \ldots,$$

we have

$$\mathbb{E}_{j,u} N_0(x) = \frac{p_j(u, x)}{1 - p_0(x, x)}.$$

Furthermore, we have

$$\mathbb{E}_{j,u} N_1(x) = \begin{cases} \mathbb{E}_{j,u} N_0(x), & \text{if } x > u, \\ \mathbb{E}_{j,u} N_0(x) - 1, & \text{if } x < u. \end{cases}$$

The desired result now follows immediately using (4.40) and (4.41). □

We are now ready to specify the densities μ_i, $i = 0, 1$.

Proposition 4.6. If $\alpha > 0$ and $\beta > 0$, the stationary joint distribution $\mathbf{F}$ of the process (M_t, D_t, C_t) on the set $\{0, 1\} \times S_2$ is given by

$$F_i\big(\mathrm{d}x, \{0\}\big) = \mu_i(x)\,\mathrm{d}x, \quad x > 0,\ i = 0, 1, \tag{4.42}$$

where μ_0 and μ_1 are given in (4.17).

Proof. First we denote the sojourn time of the process (M_t, D_t, C_t) in the set $\{i\} \times [x, x + \varepsilon] \times \{0\}$ by $W_i(x, \varepsilon)$, that is,

$$W_i(x, \varepsilon) = \int_{t=T_1}^{T} \mathbf{1}_{\{M_t = i, D_t \in [x, x+\varepsilon]\}}\,\mathrm{d}t.$$

Clearly, we have

$$\mathbb{E}_{j,u} W_0(x, \varepsilon) = \frac{\mathbb{E}_{j,u} N_0(x)}{d_-^0}\varepsilon + \mathrm{o}(\varepsilon), \tag{4.43}$$

$$\mathbb{E}_{j,u} W_1(x, \varepsilon) = \frac{\mathbb{E}_{j,u} N_1(x)}{d_+^0} \varepsilon + \mathrm{o}(\varepsilon). \tag{4.44}$$

Combining

$$\mu_i(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon \mathbb{E} T} \sum_{j=0}^{1} \int_{u=0}^{K d_+ / c_-} \mathbb{E}_{j,u} W_i(x, \varepsilon) G_j(\mathrm{d}u), \quad x > 0, \; i = 0, 1,$$

with (4.43) and (4.44) and then using lemmas 4.4 and 4.5 leads to the result. □

It is not difficult to check that propositions 4.3 and 4.6 together lead to the conclusion that the distribution $\mathbf{F}$ given in theorem 4.2 indeed is the stationary distribution of the process (M_t, D_t, C_t) when $\alpha > 0$ and $\beta > 0$.

As a side result in this subsection, we find an expression for $\mathbb{E}T$, namely $\mathbb{E}T = 1/J_2(0)$. This can be found by normalization of the distribution G in lemma 4.4.

The case $\beta \leqslant 0$

In this last step it remains to be shown that the distribution in theorem 4.2 not only represents the stationary distribution of the process (M_t, D_t, C_t) when $\alpha > 0$ and $\beta > 0$, as we showed in the previous steps, but also when $\alpha > 0$ and $\beta \leqslant 0$.

We fix the parameters b, d_+, d_-, d_+^0, d_-^0, c_+, c_- and K, and let a vary. Then we have $\alpha > 0$ if and only if $a < a_1 = b d_-^0 / d_+^0$, while $\beta > 0$ is equivalent to $a < a_0 = (b d_- c_+)/(c_- d_- + c_- d_+ + c_+ d_+)$, see figure 16. We will assume that $a_0 < a_1$, otherwise $\alpha > 0$ would imply $\beta > 0$.

In what follows we will need the infinitesimal generator $\mathcal{A}$ of the process (M_t, D_t, C_t), which is an operator mapping a function $\mathbf{h} : \mathbb{R}^2 \to \mathbb{R}^2$ to another function $\mathcal{A}\mathbf{h} : \mathbb{R}^2 \to \mathbb{R}^2$, with, for $x, y \geqslant 0$,

$$(\mathcal{A}\mathbf{h})(x, y) = \lim_{t \downarrow 0} t^{-1} \begin{pmatrix} \mathbb{E}[h_{M_t}(D_t, C_t) - h_0(x, y) \mid M_0 = 0, \; D_0 = x, \; C_0 = y] \\ \mathbb{E}[h_{M_t}(D_t, C_t) - h_1(x, y) \mid M_0 = 1, \; D_0 = x, \; C_0 = y] \end{pmatrix}.$$

It is not difficult to see that

$$(\mathcal{A}\mathbf{h})(x, y) = Q\mathbf{h}(x, y) + (\mathcal{A}_0\mathbf{h})(x, y), \quad x > 0, 0 < y < K, \tag{4.45}$$

$$(\mathcal{A}\mathbf{h})(0, y) = Q\mathbf{h}(0, y) + (\mathcal{A}_1\mathbf{h})(0, y), \quad 0 < y < K, \tag{4.46}$$

$$(\mathcal{A}\mathbf{h})(x, 0) = Q\mathbf{h}(x, 0) + (\mathcal{A}_2\mathbf{h})(x, 0), \quad x > 0, \tag{4.47}$$

$$(\mathcal{A}\mathbf{h})(0, K) = Q\mathbf{h}(0, K) + (\mathcal{A}_3\mathbf{h})(0, K), \tag{4.48}$$

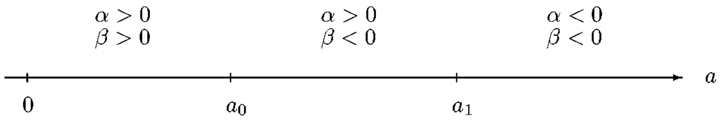

Figure 16. Behaviour of α and β as functions of a.

where Q is the generator of the process (M_t),

$$\mathcal{A}_0 = \begin{pmatrix} -d_- \frac{\partial}{\partial x} - c_- \frac{\partial}{\partial y} & 0 \\ 0 & d_+ \frac{\partial}{\partial x} - c_- \frac{\partial}{\partial y} \end{pmatrix}, \tag{4.49}$$

$$\mathcal{A}_1 = \begin{pmatrix} c_+ \frac{\partial}{\partial y} & 0 \\ 0 & d_+ \frac{\partial}{\partial x} - c_- \frac{\partial}{\partial y} \end{pmatrix}, \tag{4.50}$$

$$\mathcal{A}_2 = \begin{pmatrix} -d_-^0 \frac{\partial}{\partial x} & 0 \\ 0 & d_+^0 \frac{\partial}{\partial x} \end{pmatrix}, \tag{4.51}$$

and

$$\mathcal{A}_3 = \begin{pmatrix} c_+ \frac{\partial}{\partial y} & 0 \\ 0 & d_+ \frac{\partial}{\partial x} - c_- \frac{\partial}{\partial y} \end{pmatrix}. \tag{4.52}$$

The operator $\mathcal{A}$ can be viewed as a generalization of the Q-matrix corresponding to a continuous-time Markov process with a finite state space. In the latter context a probability measure π is stationary if and only if it satisfies $\pi Q = \mathbf{0}$, i.e., if $\pi Q\mathbf{v} = 0$ for all vectors $\mathbf{v}$. Likewise, here a measure $\mathbf{F}$ is stationary if and only if it satisfies $\mathbf{F}\mathcal{A}\mathbf{h} = 0$ for all (vector-valued) functions $\mathbf{h}$, i.e.,

$$\int_0^\infty \int_0^\infty \mathbf{F}^T(\mathrm{d}x, \mathrm{d}y)(\mathcal{A}\mathbf{h})(x, y) = 0 \tag{4.53}$$

(see, e.g., [16, p. 239]). According to theorem 4.1 a unique limiting distribution exists for any $a \in (0, a_1)$, regardless of the value of β. Moreover, we know that for $a \in (0, a_0)$ this distribution is given by the specific distribution we found in the first two steps. We will designate this distribution here by $\mathbf{F}_a$ to emphasize its dependence on the parameter a. Because the limiting distribution is stationary, we can conclude that for any suitable function $\mathbf{h}$ and any $a \in (0, a_0)$, equation (4.53) holds for $\mathbf{F} = \mathbf{F}_a$, that is,

$$\begin{aligned}
0 = {} & P_{CK} a(h_1 - h_0)(0, K) + \int_0^K \sigma_0(y)\left(a(h_1 - h_0)(0, y) + c_+ \frac{\partial h_0}{\partial y}(0, y)\right) \mathrm{d}y \\
& + \int_0^{Kd_+/c_-} \sigma_1(x)\left(-b(h_1 - h_0)(x, K - c_- x/d_+) + d_+ \frac{\partial h_1}{\partial x}(x, K - c_- x/d_+)\right. \\
& \qquad \left. - c_- \frac{\partial h_1}{\partial y}(x, K - c_- x/d_+)\right) \mathrm{d}x \\
& + \int_0^K \int_0^{(K-y)d_+/c_-} \left[f_0(x, y)\left(a(h_1 - h_0)(x, y) - d_- \frac{\partial h_0}{\partial x}(x, y) - c_- \frac{\partial h_0}{\partial y}(x, y)\right)\right. \\
& \left. + f_1(x, y)\left(-b(h_1 - h_0)(x, y) + d_+ \frac{\partial h_1}{\partial x}(x, y) - c_- \frac{\partial h_1}{\partial y}(x, y)\right)\right] \mathrm{d}x\, \mathrm{d}y \\
& + \int_0^\infty \left[\mu_0(x)\left(a(h_1 - h_0)(x, 0) - d_-^0 \frac{\partial h_0}{\partial x}(x, 0)\right)\right.
\end{aligned}$$

$$+\mu_1(x)\bigg(-b(h_1-h_0)(x,0)+d_+^0\frac{\partial h_1}{\partial x}(x,0)\bigg)\bigg]\,\mathrm{d}x. \tag{4.54}$$

To show that the above is also true for $a \in [a_0, a_1)$, we prove the following lemma, in which we will show that for certain $a \in \mathbb{C}$ the right hand side of (4.54) is a complex analytic function of a. Because it is hard to check whether the normalization constant P_{CK} is an analytic function of a, we set $P_{CK} = 1$ for a moment, thereby ignoring the probabilistic interpretation of $\mathbf{F}_a$ (and of P_{CK} itself).

Lemma 4.7. For any entire function $\mathbf{h} : \mathbb{C}^2 \to \mathbb{C}^2$, the function

$$a \mapsto \int_0^\infty \int_0^\infty \mathbf{F}_a^T(\mathrm{d}x, \mathrm{d}y)(\mathcal{A}\mathbf{h})(x, y)$$

with $P_{CK} = 1$ is complex analytic for $a \in \{z \in \mathbb{C} \mid \mathrm{Re}(z) < a_1\}$.

Proof. First we note that the singularities of the functions H_0 and H_1 in (4.19) and (4.20) can be removed by writing

$$H_0(x, y) = \frac{1}{2}\sum_{k=0}^{\infty}\frac{(z/4)^k}{k!(k+1)!}, \tag{4.55}$$

$$H_1(x, y) = H_0(x, y) + \frac{\omega x y \gamma}{4}\sum_{k=0}^{\infty}\frac{(z/4)^k}{k!(k+2)!}, \tag{4.56}$$

with

$$z = \omega\big(y^2 + 2xy\gamma\big) = \frac{4bd_-d_+}{c_-^2(d_- + d_+)^2}\big(y^2 + 2xy\gamma\big)a.$$

Since the power series in (4.55) and (4.56) are uniformly converging for all $z \in \mathbb{C}$, they are entire functions of z. Furthermore, since

$$(a, u) \mapsto \frac{4bd_-d_+}{c_-^2(d_- + d_+)^2}au^2,$$

is an entire function of a for fixed u, but also of u for fixed a ($a, u \in \mathbb{C}$), and since sums, products and concatenations of entire functions are again entire functions, we conclude that the integrand in (4.13) is also an entire function of a (for fixed u) and of u (for fixed a). But then the integral in (4.13), and hence $(a, y) \mapsto \sigma_0(y)$ is an entire function of a for fixed y and of y for fixed a, since the same holds in general for

$$(a, y) \mapsto \int_0^y g(a, u)\,\mathrm{d}u$$

when g is an entire function of a for fixed u and of u for fixed a. Similar statements can be shown to hold for σ_1, f_0, f_1, J_1, J_2, μ_0 and μ_1.

The lemma now follows readily because the partial derivatives of **h** are entire functions of x for fixed y and of y for fixed x. The restriction to $\mathrm{Re}(a) < a_1$ is due to the divergence of the last integral in (4.54) for other values of a. □

By analytic continuation we can now conclude that equation (4.54) holds, for any $a \in \mathbb{C}$ with $\mathrm{Re}(a) < a_1$, even for general P_{CK}. In particular, for a real, $a \in [a_0, a_1)$, we find $\mathbf{F}_a$ to be a stationary distribution, when we choose P_{CK} such that the total probability is 1, as before. The fact that $\mathbf{F}_a$ is the *only* stationary distribution is immediate, since we know that the process has a unique limiting distribution, regardless of the initial distribution.

This concludes the proof of theorem 4.2.

5. Special cases and generalizations

In this section we elaborate on the feedback model, and discuss some special cases and generalizations.

5.1. The normalizing constant P_{CK} and the distribution of C

Although it is in principle possible to derive the normalizing constant P_{CK} in theorem 4.2 by a laborious process of integration and summation, this is practically not a desirable option. Fortunately, it is possible, using the techniques of sections 2 and 3 to find the distribution of C, and hence also $P_{CK} = \mathbb{P}[C = K]$. We remark that functions and parameters that are not introduced in this section are the same as in theorem 4.2, e.g., H_0, β, ω, etcetera.

Recall our assumption that $M_0 = 0$, $D_0 = 0$ and $C_0 = 0$, and let $I_0, I_1, \dots$ and $B_0, B_1, \dots$ denote respectively the lengths of the *idle* periods and the *busy* periods of (D_t). Note that $\{I_i\}$ is an i.i.d. sequence with generic idle period I that is exponentially distributed with parameter a, whereas the sequence $\{B_i\}$ is not i.i.d. Let Z_k be the content of the credit buffer at the end of the kth idle period, $k = 0, 1, \dots$. Finally, let Y be distributed as a busy period of (D_t) when we forget the effect of an empty credit buffer. Specifically, the Laplace–Stieltjes transform L_Y of Y is given by the right hand side of (2.6), i.e.,

$$L_Y(s) = \frac{b}{s + b - \lambda_1(s)d_+},$$

with λ_1 as in (2.7).

The behaviour of the process $\{Z_k\}$ is given by $Z_0 = c_+ I_0$ and

$$Z_{k+1} = K - \left[K - c_+ I_{k+1} - [Z_k - c_- B_k]^+\right]^+, \quad k = 0, 1, \dots, \tag{5.1}$$

where $[x]^+$ denotes the maximum of x and 0. Direct analysis of (5.1) is problematic, because the variables B_k are not independent, and their distributions are unknown. Fortunately, the distribution of Z_k is the same as that of Z'_k when we define $Z'_0 = c_+ I_0$ and

$$Z'_{k+1} = K - \big[K - c_+ I_{k+1} - [Z'_k - c_- Y_k]^+\big]^+, \quad k = 0, 1, \ldots, \tag{5.2}$$

where $\{I_k\}$ and $\{Y_k\}$ are independent i.i.d. sequences distributed as I and Y respectively. This identifies Z'_k as the virtual waiting time immediately after arrival of a customer in a $G/M/1$-queue with uniformly bounded virtual waiting time.Specifically, the capacity of the waiting room is K, the interarrival times are $c_- Y_0, c_- Y_1, \ldots$ and the service times $c_+ I_0$, $c_+ I_1, \ldots$. The distribution of the stationary content immediately after an arrival, Z say, is given by $U(z)$ in (5.104) of [12, Part III] or in (6.10) of [11]. In our case,

$$\mathbb{P}[Z \leqslant y] = \begin{cases} 1 - \dfrac{G(K-y)}{G(K)}, & y \in [0, K), \\ 1, & y \in [K, \infty), \end{cases} \tag{5.3}$$

where the function G is the inverse Laplace transform of the function

$$L_G(s) = \frac{1}{1 - sc_+/a - L_Y(sc_-)}. \tag{5.4}$$

Laplace inversion of L_G shows that the distribution of Z is given by

$$\begin{aligned} &\mathbb{P}[Z = K] = P_{ZK}, \\ &\mathbb{P}[Z \in \mathrm{d}y] = f_Z(y)\,\mathrm{d}y, \quad y \in (0, K), \end{aligned}$$

where

$$P_{ZK} = \left(1 + \frac{a}{c_+\beta}\big(1 - \mathrm{e}^{-\beta K}\big) + \frac{c_-\nu\omega}{2\beta}\int_0^K \big(\mathrm{e}^{-\beta(K-u)} - 1\big)\mathrm{e}^{-\theta u} H_0(0,u)\,\mathrm{d}u\right)^{-1}, \tag{5.5}$$

and

$$f_Z(y) = \mathrm{e}^{-\beta(K-y)}\left(\frac{a}{c_+} - \frac{c_-\nu\omega}{2}\int_0^{K-y} \mathrm{e}^{-(\theta-\beta)u} H_0(0,u)\,\mathrm{d}u\right). \tag{5.6}$$

It follows that the Laplace transform L_Z of Z is given by

$$\begin{aligned} L_Z(s) = P_{ZK}\Bigg\{\mathrm{e}^{-sK} + \frac{1}{s-\beta}\Bigg\{&\mathrm{e}^{-\beta K}\left(\frac{a}{c_+} - \frac{c_-\nu\omega}{2}\int_0^K \mathrm{e}^{-(\theta-\beta)u} H_0(0,u)\,\mathrm{d}u\right) \\ &- \mathrm{e}^{-sK}\left(\frac{a}{c_+} - \frac{c_-\nu\omega}{2}\int_0^K \mathrm{e}^{(s-\theta)u} H_0(0,u)\,\mathrm{d}u\right)\Bigg\}\Bigg\}. \end{aligned} \tag{5.7}$$

The second step in the methodology of sections 2 and 3 was the derivation of an algebraic expression for

$$q_i(p, s) = \mathbb{E}\mathbf{1}_{\{M=i\}}\mathrm{e}^{-pD-sC}, \quad i = 0, 1.$$

For the feedback model it can be shown, using arguments analogous to the ones leading to lemma 2.3, that

$$\mathbf{q}(p,s) = \frac{H(p,s)}{\det A(p,s)} \begin{pmatrix} f(p,s) \\ q_0(p,\infty) \\ q_1(p,\infty) \end{pmatrix}, \tag{5.8}$$

where $A(p,s)$ is the same as in (3.16),

$$H(p,s) = \begin{pmatrix} -b - d_+ p + c_- s & -b \\ -a & -a + d_- p + c_- s \end{pmatrix} B(p,s),$$

with

$$B(p,s) = \begin{pmatrix} 1 & d_- p - d_-^0 p + c_- s & 0 \\ 0 & 0 & -d_+ p + d_+^0 p + c_- s \end{pmatrix},$$

and

$$f(p,s) = (d_- p + c_+ s + c_- s) q_0(\infty, s) - c_+ s \mathrm{e}^{-sK} P_{CK}.$$

We recall that for fixed $p \geqslant 0$ the zeros of $\det A(p,s)$ satisfy $s_1(p) \leqslant 0 \leqslant s_2(p)$ (see figure 7). As in remark 2.4 we can now use the fact that $\mathbf{q}(p,s)$ must remain bounded for all $p \geqslant 0$. Notice in particular that this must also be true when $s \leqslant 0$, since then $q_i(p,s) < \mathbb{E}\mathrm{e}^{-sC} < \mathrm{e}^{-sK}$. Thus, we are able to express $q_0(p,\infty)$ and $q_1(p,\infty)$ in terms of $f_1(p) = f(p, s_1(p))$ and $f_2(p) = f(p, s_2(p))$, and find

$$\begin{aligned} q_0(p,\infty) &= \frac{(f_1(p) + f_2(p))(b + d_+^0 p) g(p) + c_-(f_1(p) - f_2(p)) g_0(p)}{2p(bd_-^0 - ad_+^0 + d_-^0 d_+^0 p) g(p)}, \\ q_1(p,\infty) &= a \frac{(f_1(p) + f_2(p)) g(p) + c_-(f_1(p) - f_2(p)) g_1(p)}{2p(bd_-^0 - ad_+^0 + d_-^0 d_+^0 p) g(p)}, \end{aligned}$$

where

$$\begin{aligned} g(p) &= c_- \sqrt{(-a - b + d_- p - d_+ p)^2 - 4p(-bd_- + ad_+ - d_- d_+ p)}, \\ g_0(p) &= ab + b(d_- + d_+) p + d_+^0 p(b - a) + d_+^0 p^2 (d_- + d_+), \end{aligned}$$

and

$$g_1(p) = a + b + d_- p - 2d_-^0 p + d_+ p.$$

Evaluating (5.8) for $p = 0$ and summing q_0 and q_1 now gives,

$$\begin{aligned} \mathbb{E}\mathrm{e}^{-sC} &= \frac{c_- + c_+}{c_-} q_0(\infty, s) - \frac{c_+}{c_-} P_{CK} \mathrm{e}^{-sK} \\ &\quad + \frac{a(c_- d_- + c_- d_+ + c_+ d_+) - bc_+ d_-}{c_-(bd_-^0 - ad_+^0)} q_0(\infty, 0) \\ &\quad + \frac{c_+ P_{CK}}{c_-(bd_-^0 - ad_+^0)} \big(a\mathrm{e}^{-(a+b)K/c_-} \big(d_- - d_-^0 + d_+ - d_+^0 \big) + bd_- - ad_+ \big) \end{aligned}$$

$$-\frac{a(c_- + c_+)(d_- - d_-^0 + d_+ - d_+^0)}{c_-(bd_-^0 - ad_+^0)} q_0(\infty, (a+b)/c_-).$$

Copying the arguments following (2.14) we observe that the conditional distribution of $(C|D = 0)$, is the same as the distribution of Z. Consequently,

$$q_0(\infty, s) = \mathbb{P}[D = 0] L_Z(s), \tag{5.9}$$

and in particular

$$\mathbb{P}[C = K] = P_{ZK} \mathbb{P}[D = 0]. \tag{5.10}$$

Combining these results, gives the following proposition.

Proposition 5.1. The Laplace–Stieltjes transform of C is given by

$$L_C(s) = \mathbb{E}\mathrm{e}^{-sC} = \mathbb{P}[D = 0]\left\{\frac{c_- + c_+}{c_-} L_Z(s) - \frac{c_+}{c_-} P_{ZK} \mathrm{e}^{-sK} + \frac{\chi}{c_-}\right\}, \tag{5.11}$$

where $L_Z(s)$, the Laplace–Stieltjes transform of Z, is given in (5.7), P_{ZK} is given in (5.5) and

$$\begin{aligned}\chi = \big\{&a(c_- d_- + c_- d_+ + c_+ d_+) - bc_+ d_- + c_+(bd_- - ad_+) P_{ZK} \\ &+ ac_+\big(d_- - d_-^0 + d_+ - d_+^0\big) P_{ZK} \mathrm{e}^{-(a+b)K/c_-} \\ &- a(c_- + c_+)\big(d_- - d_-^0 + d_+ - d_+^0\big) L_Z\big((a+b)/c_-\big)\big\} / \big(bd_-^0 - ad_+^0\big).\end{aligned} \tag{5.12}$$

Inversion of (5.11) is not difficult, since we know the distribution of Z. In particular we find the following by taking $s = 0$ and $s \to \infty$ respectively in equation (5.11), and using (5.10).

Corollary 5.2. The following equalities hold,

$$\mathbb{P}[D = 0] = \frac{c_-}{c_- + c_+(1 - P_{ZK}) + \chi}, \tag{5.13}$$

$$\mathbb{P}[C = 0] = \frac{\chi}{c_- + c_+(1 - P_{ZK}) + \chi}, \tag{5.14}$$

$$P_{CK} = \frac{c_- P_{ZK}}{c_- + c_+(1 - P_{ZK}) + \chi}, \tag{5.15}$$

where P_{ZK} and χ are given in (5.5) and (5.12) respectively.

5.2. Application 1: Two-level traffic shaper

In this section we will indicate how a two-level traffic shaper may be analyzed using the general feedback model. Instead of six parameters $d_+, d_-, d_+^0, d_-^0, c_+, c_-$ for the behaviour of both buffers, we take three parameters v_0, v_1 and v_2 such that $v_0 > v_1 > v_2 > 0$ and choose

$$\begin{aligned} d_+ &= v_0 - v_1, & d_- &= v_1, \\ d_+^0 &= v_0 - v_2, & d_-^0 &= v_2, \\ c_+ &= v_2, & c_- &= v_1 - v_2. \end{aligned} \tag{5.16}$$

The interpretation is the following. The data buffer only receives data when the on-off source is in the on-state, at rate v_0. The output rate is v_1 if credit is available and $v_2 (< v_1)$ otherwise. We can think of v_2 as the long term average rate at which the data buffer is allowed to send. The rate v_1 is a higher rate that may be used for a limited period of time, namely as long as credit is available. The particular values of c_+ and c_- can be explained by arguing that whenever the data buffer is not sending (i.e., when it is empty), the "unused capacity" v_2 is saved up for later use in the form of credit, while this credit is consumed when the data buffer is sending at high rate; the "extra capacity" $v_1 - v_2$ that is used by the data buffer is taken from the credit buffer. Note that the above is equivalent to saying that the credit buffer is constantly filled at rate v_2, while it it is drained at the same rate as the data buffer (0, v_1 or v_2) at any time. For further information on two-level traffic shapers and their relation to leaky bucket traffic shapers we refer to [3] and the references mentioned there.

Simple expressions for the probabilities in corollary 5.2 are easily obtained for this case in an alternative way. Balancing the long term input and output of the credit buffer yields

$$v_2(1 - P_{CK}) = v_1 \mathbb{P}[D > 0, C > 0] + v_2 \mathbb{P}[D > 0, C = 0], \tag{5.17}$$

while a similar balance for the data buffer gives

$$\frac{a}{a+b} v_0 = v_1 \mathbb{P}[D > 0, C > 0] + v_2 \mathbb{P}[D > 0, C = 0]. \tag{5.18}$$

It follows immediately that

$$P_{CK} = 1 - \frac{a}{a+b}\frac{v_0}{v_2}. \tag{5.19}$$

Using (5.10) we now also find a simple expression for $\mathbb{P}[D = 0]$,

$$\mathbb{P}[D = 0] = P_{ZK}^{-1}\left(1 - \frac{a}{a+b}\frac{v_0}{v_2}\right), \tag{5.20}$$

while from (5.17) or (5.18) we find

$$\mathbb{P}[C = 0] = \frac{v_1}{v_1 - v_2}\left\{\left(1 - \frac{a}{a+b}\frac{v_0}{v_1}\right) - P_{ZK}^{-1}\left(1 - \frac{a}{a+b}\frac{v_0}{v_2}\right)\right\}. \tag{5.21}$$

The constant P_{ZK} in these expressions can clearly be expressed in the parameters of the model by combining (5.5) with (5.16).

The fact that P_{CK} is independent of K and v_1, may be surprising at first sight, but this can easily be understood by considering the process $(M_t, D_t - C_t + K)$. This process describes an elementary Markov-modulated fluid system in which an infinitely large fluid buffer receives fluid at rate v_0 at times when $M_t = 1$, while there is a constant output rate v_2, as long as $D_t - C_t + K > 0$. Since the credit buffer can be completely filled only at times when the data buffer is empty, we have that $\mathbb{P}[C = K] = \mathbb{P}[D - C + K = 0]$. This leads to an alternative derivation of (5.19) in which the parameters K and v_1 clearly do not play any role. Also, this viewpoint gives us a means to find the expected data buffer occupancy, since we can derive that

$$\mathbb{E}[D - C + K] = \frac{av_0}{a+b}\frac{v_0 - v_2}{bv_2 - a(v_0 - v_2)},$$

while $\mathbb{E}C$ follows from (5.11).

5.3. Application 2: Tandem queue with finite buffer(s)

A second way in which the general model may be applied is given by the following choice of parameters. Again we have three parameters for the flow rates, v_0, v_1 and v_2, such that $v_0 > v_1 > v_2 > 0$, but now we take

$$\begin{aligned} d_+ &= d_+^0 = v_0 - v_1, \qquad & d_- &= d_-^0 = v_1, \\ c_+ &= v_2, & c_- &= v_1 - v_2. \end{aligned} \tag{5.22}$$

Notice that the feedback has disappeared now, since $d_+ = d_+^0$ and $d_- = d_-^0$. Furthermore we define the process $(\overline{C}_t)$ by $\overline{C}_t \equiv K - C_t$. We can interpret $\overline{C}_t$ as the content of a buffer which receives fluid from the data buffer at rate v_1 whenever $D_t > 0$ and $\overline{C}_t < K$, while it releases fluid at rate v_2 when $\overline{C}_t > 0$. Hence the process $(M_t, D_t, \overline{C}_t)$ describes a fluid tandem queue as in section 2, but with finite second buffer.

Since the process (M_t, D_t) is not influenced by $(\overline{C}_t)$, it follows from (2.5) or directly from the balance equation for the data buffer, that

$$\mathbb{P}[D = 0] = 1 - \frac{a}{a+b}\frac{v_0}{v_1}. \tag{5.23}$$

As a consequence, we immediately find from (5.10),

$$\mathbb{P}[\overline{C} = 0] = P_{CK} = P_{ZK}\mathbb{P}[D = 0], \tag{5.24}$$

and, from the balance equation for the second buffer,

$$\mathbb{P}[\overline{C} = K] = \mathbb{P}[C = 0] = 1 - \mathbb{P}[D = 0]\left(\frac{v_1}{v_1 - v_2} - \frac{v_2}{v_1 - v_2}P_{ZK}\right), \tag{5.25}$$

where P_{ZK} can be found from (5.5) and (5.22).

In the following section we extend the (general) model to the case where the data buffer is also finite, although it must in some sense be larger than the credit buffer. This provides us with the stationary distribution for a tandem fluid queue in which both buffers are finite, provided that the fluid rates are such that during long on-periods of the fluid source, the second buffer will be completely filled *before* the first buffer is.

5.4. Extension: Finite data buffer

We shortly discuss the extension of the general feedback model in which both the credit buffer *and* the data buffer have finite sizes, K and L respectively, while the rest of the system remains unchanged, as in section 4. The process of interest is denoted as $(M_t^{(L)}, D_t^{(L)}, C_t^{(L)})$. We will only consider the case for which $L > Kd_+/c_-$, since then the analysis carries through almost identically. The main result is stated in the following theorem, where all quantities without superscript (L) are the same as in section 4.

Theorem 5.3. If the size of the data buffer is $L > Kd_+/c_-$ and $\alpha > 0$, the stationary joint distribution $\mathbf{F}^{(L)}$ of the process $(M_t^{(L)}, D_t^{(L)}, C_t^{(L)})$ satisfies

$$F_i^{(L)}(\mathrm{d}x, \mathrm{d}y) = \psi F_i(\mathrm{d}x, \mathrm{d}y), \quad 0 \leqslant x < L,\ y \geqslant 0,\ i = 0, 1, \tag{5.26}$$

$$F_1^{(L)}\big(\{L\}, \{0\}\big) = \psi \frac{d_-^0}{b} \mu_0(L), \tag{5.27}$$

with

$$\psi = \left(1 - \frac{a+b}{\alpha b} \mu_0(L)\right)^{-1}.$$

The proof is omitted for brevity; an outline can be found in [25]. Obviously, the stationary distribution can also be shown to exist when $\alpha \leqslant 0$. If we set $P_{CK} = 1$, the expressions for the various densities remain valid for some normalization constant ψ (if we replace $\eta_0(u)$ in (4.21) by au/d_-^0 for $\alpha = 0$). Since theorem 4.2 does not hold for this case, it is more difficult to find an explicit expression for this normalization constant.

References

[1] S. Aalto, Characterization of the output rate process for a Markovian storage model, J. Appl. Probab. 35(1) (1998) 184–199.

[2] S. Aalto and W.R.W. Scheinhardt, Tandem fluid queues driven by homogeneous on-off sources, Cosor memorandum, Department of Mathematics and Computing Science, Eindhoven University of Technology, The Netherlands, to appear.

[3] I.J.B.F. Adan, E.A. van Doorn, J.A.C. Resing, and W.R.W. Scheinhardt, Analysis of a single-server queue interacting with a fluid reservoir, Queueing Systems 29 (1998) 313–336.

[4] I.J.B.F. Adan and J.A.C. Resing, Simple analysis of a fluid queue driven by an M/M/1 queue, Queueing Systems 22 (1996) 171–174.

[5] I.J.B.F. Adan and J.A.C. Resing, A two-level traffic shaper for an on-off source, Cosor memorandum 99–07, Department of Mathematics and Computing Science, Eindhoven University of Technology, The Netherlands (1999).

[6] D. Anick, D. Mitra, and M.M. Sondhi, Stochastic theory of a data-handling system with multiple sources, Bell Syst. Tech. J. 61(8) (1982) 1871–1894.

[7] S. Asmussen, *Applied Probability and Queues* (Wiley, New York, 1987).

[8] S. Asmussen and O. Kella, A multi-dimensional martingale for Markov additive processes and its applications, Submitted.

[9] H. Chen and D.D. Yao, A fluid model for systems with random disruptions, Oper. Res. 40(S2) (1992) S239–S247.
[10] B.D. Choi and K.B Choi, A markov modulated fluid queueing system with strict priority, Telecom. Systems 9 (1998) 79–95.
[11] J.W. Cohen, Single server queue with uniformly bounded virtual waiting time, J. Appl. Probab. 5 (1968) 93–122.
[12] J.W. Cohen, *The Single Server Queue* (North-Holland, Amsterdam, 1982).
[13] E.A. van Doorn and W.R.W. Scheinhardt, Analysis of birth-death fluid queues, in: *Proc. KAIST Applied Mathematics Workshop*, ed. B.D. Choi (Taejon, Korea, 1996) pp. 13–29.
[14] E.A. van Doorn and W.R.W. Scheinhardt, A fluid queue driven by an infinite-state birth-death process, in: *Teletraffic Contributions for the Information Age, Proc. ITC 15*, eds. V. Ramaswami and P.E. Wirth (Amsterdam, Elsevier, 1997) pp. 465–475.
[15] A. Erdelyi, ed., *Bateman Manuscript Project, Tables of Integral Transform* (McGraw Hill, New York, 1954).
[16] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterisation and Convergence* (Wiley, New York, 1986).
[17] O. Kella, Parallel and tandem fluid networks with dependent Lévy inputs, Ann. Appl. Probab. 3(3) (1993) 682–695.
[18] O. Kella, Stability and nonproduct form of stochastic fluid networks with Lévy inputs, Ann. Appl. Probab. 6(1) (1996) 186–199.
[19] O. Kella, Non-product form of two-dimensional fluid networks with dependent Lévy inputs, Submitted.
[20] O. Kella and W. Whitt, A storage model with a two-state random environment, Oper. Res. 40(S2) (1992) S257–S262.
[21] O. Kella and W. Whitt, A tandem fluid network with Lévy input, in: *Queues and Related Models*, eds. I. Basawa and U. Bhat (Oxford University Press, Oxford, 1992) pp. 112–128.
[22] D.P. Kroese and W.R.W. Scheinhardt, A fluid queue driven by a fluid queue, in: *Teletraffic Theory as a Base for QOS: Monitoring, Evaluation, Decisions*, eds. B. Goldstein, A. Koucheryavy and M. Shneps-Shneppe (LONIIS, St. Petersburg, 1998) pp. 389–400.
[23] R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes II: Applications* (Springer, New York, 1978).
[24] M. Miyazawa, Rate conservation laws: a survey, Queueing Systems 15 (1994) 1–58.
[25] W.R.W. Scheinhardt, Markov-modulated and feedback fluid queues, Ph.D. thesis, University of Twente, Enschede, The Netherlands (1998).
[26] J.T. Virtamo and I. Norros, Fluid queue driven by an $M/M/1$ queue, Queueing Systems 16 (1994) 373–386.
[27] J. Zhang, Performance study of Markov modulated fluid flow models with priority traffic, in: *Proc. IEEE INFOCOM ’93* (1993) pp. 10–17.