Probability in the Engineering and Informational Sciences, **16**, 2002, 29–45. Printed in the U.S.A.

A TANDEM FLUID QUEUE WITH GRADUAL INPUT

Werner R. W. Scheinhardt
Faculty of Mathematical Sciences
University of Twente
Enschede, The Netherlands
E-mail: werner@math.utwente.nl
and
CWI, Amsterdam, The Netherlands

Bert Zwart
Department of Mathematics & Computing Science
Eindhoven University of Technology
Eindhoven, The Netherlands
E-mail: zwart@win.tue.nl

For a two-node tandem fluid model with gradual input, we compute the joint steady-state buffer-content distribution. Our proof exploits martingale methods developed by Kella and Whitt. For the case of finite buffers, we use an insightful sample-path argument to extend an earlier proportionality result of Zwart to the network case.

1. INTRODUCTION

In this article, we study a tandem fluid network which operates in a two-state random environment. Depending on the state of the environment, the content in the first buffer either increases according to some general stochastic process or it decreases linearly. The output of the first buffer is fed into a second buffer, after which it leaves the system. For this model, we compute the Laplace–Stieltjes transform of the joint steady-state buffer-content distribution.

The model in this article can be put in the context of tandem queues where the service at the various queues is deterministic, and the probabilistic behavior is only due to the stochastic arrival process(es). These systems may typically be used to

 0269-9648/02 $12.50

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

model a sequence of multiplexers in a communication network or a sequence of production lines that operate in a deterministic manner.

The first of these systems to be analyzed were classical tandem queues with deterministic service times; see, for example, Rubin [22], Shalmon and Kaplan [24], Boxma and Resing [8] and references therein. These may be viewed as slotted (discrete-time) versions of the model considered here. In the last decade, another class of models, operating in continuous time, was studied successfully. Here, networks of fluid queues are driven by (instantaneous) Lévy input; see, for example, Kella and Whitt [18,19], and Kella [15].

Several recent articles were concerned with a third class of models, in which fluid networks are fed by *gradual* input; this type of model is considered in the present article. Kroese and Scheinhardt [20] (see also Scheinhardt [23]) analyzed several systems of fluid queues that are driven by a two-state Markov process. Their framework included a two-node tandem system for which the joint stationary distribution of the buffer contents was found. The transform version of this result was generalized to feedforward networks with Markov-modulated input by Kella [16]. A different extension can be found in Aalto and Scheinhardt [1], where a multinode tandem fluid queue fed by homogeneous On–Off sources with general On-time distribution was analyzed.

The main results in this study are strongly related to those in [16] and [1], but there are some differences. The main difference with [1] is that we find the *joint* Laplace–Stieltjes transform of the buffer contents, whereas [1] is mainly concerned with marginal results. Compared to [16], we study a simpler network topology. On the other hand, our input process is more general than the (Markov-additive) input process of [16]. In particular, our assumptions allow one to consider non-Markovian input. For example, the semi-Markov input process as considered recently by Boxma, Kella, and Perry [6] falls within the framework considered here; see Section 4.2. Non-Markovian input processes are currently particularly relevant in communication networks, where it is now quite common to assume that On-periods of On–Off sources are heavy tailed, hence not of phase type. We refer to Boxma and Dumas [5] for a survey on fluid queues with heavy-tailed input characteristics; see also the recent book by Park and Willinger [21]. In addition to its intrinsic interest, the tandem fluid queue considered here seems to play a key role in more complicated networks of fluid queues; see, for example, Van Uitert and Borst [26], which is concerned with networks of fluid queues under the generalized processor-sharing discipline.

The way in which we derive our results is as follows. First, we show that the joint steady-state buffer-content distribution satisfies a decomposition property; this distribution can be written as the sum of two random vectors (see also [6] for a similar result for the single-buffer case). The first term can be viewed as the steady-state buffer-content distribution of a tandem network with instantaneous Lévy input at both nodes. The joint buffer-content distribution of this particular tandem network is obtained by applying the powerful martingale that was introduced by Kella and Whitt [19], which is also applied in [15,16]. The second term in the decomposition is associated with the stationary distribution of a clearing model.

We also treat the case in which the buffer sizes are finite. By means of an insightful sample-path argument, it is shown that the steady-state distributions of the finite and infinite buffer models are proportional. This extends the approach in Zwart [27], where the corresponding result for the single-node case was obtained. The intuition behind the proof is reminiscent of many articles dealing with traditional (i.e., nonfluid) finite-capacity systems, such as those in Boots and Tijms [3,4], Gouweleeuw and Tijms [11], Hooghiemstra [12], and Keilson and Servi [13,14]. Our approach can also be applied to the finite-buffer equivalents of the networks considered in [15,16,18].

The article is organized as follows. Section 2 provides a detailed model description and states a number of preliminary results. Our main results are in Section 3, where we show the decomposition property. Furthermore, we use this property to find an expression for the transform of the joint distribution. In Section 4, we apply the results of Section 3 to some examples which allow for explicit computations, namely the two respective cases where the input into the first buffer is regulated by an On–Off process and by a semi-Markov process. Section 5 treats the finite buffer case.

2. MODEL DESCRIPTION AND PRELIMINARIES

We start with a detailed model description. The content process of the first buffer falls within the framework of Kella and Whitt [17], because it operates in a two-state random environment. In particular, the first buffer is fed by a general source which operates in two modes, which we call On and Off, and it has a constant output rate c_1. During Off periods of the source, which are exponentially distributed with parameter λ, no fluid enters the buffer, so its content decreases linearly with slope c_1 as long as the buffer is not empty. When the source is On, the buffer content has the same increments (in distribution) as the generic stochastic process $\mathcal{X} = \{X(t), t \geq 0\}$, which has nondecreasing sample paths. Note that $X(t)$ is distributed as the rise of the fluid level in the first t time units of an On period, so the total amount of fluid that was added is distributed as $X(t) + c_1 t$. Obviously, during different On periods, the fluid level behaves according to different realizations of $\mathcal{X}$, all starting at $t = 0$. Furthermore, an On period is terminated after some (generic) time A, which may depend on $\mathcal{X}$ and has finite mean. For Re $u, v \geq 0$, we define

$$\gamma(u,v) = \mathbb{E}\{e^{-uX(A)-vA}\} \tag{2.1}$$

as the Laplace–Stieltjes transform (LST) of $(X(A), A)$. It is easy to see that the steady-state probability that the source is On, which we denote by p, is given by

$$p = \frac{\lambda \mathbb{E}\{A\}}{1 + \lambda \mathbb{E}\{A\}}. \tag{2.2}$$

As long as the first buffer is not empty, the processed fluid is fed into a second buffer at rate c_1. The second buffer also has a constant output rate, namely c_2, as long as it

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

is not empty. To avoid a trivial model, we will assume that $c_1 > c_2$, so that the second buffer is the bottleneck.

The content of buffer i $(i = 1,2)$ at time t is denoted by $V_i(t)$. The process of interest is then given by $\mathcal{V} = \{V(t), t \geq 0\}$, where $V(t) = (V_1(t), V_2(t))$. A typical sample path is depicted in the first part of Figure 1. It is clear that both buffer-content

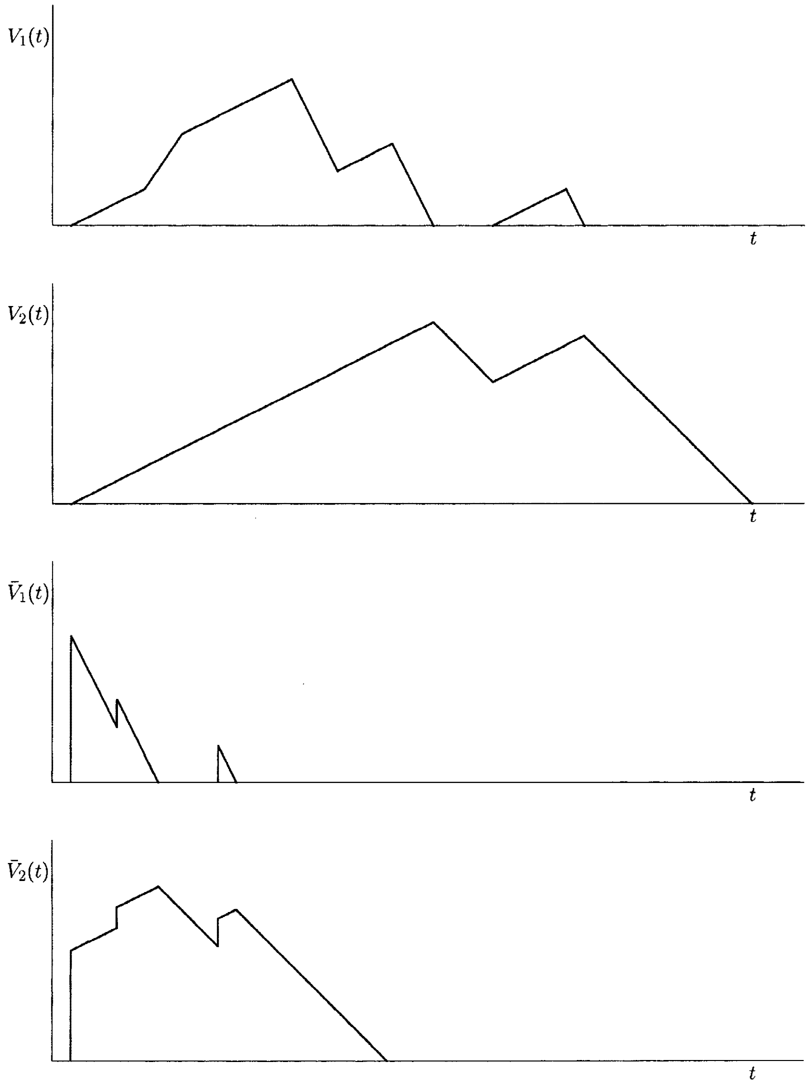

FIGURE 1. Construction of the process $\overline{\mathcal{V}}$ from $\mathcal{V}$.

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

processes have negative drift if and only if the expected amount of fluid that flows into the first buffer per unit of time is less than c_2; that is, iff

$$\rho = p\,\frac{\mathbb{E}\{X(A)\} + c_1\mathbb{E}\{A\}}{\mathbb{E}\{A\}} < c_2. \tag{2.3}$$

This may be rewritten, using (2.2), as

$$\lambda\mathbb{E}\{X(A)\} + \lambda(c_1 - c_2)\mathbb{E}\{A\} < c_2, \tag{2.4}$$

which we assume to hold throughout the remainder of the article. Clearly, the process $\mathcal{V}$ is regenerative; as regeneration epochs, we take the instants when the On–Off source starts an On period in an empty network. Using standard regenerative process theory (see, e.g., Asmussen [2] and Cohen [10]) it now follows that $V(t)$ converges in distribution to a random vector $V = (V_1, V_2)$. Choosing 0 to be a regeneration epoch and denoting a (generic) regeneration cycle by C, the distribution of V can be written as

$$\mathbb{P}\{V_1 > x_1; V_2 > x_2\} = \frac{1}{\mathbb{E}\{C\}}\,\mathbb{E}\left\{\int_0^C 1_{[V_1(t)>x_1, V_2(t)>x_2]}\,dt\right\}, \tag{2.5}$$

where $1_{[S]}$ is the indicator function of the event S. For later reference, we find the probability that the first buffer is empty:

$$\mathbb{P}\{V_1 = 0\} = 1 - \frac{\rho}{c_1} = \frac{c_1 - \lambda\mathbb{E}\{X(A)\}}{c_1 + c_1\lambda\mathbb{E}\{A\}}; \tag{2.6}$$

see [17]. Similarly, we find for the second buffer that

$$\mathbb{P}\{V_2 = 0\} = 1 - \frac{\rho}{c_2} = \frac{c_2 - \lambda\mathbb{E}\{X(A)\} - \lambda(c_1 - c_2)\mathbb{E}\{A\}}{c_2 + c_2\lambda\mathbb{E}\{A\}}. \tag{2.7}$$

Since we assumed $c_1 > c_2$, the latter probability is, in fact, equal to $\mathbb{P}\{V_1 = 0, V_2 = 0\}$, the probability that the entire system is empty.

We define the joint LST of V as $\omega(u,v) = \mathbb{E}\{e^{-uV_1 - vV_2}\}$. As mentioned in Section 1, the main goal of this article is to compute $\omega(u,v)$. In doing so, we need two additional random variables that are closely related to A and $X(A)$. We define them as follows. First, A^* is distributed as the equilibrium distribution of A; that is,

$$\mathbb{P}\{A^* \le x\} = \frac{1}{\mathbb{E}\{A\}}\int_0^x \mathbb{P}\{A > y\}\,dy. \tag{2.8}$$

We interpret A^* as the elapsed time that the source is On, when we observe the system in steady state at an arbitrary epoch during an On period. At the same time epoch, one can also observe the increase of the buffer content since the beginning of that On period, which we denote by $X(A^*)$. The distribution $X(A^*)$ is also known, see [17]; however, in the sequel, we will need the joint distribution of $(X(A^*), A^*)$ as

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

well. For completeness sake, we give expressions for this distribution and its transform, which we denote by $\gamma^*(u,v)$:

$$\mathbb{P}\{X(A^*) > x, A^* > y\} = \frac{1}{\mathbb{E}\{A\}}\,\mathbb{E}\left\{\int_0^A 1_{[X(t)>x,\, t>y]}\,dt\right\}. \tag{2.9}$$

$$\gamma^*(u,v) = \mathbb{E}\{e^{-uX(A^*)-vA^*}\} = \frac{1}{\mathbb{E}\{A\}}\,\mathbb{E}\left\{\int_0^A e^{-uX(t)-vt}\,dt\right\}. \tag{2.10}$$

In the sequel, we will assume that γ^* is known; Section 4 provides explicit expressions for γ^* in some special cases. As an aside, we note that $X(A^*)$ can be interpreted as the stationary workload of a (fluid) queue fed by $\mathcal{X}$, where all the work is removed after a random time A. Such a model is called a clearing model; see, for example, Stidham [25]. The random variable A^* can then be interpreted as the time elapsed since the last clearing.

Finally, we need an expression for the transform $\pi(s) = \mathbb{E}\{e^{-sP}\}$, where the random variable P is a generic busy period of the first buffer. It can be shown as in [5,7] that $\pi(s)$ is the unique solution in the unit circle of the equation

$$\pi(s) = \gamma\left(\frac{s + \lambda(1 - \pi(s))}{c_1}, s\right). \tag{2.11}$$

Note that it follows immediately from (2.11) that

$$\mathbb{E}\{P\} = \frac{c_1\mathbb{E}\{A\} + \mathbb{E}\{X(A)\}}{c_1 - \lambda\mathbb{E}\{X(A)\}}. \tag{2.12}$$

3. THE JOINT STEADY-STATE BUFFER-CONTENT DISTRIBUTION

In this section, we give our main result, which is an explicit expression for the transform $\omega(u,v)$ of the steady-state buffer-content distribution. This expression is obtained in two steps: First, we give a decomposition property of V, which reduces the problem to the computation of the steady-state distribution of V, *given* that the source is Off. In the second step, this problem is solved following the approach in [15] by applying the martingale that was introduced in [19].

For the first step, we define $J(t)$ to be a 0–1 variable, which equals 1 if the source is On at time t (i.e., if the content of the first buffer is increasing) and 0 otherwise. Clearly, in steady state, $J(t)$ is distributed as a random variable J, which is 1 with probability p and 0 with probability $1 - p$, where p is given in (2.2). Also, we introduce the process $\bar{\mathcal{V}} = \{\bar{V}(t), t \geq 0\}$, with $\bar{V}(t) = (\bar{V}_1(t), \bar{V}_2(t))$, as the process obtained from $\mathcal{V}$ after deleting the On periods. As an illustration of this "deleting procedure," we refer to Figure 1, rather than giving the precise details. We refer to [17] or [27] for a detailed description of this procedure in the single-node case.

It can be shown that this process $\bar{\mathcal{V}}$ also has a steady-state distribution. Let $\bar{V} = (\bar{V}_1, \bar{V}_2)$ denote a generic random vector with this distribution, and let $\bar{\omega}(u,v)$ denote the corresponding LST. We are now ready to state the decomposition result, where we will use $\stackrel{d}{=}$ to indicate equality in distribution.

THEOREM 3.1: *The stationary buffer content V can be written as*

$$V \stackrel{d}{=} \bar{V} + J \times (X(A^*), (c_1 - c_2)A^*), \tag{3.1}$$

where $\bar{V}$, J, *and* $(X(A^*), (c_1 - c_2)A^*)$ *are independent. In terms of transforms,*

$$\omega(u,v) = \bar{\omega}(u,v)(1 - p + p\gamma^*(u, (c_1 - c_2)v)). \tag{3.2}$$

PROOF: Note that J can be identified with the indicator of the event that the input source is On in steady state. Observe that $(V \mid J = 0) \stackrel{d}{=} \bar{V}$. Using PASTA, the steady-state buffer-content distribution observed at the end of Off periods has the same distribution as $\bar{V}$. This implies (using the definitions of A^* and $X(A^*)$) that

$$(V \mid J = 1) \stackrel{d}{=} \bar{V} + (X(A^*), (c_1 - c_2)A^*),$$

with $\bar{V}$ and $(X(A^*), (c_1 - c_2)A^*)$ independent. Combining these results yields (3.1), from which (3.2) follows easily. ■

In view of this, it suffices to compute $\bar{\omega}(u,v)$. Hence, in the remainder of this section, we concentrate on the steady-state distribution of $\bar{\mathcal{V}}$.

The crucial observation is that $\bar{\mathcal{V}}$ can be identified with the joint buffer-content process of a tandem network with dependent Lévy input as studied in [15]. In order to apply the results of [15], we define $Z_1(t) = \bar{V}_1(t)$, $Z_2(t) = \bar{V}_1(t) + \bar{V}_2(t)$, and $Z(t) = (Z_1(t), Z_2(t))$. Observe that $\{Z_2(t)\}$ can be identified with the buffer-content process of an $M/G/1$ queue with Poisson(λ) arrivals, generic service time $X(A) + (c_1 - c_2)A$, and service speed c_2.

We now find the following useful martingale from the fact that $\{Z(t)\}$ is a two-dimensional reflected Lévy process (cf. Lemma 2.1 of [15]).

LEMMA 3.1: *The process* $\mathcal{M} = \{M(t)\}$, *given by*

$$\begin{aligned} M(t) = \phi(u,v) \int_0^t e^{-uZ_1(s) - vZ_2(s)} ds + 1 - e^{-uZ_1(t) - vZ_2(t)} \\ - uc_1 \int_0^t e^{-vZ_2(s)} 1_{[Z_1(s)=0]}\, ds - vc_2 \int_0^t 1_{[Z_2(s)=0]}\, ds, \end{aligned} \tag{3.3}$$

with

$$\phi(u,v) = uc_1 + vc_2 - \lambda(1 - \gamma(u + v, (c_1 - c_2)v)),$$

is a martingale.

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

PROOF: Let $Y(t) = (Y_1(t), Y_2(t))$ be a two-dimensional Lévy process with exponent $\phi(u,v)$; that is,

$$\mathbb{E}\{e^{-uY_1(t)-vY_2(t)}\} = e^{\phi(u,v)t}.$$

Furthermore, we define

$$I_i(t) = \max_{0\le s\le t}(-Y_i(s)), \qquad i = 1,2.$$

Then, $Z(t)$ may be represented as follows (note that $Z(0) \equiv 0$):

$$Z_i(t) = Y_i(t) + I_i(t), \qquad i = 1,2.$$

Noting that $dI_i(t) = c_i 1_{[Z_i(t)=0]}\, dt$ for $i = 1,2$, the lemma follows from [19]. ■

Using this martingale, it is possible to obtain an expression for the LST of the stationary distribution of $\{Z(t)\}$, which is given in the following theorem.

THEOREM 3.2: *The joint LST of Z is given by*

$$\mathbb{E}\{e^{-uZ_1-vZ_2}\} = \frac{u(\lambda\mathbb{E}\{X(A)\} - c_1)\mathbb{E}\{e^{-vZ_2} | Z_1 = 0\} + v(\lambda\mathbb{E}\{X(A)\} + \lambda(c_1 - c_2)\mathbb{E}\{A\} - c_2)}{uc_1 + vc_2 - \lambda(1 - \gamma(u+v,(c_1-c_2)v))}.$$

PROOF: We mimic the proof of Corollary 2.3 in [15]. As a stopping time we take some epoch T with $Z_1(T) = Z_2(T) = 0$. Applying Doob's optional stopping theorem as in [15] and using regenerative process theory as in (2.5), one gets for Re $u,v \ge 0$,

$$\phi(u,v)\mathbb{E}\{e^{-uZ_1-vZ_2}\} = u\,\mathbb{P}\{Z_1 = 0\}\,\mathbb{E}\{e^{-vZ_2} | Z_1 = 0\} + v\,\mathbb{P}\{Z_2 = 0\}\,\mathbb{E}\{e^{-uZ_1} | Z_2 = 0\}. \tag{3.4}$$

Keeping the definitions of Z_1 and Z_2 in mind, the two respective probabilities in (3.4) can be found by dividing the right-hand sides of (2.6) and (2.7) by $1 - p$. The result now follows after noting that $\mathbb{E}\{e^{-uZ_1} | Z_2 = 0\} = 1$. ■

The translation of Theorem 3.2 to the transform of $(\bar{V}_1, \bar{V}_2)$ is done by noting that $\mathbb{E}\{e^{-uZ_1-vZ_2}\} = \bar{\omega}(u+v,v)$ and $\mathbb{E}\{e^{-vZ_2} | Z_1 = 0\} = \mathbb{E}\{e^{-v\bar{V}_2} | \bar{V}_1 = 0\}$. Hence, the only unknown we have to find is $\mathbb{E}\{e^{-v\bar{V}_2} | \bar{V}_1 = 0\}$. By Theorem 3.1, $(\bar{V}_1, \bar{V}_2) \stackrel{d}{=} ((V_1, V_2) | J = 0)$. Hence (noting that $V_1 = 0$ implies $J = 0$),

$$\mathbb{E}\{e^{-v\bar{V}_2} | \bar{V}_1 = 0\} = \mathbb{E}\{e^{-vV_2} | V_1 = 0\}.$$

Note that the second buffer can be identified with a fluid queue fed by a single On–Off source having constant input rate c_1 during On periods. These On periods are busy periods of the first buffer. Appropriately scaling time, such that the output rate becomes 1, this means that the distribution of $(V_2 | V_1 = 0)$ can be identified with

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

the steady-state workload distribution of an $M/G/1$ queue with arrival rate λ/c_2 and service times $(c_1 - c_2)P$ (see also [1,18]). Hence, we have

$$\mathbb{E}\{e^{-vV_2} \mid V_1 = 0\} = \frac{(c_2 - \lambda(c_1 - c_2)\mathbb{E}\{P\})v}{c_2 v - \lambda(1 - \pi((c_1 - c_2)v))}. \tag{3.5}$$

If we combine our findings, we arrive at the main conclusion of this section:

THEOREM 3.3: *The LST of (V_1, V_2) is given by*

$$\omega(u,v) = \bar{\omega}(u,v)(1 - p + p\gamma^*(u,(c_1 - c_2)v)),$$

with

$$\bar{\omega}(u,v) = \frac{(u-v)(\lambda\mathbb{E}\{X(A)\} - c_1)\mathbb{E}\{\mathrm{e}^{-vV_2}|V_1 = 0\} + v(\lambda\mathbb{E}\{X(A)\} + \lambda(c_1 - c_2)\mathbb{E}\{A\} - c_2)}{(u-v)c_1 + vc_2 - \lambda(1 - \gamma(u,(c_1 - c_2)v))},$$

and $\mathbb{E}\{e^{-vV_2}|V_1 = 0\}$ given in (3.5).

From Theorem 3.3, it is straightforward to derive expressions for the moments, marginal distributions, and correlations. To compute the original steady-state probabilities from Theorem 3.3, one may use the multidimensional transform-inversion technique described in Choudhury, Lucantoni, and Whitt [9].

We end this section with a brief outline of how to extend Theorem 3.3 to the multinode tandem case. Consider n nodes with capacities $c_1 > c_2 > \cdots > c_n$ and assume that the stability condition (2.4) holds with c_2 replaced by c_n. If we let V_i denote the steady-state buffer content of buffer i, $i = 1, \ldots, n$, we find a decomposition result as in Theorem 3.1, which leads to

$$\begin{aligned}\omega(u_1, \ldots, u_n) &\equiv \mathbb{E}\{e^{-u_1V_1 - \cdots - u_nV_n}\} \\ &= \bar{\omega}(u_1, \ldots, u_n)(1 - p + p\gamma^*(u_1, (c_1 - c_2)u_2 + \cdots + (c_{n-1} - c_n)u_n)),\end{aligned}$$

where $\bar{\omega}$ is defined in the obvious way. To find $\bar{\omega}$, one can study the multidimensional martingale $\mathcal{M} = \{M(t)\}$, given by

$$\begin{aligned}M(t) = \phi(u_1, \ldots, u_n)&\int_0^t e^{-u_1Z_1(s) - \cdots - u_nZ_n(s)}\,ds + 1 - e^{-u_1Z_1(t) - \cdots - u_nZ_n(t)} \\ &- u_1c_1\int_0^t e^{-u_2Z_2(s) - \cdots - u_nZ_n(s)}1_{[Z_1(s)=0]}\,ds \\ &- u_2c_2\int_0^t e^{-u_3Z_3(s) - \cdots - u_nZ_n(s)}1_{[Z_2(s)=0]}\,ds - \cdots - u_nc_n\int_0^t 1_{[Z_n(s)=0]}\,ds,\end{aligned} \tag{3.6}$$

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

where the $Z_i(t), i = 1,\dots,n$, are defined similarly as earlier and ϕ is given by

$$\phi(u_1,\dots,u_n) = u_1 c_1 + \cdots + u_n c_n - \lambda(1 - \gamma(u_1 + \cdots + u_n, (c_1 - c_2)u_2 + \cdots + (c_{n-1} - c_n)u_n)).$$

This martingale leads to a generalized version of (3.4) (one can also directly apply Corollary 2.3 of [15]). This equation can be solved in a similar way as in the proof of Theorem 3.2.

4. EXAMPLES

In the previous section, we derived an expression for $\omega(u,v)$ in terms of $\gamma(\cdot,\cdot)$, $\gamma^*(\cdot,\cdot)$, and $\pi(\cdot)$. The main goal of this section is to give some examples of the input process $\mathcal{X}$ for which it is possible to get tractable expressions for these transforms. Together with Theorem 3.3, this provides an explicit expression for $\omega(u,v)$ in these cases. In the next two subsections, we treat (i) input from an On–Off source and (ii) semi-Markov input.

4.1. Input from a Simple On–Off Source

Our first example, which was the original motivation for this work, is the case where the first buffer is fed by a single On–Off source. If this source is On, it feeds fluid into the first buffer with constant rate $r > c_1$. For this special case, we take $X(t) = (r - c_1)t$, $t \geq 0$. If we denote the LST of A by $\alpha(s) = \mathbb{E}\{e^{-sA}\}$, we get

$$\mathbb{E}\{X(A)\} = (r - c_1)\mathbb{E}\{A\}, \tag{4.1}$$

$$\gamma(u,(c_1 - c_2)v) = \alpha((r - c_1)u + (c_1 - c_2)v), \tag{4.2}$$

$$\gamma^*(u,(c_1 - c_2)v) = \frac{1 - \alpha((r - c_1)u + (c_1 - c_2)v)}{\mathbb{E}\{A\}((r - c_1)u + (c_1 - c_2)v)}. \tag{4.3}$$

The latter equation follows immediately from the obvious identity $(X(A^*),A^*) \equiv ((r - c_1)A^*,A^*)$. An explicit expression for the LST of V_1 and V_2 follows by combining (4.2) and (4.3) with Theorem 3.3. Finally, $\pi(\cdot)$ follows from

$$\pi(s) = \alpha((r - c_1)(s + \lambda(1 - \pi(s))) + s).$$

Several other studies contain results for this canocical model which are strongly related to the problem addressed here: The marginal distributions of V_1 and V_2 and the correlation between V_1 and V_2 have been computed in [1]. The joint distribution of (V_1,V_2) in case A has a phase-type distribution has been found in [16]. When A is exponentially distributed, it is possible to invert ω to find an expression for the distribution of V, see [20,23].

4.2. Semi-Markov Input

In this subsection, we assume that the content of the first buffer is regulated by a semi-Markov process. This is motivated by a recent study [6], in which a single fluid buffer is analyzed that is fed by the same type of input. Hence, we will follow [6] and consider the Markov renewal process $\{(Y_n, \tau_{n+1}) \mid n \geq 0\}$ with state space $\{0, \dots, K\} \times [0, \infty)$. We let $T_0 = 0$ and $T_n = \sum_{k=1}^{n} \tau_k$, $n \geq 1$, and introduce the corresponding counting process by $N(t) = \sup\{n : T_n \leq t\}$. Then the semi-Markov process (SMP) $\{Y(t), t \geq 0\}$ is defined by $Y(t) = Y_{N(t)}$. The behavior of this process is given by the stochastic matrix P consisting of the transition probabilities $p_{ij} = \mathbb{P}\{Y_1 = j \mid Y_0 = i\}$, $0 \leq i, j \leq K$ (we assume that $p_{ii} = 0$), and the functions $F_{ij}(t)$, defined by

$$F_{ij}(t) = \mathbb{P}\{\tau_1 \leq t \mid Y_0 = i, Y_1 = j\}. \tag{4.4}$$

It is convenient to also define

$$\begin{aligned}
F_i(t) &= \mathbb{P}\{\tau_1 \leq t \mid Y_0 = i\} = \sum_{j=0}^{K} p_{ij} F_{ij}(t), \\
\mathbb{E}_{ij}\{\cdot\} &= \mathbb{E}\{\cdot \mid Y_0 = i, Y_1 = j\}, \\
\mathbb{E}_i\{\cdot\} &= \mathbb{E}\{\cdot \mid Y_0 = i\}, \\
\tau_{ij}(u) &= \mathbb{E}_{ij}\{e^{-u\tau_1}\}, \\
\tau_i(u) &= \mathbb{E}_i\{e^{-u\tau_1}\}, \\
\tau_i^e(u) &= \frac{1 - \tau_i(u)}{u\mathbb{E}_i\{\tau_1\}}, \\
m_{ij} &= \mathbb{E}_{ij}\{\tau_1\}, \\
m_i &= \mathbb{E}_i\{\tau_1\}.
\end{aligned}$$

An important assumption is that the sojourn time in state 0 (say) is exponentially distributed and independent of the next jump [i.e., $F_{0j}(t) = F_0(t) = 1 - e^{-\lambda t}$].

The SMP regulates the content of the first buffer in our tandem queue in the following way. If $Y(t) = i$, $i \geq 1$, then the buffer content increases at rate $q_i = r_i - c_1$, where $r_i \geq c_1$. When the SMP is in the special state 0, the buffer content decreases at rate c_1. Hence, we can construct our process $\mathcal{X}$ as follows. Suppose that the SMP jumps from state 0 at time 0. Then,

$$X(t) = \int_0^t q_{Y(u)}\, du, \quad t \geq 0,$$

$$A = \inf\{t > 0 : Y(t) = 0\}.$$

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

We now compute the LSTs of $(X(A), A)$ and $(X(A^*), A^*)$, extending the approach of [6] by which the marginal LSTs of $X(A)$ and $X(A^*)$ were found. Keeping (2.10) in mind, we define for $1 \le i \le K$,

$$\beta_i^*(u,v) = \mathbb{E}_i\left\{\int_0^A e^{-u\int_0^t q_{Y(s)}\,ds - vt}\,dt\right\}. \tag{4.5}$$

By conditioning upon Y_1 and τ_1, we obtain

$$\beta_i^*(u,v) = m_i \tau_i^e(q_i u + v) + \sum_{j=1}^K p_{ij}\tau_{ij}(q_i u + v)\beta_j^*(u,v), \quad 1 \le i \le K. \tag{4.6}$$

This system of equations has a unique solution. To obtain an expression for $\gamma^*(u,v)$, note that $\mathbb{E}\{A\}$ can be computed as

$$\mathbb{E}\{A\} = \sum_{i=1}^K p_{0i}\, a_i, \tag{4.7}$$

where the $a_i = \mathbb{E}_i\{A\}, i = 1,\ldots,K$, form the unique solution of

$$a_i = m_i + \sum_{j=1}^K p_{ij}\, a_j. \tag{4.8}$$

Combining (2.10), (4.6), and (4.7), we obtain

$$\gamma^*(u,v) = \frac{\sum_{j=1}^K p_{0j}\,\beta_j^*(u,v)}{\sum_{j=1}^K p_{0j}\,a_j}. \tag{4.9}$$

The computation of γ is similar but easier (see also [6]) so we only state the final result: γ can be written as

$$\gamma(u,v) = \sum_{j=1}^K p_{0j}\,\beta_j(u,v), \tag{4.10}$$

with $\beta_j(u,v), j = 1,\ldots,K$, the unique solution of

$$\beta_i(u,v) = p_{i0}\tau_{i0}(q_i u + v) + \sum_{j=1}^K p_{ij}\tau_{ij}(q_i u + v)\beta_j(u,v), \quad 1 \le i \le K. \tag{4.11}$$

Recursive expressions for the moments of A, $X(A)$, and $X(A^*)$ can be found in [6].

5. FINITE BUFFERS

In this section, we look at the case in which the buffers have respective sizes K_1 and K_2. Using obvious notation, we will denote the transient process that describes both

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

buffer contents by $\mathcal{V}^{K_1,K_2}$. It can be shown that this process has a stationary distribution and we let $V^{K_1,K_2} = (V_1^{K_1,K_2}, V_2^{K_1,K_2})$ be distributed accordingly.

The main result of this section is Theorem 5.1. In this theorem, we relate the steady-state distribution of $\mathcal{V}^{K_1,K_2}$ to that of $\mathcal{V}$. Hence, it is still assumed that (2.4) holds, even though this is no longer required for stability. Furthermore, we need to make the following additional assumption.

ASSUMPTION 5.1: K_1, K_2, and $\mathcal{X}$ *are such that the second buffer fills before the first one does*; *that is*, *for all* t,

$$\mathbb{P}\{V_2^{K_1,K_2}(t) = K_2 \mid V_1^{K_1,K_2}(t) = K_1\} = 1.$$

If $X(t) \equiv (r - c_1)t$ (the scenario considered in Section 4.1), and if the system is empty at time $t = 0$, this assumption is satisfied iff

$$\frac{K_1}{r - c_1} \geq \frac{K_2}{c_1 - c_2}. \tag{5.1}$$

A similar characterization holds for the model considered in Section 4.2.

The main result of this section now states that the distributions of V and V^{K_1,K_2} are *proportional*:

THEOREM 5.1: *If Assumption 5.1 holds, then for* $0 \leq x < K_1$, $0 \leq y < K_2$,

$$\mathbb{P}\{V_1^{K_1,K_2} \leq x; V_2^{K_1,K_2} \leq y\} = \frac{\mathbb{P}\{V_1 \leq x; V_2 \leq y\}}{\mathbb{P}\{W_1 \leq K_1\}\mathbb{P}\{W_2 \leq K_2\}}, \tag{5.2}$$

with $W_1 \stackrel{d}{=} (V_1 \mid J = 0)$ *and* $W_2 \stackrel{d}{=} (V_2 \mid V_1 = 0)$.

Both this theorem and its proof below are an extension of the single node case which is treated in [27].

PROOF: The proof consists of two steps:

1. First, we consider the fluid tandem queue with buffer sizes $K_1 = \infty$ and $K_2 < \infty$. Denote this process by $\mathcal{V}^{\infty,K_2}$ and let $V^{\infty,K_2} = (V_1^{\infty,K_2}, V_2^{\infty,K_2})$ be distributed according to its stationary distribution. We show that, for $y < K_2$,

$$\mathbb{P}\{V_1^{\infty,K_2} \leq x; V_2^{\infty,K_2} \leq y\} = \frac{\mathbb{P}\{V_1 \leq x; V_2 \leq y\}}{\mathbb{P}\{W_2 \leq K_2\}}. \tag{5.3}$$

2. In our second step, we show that if Assumption 5.1 holds, for $x < K_1$ and $y < K_2$,

$$\mathbb{P}\{V_1^{K_1,K_2} \leq x; V_2^{K_1,K_2} \leq y\} = \frac{\mathbb{P}\{V_1^{\infty,K_2} \leq x; V_2^{\infty,K_2} \leq y\}}{\mathbb{P}\{W_1 \leq K_1\}}. \tag{5.4}$$

The proof is then completed by combining (5.3) and (5.4).

Step 1. From each sample path of $\mathcal{V}$, we construct a sample path of $\mathcal{V}^{\infty, K_2}$. This construction is done as follows (see also Fig. 2): Given a sample path of $\mathcal{V}$, consider the excursions of $\{V_2(t)\}$ above level K_2. These excursions consist of two parts (a) and (b), corresponding to (a) and (b) in Figure 2:

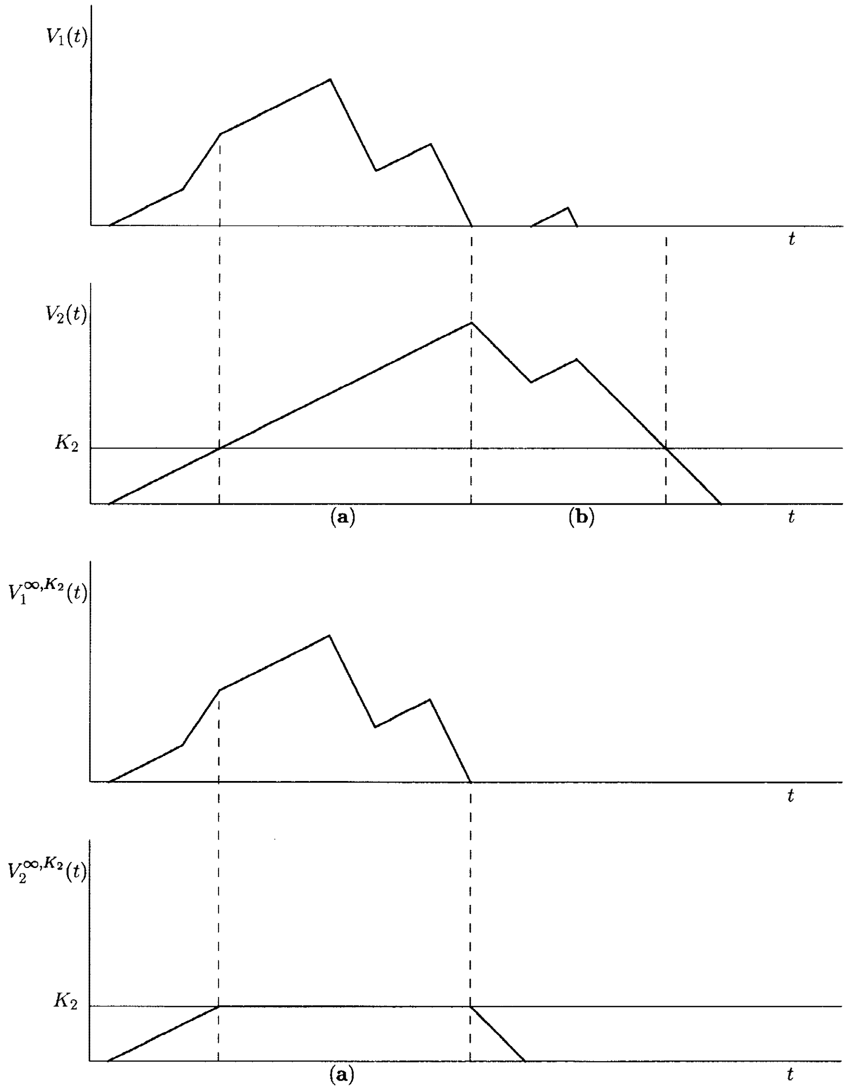

FIGURE 2. Construction of the process $\mathcal{V}^{\infty, K_2}$ from $\mathcal{V}$.

https://doi.org/10.1017/S0269964802161031 Published online by Cambridge University Press

(a) is the time it takes until the first buffer becomes empty (i.e., the remaining busy period of the first buffer)
(b) is the remaining part of the excursion.

Now, construct a sample path of a process $\hat{\mathcal{V}}$ from a sample path of $\mathcal{V}$ as follows:

- Time epochs where $V_2(t) \leq K_2$ remain unchanged.
- Part (a) of the excursions as described earlier is modified as follows: $\hat{V}_2(t) = K_2$ and $\hat{V}_1(t) = V_1(t)$;
- Delete the remaining parts of the excursions of $V_2(t)$.

The constructed process $\hat{\mathcal{V}}$ has the same law as $\mathcal{V}^{\infty,K_2}$: Every time $\hat{V}_2(t)$ leaves state K_2, the environment process is Off (in fact, the first buffer is empty), and the remaining Off time is exponentially distributed with rate λ—as it should be. Henceforth, take $\mathcal{V}^{K_1,K_2} \equiv \hat{\mathcal{V}}$. Denote a regeneration cycle of this process by C^{∞,K_2}. An immediate consequence of the construction of $\mathcal{V}^{K_1,K_2}$ is that, sample-path-wise, for $y < K_2$,

$$\int_0^{C^{\infty,K_2}} 1_{[V_1^{\infty,K_2}(t) \leq x, V_2^{\infty,K_2}(t) \leq y]}\, dt = \int_0^{C} 1_{[V_1(t) \leq x, V_2(t) \leq y]}\, dt. \tag{5.5}$$

Combining this with regenerative process theory (as in (2.5)), we get, for all x and $y < K_2$,

$$\mathbb{P}\{V_1^{\infty,K_2} \leq x; V_2^{\infty,K_2} \leq y\} = \frac{\mathbb{E}\{C\}}{\mathbb{E}\{C^{\infty,K_2}\}}\, \mathbb{P}\{V_1 \leq x; V_2 \leq y\}. \tag{5.6}$$

In particular, for $x \to \infty$, we get

$$\mathbb{P}\{V_2^{\infty,K_2} \leq y\} = \frac{\mathbb{E}\{C\}}{\mathbb{E}\{C^{\infty,K_2}\}}\, \mathbb{P}\{V_2 \leq y\}. \tag{5.7}$$

From Theorem 5.2 of [27], we obtain

$$\frac{\mathbb{E}\{C\}}{\mathbb{E}\{C^{\infty,K_2}\}} = \frac{1}{\mathbb{P}\{W_2 \leq K_2\}},$$

which proves (5.3).

Step 2. This step is similar to Step 1 and gives a sample-path construction of the process $\mathcal{V}^{K_1,K_2}$ from $\mathcal{V}^{\infty,K_2}$. For each sample path of the latter process, consider the excursions of $\{V_1^{\infty,K_2}(t)\}$ above level K_1. Note that Assumption 5.1 ensures that the second buffer is full during these excursions (our method would break down if this would not be the case). As earlier, divide the excursions into two parts; the first part ends when an Off period is finished. Truncate the first part of the excursion of $V_1^{\infty,K_2}(t)$ to K_1 (while $V_2^{\infty,K_2}(t)$ remains unchanged) and delete the second part of the excursion.

Due to exactly the same argument as in Step 1, the constructed process can be identified with $\mathcal{V}^{K_1,K_2}$. This construction of $\mathcal{V}^{K_1,K_2}$ implies that, sample-path-wise,

$$\int_0^{C^{K_1,K_2}} 1_{[V_1^{K_1,K_2}(t)\le x,\, V_2^{K_1,K_2}(t)\le y]}\, dt = \int_0^{C^{\infty,K_2}} 1_{[V_1^{\infty,K_2}(t)\le x,\, V_2^{\infty,K_2}(t)\le y]}\, dt. \tag{5.8}$$

Using regenerative process theory, this implies

$$\mathbb{P}\{V_1^{K_1,K_2} \le x; V_2^{K_1,K_2} \le y\} = \frac{\mathbb{E}\{C^{\infty,K_2}\}}{\mathbb{E}\{C^{K_1,K_2}\}} \mathbb{P}\{V_1^{\infty,K_2} \le x; V_2^{\infty,K_2} \le y\}. \tag{5.9}$$

What remains is to identify the prefactor on the right-hand side of (5.9). From Theorem 5.2 of [27], it follows that for $x < K_2$,

$$\mathbb{P}\{V_1^{K_1,K_2} \le x\} = \frac{\mathbb{P}\{V_1 \le x\}}{\mathbb{P}\{W_1 \le K_1\}} = \frac{\mathbb{P}\{V_1^{\infty,K_2} \le x\}}{\mathbb{P}\{W_1 \le K_1\}}. \tag{5.10}$$

Also, note that

$$\mathbb{P}\{V_1^{K_1,K_2} = 0, V_2^{K_1,K_2} = K_2\} = \mathbb{P}\{V_1^{\infty,K_2} = 0, V_2^{\infty,K_2} = K_2\} = 0. \tag{5.11}$$

Combining (5.10) and (5.11), we obtain

$$\begin{aligned}\mathbb{P}\{V_1^{K_1,K_2} = 0; V_2^{K_1,K_2} < K_2\} &= \mathbb{P}\{V_1^{K_1,K_2} = 0\} \\ &= \frac{\mathbb{P}\{V_1^{\infty,K_2} = 0\}}{\mathbb{P}\{W_1 \le K_1\}} \\ &= \frac{\mathbb{P}\{V_1^{\infty,K_2} = 0; V_2^{\infty,K_2} < K_2\}}{\mathbb{P}\{W_1 \le K_1\}}.\end{aligned}$$

Invoking (5.9) for $x = 0$ and $y = K_2$ yields that the unknown prefactor in (5.9) equals $\mathbb{P}\{W_1 \le K_1\}^{-1}$. This completes Step 2 and the proof of the theorem. ■

It can be shown that analogs of Theorem 5.1 also hold for the networks considered in [15,16,18], after obvious modifications of Assumption 5.1. These results may be derived in a similar way as Theorem 5.1.

Acknowledgments

This research was partly carried out when the first author had a postdoctoral position at Eindhoven University of Technology, supported by the Netherlands Organization for Scientific Research (NWO). The research of the second author is supported by an NWO grant. Both authors would like to thank Onno Boxma for helpful comments.

References

1. Aalto, S. & Scheinhardt, W.R.W. (2000). Tandem fluid queues fed by homogeneous On-Off sources. *Operations Research Letters* 27: 73–82.
2. Asmussen, S. (1987). *Applied probability and queues*. New York: Wiley.

3. Boots, N.K. & Tijms, H.C. (1999). A multi-server queueing system with impatient customers. *Management Science* 45: 444-448.
4. Boots, N.K. & Tijms, H.C. (1999). An $M/M/c$ queue with impatient customers. *Top* 7: 213–221.
5. Boxma, O.J. & Dumas, V. (1998). Fluid queues with heavy-tailed activity period distributions. *Computer Communications* 21: 1509–1529.
6. Boxma, O.J., Kella, O., & Perry, D. (2001). An intermittent fluid system with exponential on-times and semi-Markov input rates. *Probability in the Engineering and Informational Sciences* 15: 189–198.
7. Boxma, O.J., Perry, D., & van der Duyn Schouten, F. (1999). Fluid queues and mountain processes. *Probability in the Engineering and Informational Sciences* 13: 407–427.
8. Boxma, O.J. & Resing, J. (1995). Tandem queues with deterministic service times. *Annals of Operations Research* 49: 221–239.
9. Choudhury, G., Lucantoni, D., & Whitt, W. (1994). Multidimensional transform inversion with applications to the transient $M/G/1$ queue. *Annals of Applied Probability* 4: 719–740.
10. Cohen, J.W. (1976). *Regenerative processes in queueing theory.* Berlin: Springer-Verlag.
11. Gouweleeuw, F. & Tijms, H.C. (1998). Computing loss probabilities in discrete-time queues. *Operations Research* 46: 149–154.
12. Hooghiemstra, G. (1987). A path construction for the virtual waiting time of an $M/G/1$ queue. *Statistica Neerlandica* 41: 175–181.
13. Keilson, J. & Servi, L.D. (1989). Blocking probability for $M/G/1$ vacation systems with occupancy level dependent schedules. *Operations Research* 37: 134–140.
14. Keilson, J. & Servi, L.D. (1993). The $M/G/1/K$ blocking formula and its generalizations to state-dependent vacation systems and priority systems. *Queueing Systems* 14: 111–123.
15. Kella, O. (1993). Parallel and tandem fluid networks with dependent Lévy inputs. *Annals of Applied Probability* 3: 682–695.
16. Kella, O. (2001). Markov-modulated feedforward fluid networks. *Queueing Systems* 37: 141–161.
17. Kella, O. & Whitt, W. (1992). A storage model with a two-state random environment. *Operations Research* 40: S257–S262.
18. Kella, O. & Whitt, W. (1992). A tandem fluid network with Lévy input. In I. Basawa & U. Bhat (eds.), *Queues and related models*. Oxford: Oxford University Press, pp. 112–128.
19. Kella, O. & Whitt, W. (1992). Useful martingales for stochastic storage processes with Lévy input. *Journal of Applied Probability* 29: 396–403.
20. Kroese, D.P. & Scheinhardt, W.R.W. (2001). Joint distributions for interacting fluid queues. *Queueing Systems* 37: 99–139.
21. Park, K. & Willinger, W. (eds.) (2000). *Self-similar network traffic and performance evaluation.* New York: Wiley.
22. Rubin, I. (1974). Path delays in communication networks. *Applied Mathematics and Optimization* 1: 193–221.
23. Scheinhardt, W.R.W. (1998). Markov-modulated and feedback fluid queues. [Unpublished] Ph.D. thesis, University of Twente, Twente, The Netherlands. Also available at http://www.ub.utwente.nl/webdocs/tw/1/t0000008.pdf.
24. Shalmon, M. & Kaplan, M.A. (1984). A tandem network of queues with deterministic service and intermediate arrivals. *Operations Research* 32: 753–773.
25. Stidham, S. (1986). Clearing systems and (s, S) inventory systems with nonlinear costs and positive lead times. *Operations Research* 34: 276–280.
26. van Uitert, M. & Borst, S.C. (2001). Generalized processor sharing networks fed by heavy-tailed traffic flows. Proceedings of Infocom 2001, Anchorage, AK, pp. 269–278.
27. Zwart, A.P. (2000). A fluid queue with a finite buffer and subexponential input. *Advances in Applied Probability* 32: 221–243.