Centrum voor Wiskunde en Informatica

REPORTRAPPORT

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

A feedback fluid queue with two congestion control thresholds

R. Malhotra, M.R.H. Mandjes, W.R.W. Scheinhardt,
J.L. van den Berg

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

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

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

A feedback fluid queue with two congestion control thresholds

ABSTRACT

Feedback fluid queues play an important role in modeling congestion control mechanisms for packet networks. In this paper we present and analyze a fluid queue with a feedback-based traffic rate adaptation scheme which uses two thresholds. The higher threshold B_{1} is used to signal the beginning of congestion while the lower threshold B_{2} signals the end of congestion. These two parameters together allow to make the trade--off between maximizing throughput performance and minimizing delay. The difference between the two thresholds helps to control the amount of feedback signals sent to the traffic source. In our model the input source can behave like either of two Markov fluid processes. The first applies as long as the upper threshold B_{1} has not been hit from below. As soon as that happens, the traffic source adapts and switches to the second process, until B_{2} (smaller than B_1) is hit from above. We analyze the model by setting up the Kolmogorov forward equations, then solving the corresponding balance equations using a spectral expansion, and finally identifying sufficient constraints to solve for the unknowns in the solution. In particular, our analysis yields expressions for the stationary distribution of the buffer occupancy, the buffer delay distribution, and the throughput.

A Feedback Fluid Queue with
Two Congestion Control Thresholds

R. Malhotra, M.R.H. Mandjes, W.R.W. Scheinhardt, J.L. van den Berg *

Abstract

Feedback fluid queues play an important role in modeling congestion control mechanisms for packet networks. In this paper we present and analyze a fluid queue with a feedback-based traffic rate adaptation scheme which uses two thresholds. The higher threshold B_1 is used to signal the beginning of congestion while the lower threshold B_2 signals the end of congestion. These two parameters together allow to make the trade–off between maximizing throughput performance and minimizing delay. The difference between the two thresholds helps to control the amount of feedback signals sent to the traffic source. In our model the input source can behave like either of two Markov fluid processes. The first applies as long as the upper threshold B_1 has not been hit from below. As soon as that happens, the traffic source adapts and switches to the second process, until B_2 (smaller than B_1) is hit from above. We analyze the model by setting up the Kolmogorov forward equations, then solving the corresponding balance equations using a spectral expansion, and finally identifying sufficient constraints to solve for the unknowns in the solution. In particular, our analysis yields expressions for the stationary distribution of the buffer occupancy, the buffer delay distribution, and the throughput.

FLUID QUEUES ● FEEDBACK REGULATION ● CONGESTION CONTROL ● SPECTRAL EXPANSION

1 Introduction

In performance analysis of communication systems, fluid queues are a frequently used modeling framework. These models approximate packet streams by flows of fluid. *Feedback* fluid queues are of special significance when modeling congestion control mechanisms, in particular closed-loop controls in which the input process is affected by the current value of the buffer content. Practical examples are TCP congestion control, random early detection [8], explicit congestion notification [7] and Ethernet congestion control mechanisms [13].

---

*R. Malhotra is with Alcatel-Lucent, Capitool 5, 7521 PL Enschede, the Netherlands, `rimalhotra@alcatel-lucent.com`, and also with University of Twente, the Netherlands. M.R.H. Mandjes is with Korteweg-de Vries Institute for Mathematics, Plantage Muidergracht 24, 1018 TV Amsterdam, the Netherlands, `mmandjes@science.uva.nl`, and also with CWI, Amsterdam, the Netherlands, and EURANDOM, Eindhoven, the Netherlands. W.R.W. Scheinhardt is with University of Twente, Faculty of Electrical Engineering, Mathematics, and Computer Science, P.O. Box 217, 7500 AE Enschede, the Netherlands, `w.r.w.scheinhardt@utwente.nl`, and also with CWI, Amsterdam, the Netherlands. J.L. van den Berg is with TNO ICT, P.O. Box 5050, 2600 GB Delft, the Netherlands, `j.l.vandenberg@tno.nl`, and also with University of Twente, the Netherlands.

Previous work on feedback fluid queues [15] predominantly focused on queues in which a *single* buffer threshold signals both the onset and end of congestion. In these models, depending on whether the buffer occupancy is below or above the threshold, the traffic source is allowed to transmit at a high peak rate or slowed down to a (lower) guaranteed traffic rate, respectively. The special case at which the threshold lies at 0 was also addressed in, e.g., [4, 17].

These 'single-threshold systems' have serious drawbacks. In the first place, in case the threshold is crossed often, many feedback signals have to be sent to the traffic source, and these may consume a significant part of the available bandwidth. In the second place, it can be argued that it is not optimal that a single threshold performs the function of signaling the beginning and end of congestion: to maximize throughput a full buffer is preferred (thus delaying the signal for the onset of congestion as much as possible), whereas to minimize the buffer delay requires minimizing the buffer occupancy (thus delaying the signal for the end of the congestion phase as much as possible). Therefore, to find a good trade-off between throughput and buffer delay, there is a need for mechanisms with *two* separate thresholds: one to signal the beginning of the congestion phase, and an other to signal the end. In this paper we propose and analyze such a mechanism.

The model addressed in this paper belongs to the class of *(Markovian) fluid models*. The 'classical' fluid model [3, 11] is characterized by a generator matrix Q governing a Markovian background process and a diagonal matrix $R = \text{diag}\{r_1, ..., r_d\}$: if the background process is in state i, traffic is generated at a rate $r_i \geq 0$. It was shown that the steady-state buffer content distribution obeys a system of linear differential equations, and after imposing the proper additional constraints these can be solved. ¿From a methodological standpoint, an important contribution was due to Rogers [18], who succeeded to express the steady-state buffer content distribution in terms of the fundamental Wiener-Hopf factorization. Another key paper is by Ahn and Ramaswami [2], who explicitly exploit relations with quasi-birth-death processes. We also mention that a nice (recent) literature overview on Markov fluid queues is given in, e.g., [5].

In the second half of the 1990s models emerged in which the source behavior was influenced by the buffer content; see for instance [1, 4, 15, 17]; in [9, 19], Q and R depend *continuously* on the buffer level. Importantly, in all these feedback fluid models the buffer content uniquely defines the probabilistic properties of the source. The model analyzed in the present paper departs from this property. In fact the input process has two 'modes'. The first mode applies as long as the upper threshold B_1 has not been reached from below. As soon as that happens, we switch to the second mode, until the lower threshold B_2, smaller than B_1, is hit from above, i.e., the buffer occupancy falls below B_2. In this way the threshold B_1 is used to signal the onset of congestion, and B_2 to signal the end of congestion. Important novelty of our model as compared to earlier feedback models is that the input process may behave in two different ways between the two thresholds, depending on the past.

In this paper our goal is to find the steady-state buffer content distribution and associated performance measures of the model with two thresholds, as was described above. The methodology used involves the derivation of Kolmogorov equations and balance equations, which result in a solution in

terms of a spectral expansion; then additional conditions are identified that solve for the unknowns in the solution. Although this method is in principle similar to the approach in, e.g., [15], our model poses new challenges, due to its specific features (i.e., the two thresholds, the two modes of the input process). In the analysis particularly the behavior at the thresholds should be handled with care. As a result, the derivation of the conditions to solve for the unknowns turns out to be substantially more difficult than in the model of [15].

We mentioned above that the model we investigate can be useful in detecting congestion and can be generically applied to any congestion control mechanism for packet networks such as [8], [7] and [13]. Here we explain how this kind of feedback control has special significance with respect to congestion control in Ethernet metropolitan networks. The back-pressure scheme defined in IEEE 802.3x [10], is intended to provide flow control on a hop-by-hop basis by allowing ports to *turn off* their upstream link neighbors for a period of time. For a full-duplex connection, this mechanism is based on a special frame called *pause frame* in which the pause period is specified. The end-station (or router) receiving the pause frame looks at the pause period, and does not transmit or attempt transmission for that amount of time. Alternatively, an ON/OFF pause message can be sent signaling the beginning and end of the transmission pause phase. This congestion control method is usually implemented by using a high and a low threshold in the (congested) queue. When the queue occupancy exceeds the high threshold the *PauseOn* message is sent and when the queue occupancy drops below the low threshold the *PauseOff* message is sent and consequently transmission is resumed. Previous works [12, 13, 16] on the Ethernet congestion control have concentrated on the throughput gain which can be achieved by using the scheme. We are not aware of any literature with an analytically tractable model of the mechanism. The model presented in this paper can be used to optimally configure the high and low thresholds and decide on when to initiate the transmission pause phase and when to end it, consequently, addressing an essential design criterion for the Ethernet congestion avoidance scheme.

The rest of the paper is organized as follows. In Section 2 we describe the model with two thresholds in detail. In Section 3 we analyze the model and derive the equilibrium distribution of the buffer content. This section consists of three parts. Section 3.1 gives the balance equations for the buffer occupancy. Section 3.2 uses the spectral expansion method to provide the solution to the buffer occupancy in a compact form, which involves several unknown coefficients. In Section 3.3 we derive as many constraints as there are unknowns, so that these coefficients can be identified. In Section 4 we demonstrate our analysis by considering a numerical example and graphically present the buffer content distribution; we remark that a full numerical assessment of the back-pressure mechanism, relying on the methods developed in this paper, is found in [14]. We do include here, however, numerical evidence for the claim that the two-threshold mechanism leads to a reduction of the signaling overhead. Finally in Section 5 we conclude the paper with a summary of our results, and a discussion on future work.

2   Model

In this section we provide the formal definition of the model. We consider a fluid queue with an infinite buffer and constant output rate c. Let $W(t)$ be the content of the queue at time t, which is a stochastic process due to the probabilistic way in which fluid enters the queue. A popular model for such an input process is a so-called *Markov fluid source*. This model prescribes that the rate at which fluid enters the queue per unit time depends on the current state of some background irreducible continuous-time Markov chain $X(t)$, defined on a finite state-space $\{1, \ldots, d\}$, for $d \in \mathbb{N}$. At times when $X(t) = i$, the current input rate is $r_i \geq 0$. When we let the corresponding generator matrix be $Q \equiv (q_{ij})_{i,j=1}^d$, with $\sum_{j=1}^d q_{ij} = 0$ and $q_{ij} \geq 0$ for $i \neq j$, and define the d-dimensional traffic rate vector $\boldsymbol{r} \equiv (r_1, \ldots, r_d)^{\mathrm{T}}$, we call this input process a Markov-fluid source with parameters Q and $\boldsymbol{r}$.

In our feedback fluid model, the input stream alternates between two modes. In one mode the input process behaves like a Markov fluid source with generator Q^+ (dimension $d \times d$) and traffic rate vector $\boldsymbol{r}^+$ (dimension d). Similarly, in the other mode it behaves like a Markov fluid source with generator Q^- (also dimension $d \times d$) and traffic rate vector $\boldsymbol{r}^-$ (also dimension d).

We introduce the indicator variable process $I(\cdot)$, taking values in $\{'+', '-'\}$, which gives the current mode of operation of the input source. It is important to note that whenever $I(t)$ switches from one mode to another, the background process $X(t)$ stays in the same state; only its dynamics will from that time onwards behave according to the other generator matrix. However, the rate at which the fluid buffer receives fluid *does* change instantaneously from r_i^+ to r_i^- (or vice versa), when the background process $X(t)$ is in state i at the switching instant. Which of the two modes is currently valid at some time t depends on the behavior of the content process $W(t)$ relative to two thresholds, an upper threshold B_1 and a lower threshold B_2. The first mode ('+') applies as long as $W(t)$ has not reached the upper threshold B_1 from below. As soon as that happens, $I(t)$ switches to the other mode ('−'), until $W(t)$ hits the lower threshold B_2 from above, etc.

Let us describe this in some more detail, see also Figure 1. Suppose $W(0) = 0$, i.e., the process starts with an empty buffer, and let the indicator process $I(t)$ start in '+', where it will stay as long as the process $W(t)$ has not reached B_1 from below. During this period the input process behaves as a Markov-fluid source with d-dimensional generator Q^+ and traffic rate vector $\boldsymbol{r}^+$, and the buffer is drained at a rate c. At some point the buffer content $W(t)$ reaches the upper threshold B_1. Suppose that $X(t)$ is then in state j. Then $I(t)$ switches to '−', while the background process $X(t)$ stays in state j. From then on the input process behaves as a Markov-fluid source with generator Q^- and traffic rate vector $\boldsymbol{r}^-$, while the buffer is still drained at rate c. In particular, the current flow rate will change from r_j^+, which is larger than c, to r_j^-, which may or may not be larger than c. Further on at some moment the buffer content $W(t)$ drops to the lower threshold B_2. Suppose that $X(t)$ is in state k at this moment, then $I(t)$ switches to '+' while the background process stays in state k, and the input rate changes from $r_k^- < c$ to r_k^+.

Thus, the process continues, and will converge to an equilibrium distribution, assuming the queue is stable. With π^- denoting the equilibrium distribution corresponding to Q^-, i.e., $\pi^- Q^- = 0$ and
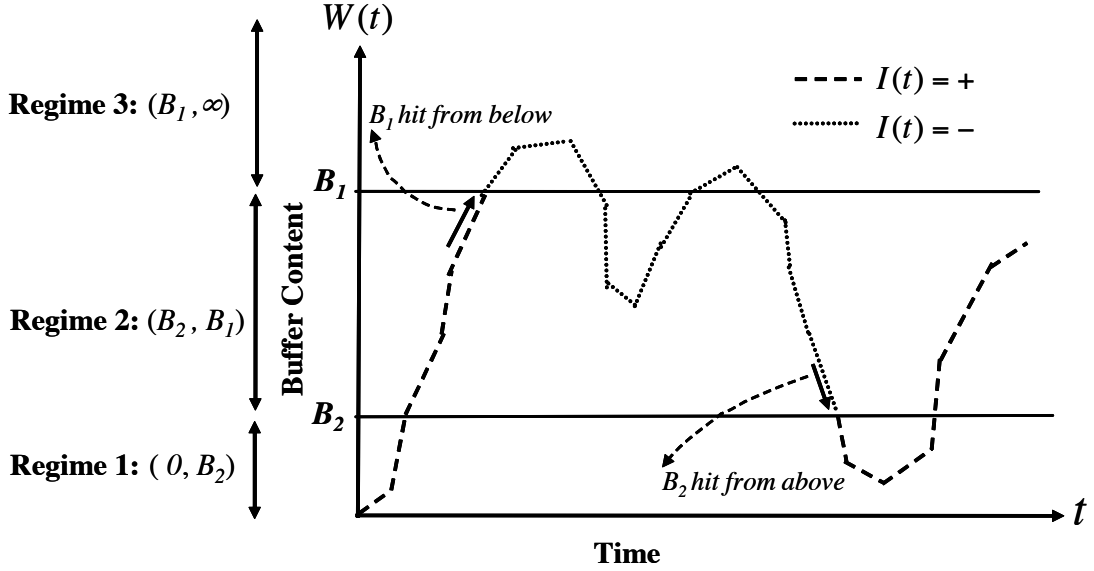
Figure 1: Different regimes for the buffer content $W(t)$

$\sum_{i=1}^{d} \pi_i^- = 1$, the equilibrium condition is

$$\sum_{i=1}^{d} \pi_i^- r_i^- < c;$$

throughout this paper we assume that this condition is satisfied. For technical reasons, we will also assume that for all states $i = 1, \ldots, d$ we have $r_i^+ \neq c$ and $r_i^- \neq c$, so that the content of the queue is never constant over time (unless it is zero). Let W be the steady-state buffer content, i.e., a random variable distributed as $\lim_{t \to \infty} W(t)$ and define I and X similarly. Define also, for $i = 1, \ldots, d$ and $x \geq 0$,

$$F_i^-(x) := \mathbb{P}(I = -, X = i, W \leq x), \quad F_i^+(x) := \mathbb{P}(I = +, X = i, W \leq x).$$

The goal of our paper is to identify $F_i^-(x)$ and $F_i^+(x)$. Having solved for these distribution functions, we can calculate two important performance measures for the system, namely the throughput ϑ and the distribution of the delay D, as follows:

$$\vartheta = \sum_{i=1}^{d} \left(r_i^+ F_i^+(\infty) + r_i^- F_i^-(\infty) \right), \tag{1}$$

and for $d \geq 0$,

$$\mathbb{P}(D \leq d) = \left(\sum_{i=1}^{d} \left(r_i^+ F_i^+(dc) + r_i^- F_i^-(dc) \right) \right) \bigg/ \left(\sum_{i=1}^{d} \left(r_i^+ F_i^+(\infty) + r_i^- F_i^-(\infty) \right) \right).$$

We define some additional notation which will be helpful in considering the various cases while solving for $F_i^-(x)$ and $F_i^+(x)$. We define the sets of 'up-states' and 'down-states' for both modes, and

5

their cardinalities, as follows:

$$S_{\mathrm{D}}^- := \{i : r_i^- < c\} \quad \text{and} \quad d_{\mathrm{D}}^- := \#\{S_{\mathrm{D}}^-\};$$
$$S_{\mathrm{U}}^- := \{i : r_i^- > c\} \quad \text{and} \quad d_{\mathrm{U}}^- := \#\{S_{\mathrm{U}}^-\};$$
$$S_{\mathrm{D}}^+ := \{i : r_i^+ < c\} \quad \text{and} \quad d_{\mathrm{D}}^+ := \#\{S_{\mathrm{D}}^+\};$$
$$S_{\mathrm{U}}^+ := \{i : r_i^+ > c\} \quad \text{and} \quad d_{\mathrm{U}}^+ := \#\{S_{\mathrm{U}}^+\}.$$

The subscript D is a mnemonic for 'Down', referring to the buffer being drained, while U stands for 'Up', referring to the buffer filling up. Evidently, $d_{\mathrm{D}}^- + d_{\mathrm{U}}^- = d_{\mathrm{D}}^+ + d_{\mathrm{U}}^+ = d$.

3 Analysis

In this section we analyze the buffer content distribution, by presenting a procedure to compute $F_i^-(x)$ and $F_i^+(x)$, for $i = 1, \ldots, d$. From the model description in the previous section we know that $F_i^-(x)$ and $F_i^+(x)$ have different characteristics in different intervals of the buffer content. Therefore, we define Regimes 1, 2, and 3, as shown in Figure 1. For each of these regimes we analyze $F_i^-(x)$ and $F_i^+(x)$. Two cases are rather straightforward, and therefore we start with these.

3^+: $F_i^+(x)$ *in Regime* 3. When the buffer occupancy reaches B_1 the indicator switches to the '$-$' state (if it was not in '$-$' already). The indicator changes to '$+$', only after the buffer occupancy would drop below B_2, where $B_2 < B_1$. Therefore, $F_i^+(x)$ is constant in the interval $[B_1, \infty)$. For $i = 1, ..., d$ and $x \geq B_1$,

$$F_i^+(x) = F_i^+(B_1). \tag{2}$$

1^-: $F_i^-(x)$ *in Regime* 1. If the buffer occupancy drops below B_2, then the indicator switches to '$+$'. Therefore, the fluid source with the '$-$' indicator can never be active below B_2 but only in the interval $[B_2, \infty)$. We have, for all $i = 1, ..., d$ and $x \leq B_2$,

$$F_i^-(x) = 0. \tag{3}$$

Nevertheless, it is important to note that even though $F_i^-(B_2) = 0$ the density at B_2, i.e.,

$$f_i^-(B_2) := \left. \frac{\mathrm{d}F_i^-(x)}{\mathrm{d}x} \right|_{x \downarrow B_2},$$

might not be equal to zero.

The other cases 1^+, 2^+, 2^- and 3^- are analyzed in the following subsections. We follow an approach similar to that in [3, 15] to find the complete solution to $F_i^-(x)$ and $F_i^+(x)$. We derive the balance equations for both $F_i^-(x)$ and $F_i^+(x)$ in Section 3.1, by first considering the Kolmogorov forward equations. What makes the Kolmogorov equations especially complicated in our case are the transitions around the thresholds B_1 and B_2. Assuming that a transition takes place somewhere in a small

time interval, the exact time of the transition is unknown, and as a consequence so is the indicator of the generator matrix in that interval. Section 3.1 deals with these issues. In Section 3.2 the spectral expansion method is used to find the solution to the differential equations. These can be written down in a rather simple form, but involve an extensive set of unknown coefficients. In Section 3.3 we find as many conditions as the number of unknowns, so that the stationary distribution of the buffer content can be determined.

3.1 Derivation of the balance equations

We found above simple solutions for $F_i^+(x)$ in Regime 3 and $F_i^-(x)$ in Regime 1, given by Eqns. (2) and (3). Our goal in this subsection is to derive the Kolmogorov forward equations for the other cases, from which we then easily obtain the corresponding balance equations. We slightly abuse notation by also using F_i^- and F_i^+ for the time-dependent distribution functions, i.e., we define $F_i^-(t, x) := \mathbb{P}(I(t) = -, X(t) = i, W(t) \leq x)$ and $F_i^+(t, x)$ analogously.

1^+: $F_i^+(x)$ *in Regime* 1. Regime 1 refers to $0 < x < B_2$, where we have for small h

$$F_i^+(t + h, x) = \left(1 - h \sum_{j \neq i} q_{i,j}^+\right) F_i^+\left(t, x - h(r_i^+ - c)\right) + h \sum_{j \neq i} q_{j,i}^+ F_j^+(t, x) + o(h).$$

Rearranging, dividing by h, and using the fact that the rows of the Q^+ matrix add up to zero, we obtain

$$\frac{F_i^+(t + h, x) - F_i^+(t, x - h(r_i^+ - c))}{h} = q_{i,i}^+ F_i^+(t, x - h(r_i^+ - c)) + \sum_{j \neq i} q_{j,i}^+ F_j^+(t, x) + \frac{o(h)}{h}.$$

By taking $h \downarrow 0$, we find

$$\frac{\partial}{\partial t} F_i^+(t, x) + (r_i^+ - c) \frac{\partial}{\partial x} F_i^+(t, x) = \sum_j q_{j,i}^+ F_j^+(t, x).$$

(Remark that, formally, these partial derivatives are not necessarily well-defined. As we are interested in the stationary behavior of the queue, this fact does not play a role – in fact we can assume the queue content has a proper density at time 0.) Assuming stationarity we set $F_i^+(t, x) = F_i^+(x)$ and in addition we set all derivatives with respect to t equal to 0. We thus obtain

$$(r_i^+ - c) \frac{\mathrm{d}}{\mathrm{d}x} F_i^+(x) = \sum_j q_{j,i}^+ F_j^+(x). \tag{4}$$

2^+: $F_i^+(x)$ *in Regime* 2. We now consider the interval $B_2 < x < B_1$. For i in S_{U}^-, we simply have the same equations as in Regime 1, leading to Eqn. (4). However, when i in S_{D}^-, we have to include the possibility that a transition can occur from the '$-$' state into the '$+$' state. This will happen when the buffer content at time t is between B_2 and $B_2 - h(r_i^- - c)$ (which is just above B_2 due to

$i \in S_D^-$), and the background process does not change its state during $(t, t + h]$. A complication here is that before the transition, Q^- is active and after the transition Q^+ is active. However, we do know that the probability that $X(t)$ remains in i during during $(t, t + h]$ is $1 + o(1)$, no matter what the precise form is, and this knowledge is sufficient. We thus find, for i in S_D^-,

$$
\begin{aligned}
F_i^+(t + h, x) = {} & \left(1 - h \sum_{j \neq i} q_{i,j}^+\right) F_i^+\left(t, x - h(r_i^+ - c)\right) + h \sum_{j \neq i} q_{j,i}^+ F_j^+(t, x) \\
& + (1 + o(1)) \left(F_i^-(t, B_2 - h(r_i^- - c)) - F_i^-(t, B_2)\right) + o(h).
\end{aligned} \tag{5}
$$

By rearranging and dividing by h on both sides we obtain

$$
\begin{aligned}
\frac{F_i^+(t + h, x) - F_i^+(t, x - h(r_i^+ - c))}{h} = {} & q_{i,i}^+ F_i^+(t, x - h(r_i^+ - c)) + \sum_{j \neq i} q_{j,i}^+ F_j^+(t, x) \\
& + (1 + o(1)) \frac{(F_i^-(t, B_2 - h(r_i^- - c)) - F_i^-(t, B_2))}{h} + \frac{o(h)}{h}.
\end{aligned}
$$

Then taking $h \downarrow 0$, and assuming stationarity we find

$$
(r_i^+ - c) \frac{\mathrm{d}}{\mathrm{d}x} F_i^+(x) = \sum_j q_{j,i}^+ F_j^+(x) - (r_i^- - c) \left. \frac{\mathrm{d}}{\mathrm{d}x} F_i^-(x) \right|_{x \downarrow B_2}. \tag{6}
$$

Since for i in S_U^-, we already found the same equations as in Regime 1, leading to Eqn. (4), we can combine the two cases to find, for $i = 1, \dots, d$,

$$
(r_i^+ - c) \frac{\mathrm{d}}{\mathrm{d}x} F_i^+(x) = \sum_j q_{j,i}^+ F_j^+(x) - A_i^-, \tag{7}
$$

where

$$
A_i^- := \begin{cases} (r_i^- - c) \left. \dfrac{\mathrm{d}}{\mathrm{d}x} F_i^-(x) \right|_{x \downarrow B_2} & \text{for } i \text{ in } S_D^-; \\ 0 & \text{for } i \text{ in } S_U^-. \end{cases} \tag{8}
$$

2^-: $F_i^-(x)$ *in Regime* 2. In Regime 2, i.e., $B_2 < x < B_1$, for i in S_U^-, we have the simple case

$$
F_i^-(t + h, x) = \left(1 - h \sum_{j \neq i} q_{i,j}^-\right) F_i^-\left(t, x - h(r_i^- - c)\right) + h \sum_{j \neq i} q_{j,i}^- F_j^-(t, x) + o(h).
$$

By rearranging, dividing by h, taking $h \downarrow 0$ and assuming stationarity we obtain

$$
(r_i^- - c) \frac{\mathrm{d}}{\mathrm{d}x} F_i^-(x) = \sum_j q_{j,i}^- F_j^-(x). \tag{9}
$$

For $i \in S_D^-$, the equation is more complicated. If we consider a time interval of h time units, then in this interval the buffer occupancy will drop by $|h(r_i^- - c)|$. We have to make sure that it does

not drop to or below B_2. If this occurs then the indicator switches to the '+' state, which is the probability we want to subtract from the equations (as it was already taken into account in (6)). Therefore, we include the term $-F_i^-(t, B_2 - h(r_i^- - c))$ which ensures that after h time units the buffer occupancy cannot drop to or below B_2. We thus obtain

$$F_i^-(t + h, x) = \left(1 - h \sum_{j \neq i} q_{i,j}^-\right) \left(F_i^-(t, x - h(r_i^- - c)) - F_i^-(t, B_2 - h(r_i^- - c))\right)$$
$$+ h \sum_{j \neq i} q_{j,i}^- F_j^-(t, x) + o(h).$$

By rearranging, taking $h \downarrow 0$ and assuming stationarity we get

$$(r_i^- - c)\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(t, x) = \sum_j q_{j,i}^- F_j^-(t, x) + (r_i^- - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(t, x)\right|_{x \downarrow B_2}. \tag{10}$$

We can now combine Eqns. (9) and (10) into one equation for i as

$$(r_i^- - c)\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x) = \sum_j q_{j,i}^- F_j^-(x) + A_i^- \tag{11}$$

where A_i^- is given by Eqn. (8).

3^-: $F_i^-(x)$ *in Regime* 3. The equations for $B_1 < x < \infty$, are the most complicated. This is because we have to take into account two aspects. Firstly, we have to exclude the possibility of a transition from the '−' into the '+' state when $X(t)$ is in S_D^- and the buffer content is just above B_2 at time t. It is clear from the explanation for $F_i^-(x)$ in Regime 2 that this is done by including a term $-F_i^-(t, B_2 - h(r_i^- - c))$. Secondly, we have to include the possibility of a transition from the '+' state into the '−' state. If at time t, the buffer content is somewhere in the interval $[B_1, B_1 - h(r_i^+ - c)]$, and the background state is in S_U^+, then in another h time units, the buffer content will increase and will definitely reach B_1 and jump into the '−' state. We therefore add a term $(1 + o(1))\left(F_i^+(t, B_1) - F_i^+(t, B_1 - h(r_i^+ - c))\right)$, similar to the term $(1 + o(1))\left(F_i^-(t, B_2 - h(r_i^- - c)) - F_i^-(t, B_2)\right)$ we added in the equation for F_i^+ in Regime 2. We find, for i in $S_D^- \cap S_U^+$,

$$F_i^-(t + h, x) = \left(1 - h \sum_{j \neq i} q_{i,j}^-\right) \left(F_i^-(t, x - h(r_i^- - c)) - F_i^-(t, B_2 - h(r_i^- - c))\right)$$
$$+ (1 + o(1))\left(F_i^+(t, B_1) - F_i^+(t, B_1 - h(r_i^+ - c))\right) + h \sum_{j \neq i} q_{j,i}^- F_j^-(t, x) + o(h). \tag{12}$$

By rearranging, dividing by h, taking $h \downarrow 0$, and assuming stationarity, we obtain

$$(r_i^- - c)\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x) = \sum_j q_{j,i}^- F_j^-(x) + (r_i^- - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x)\right|_{x \downarrow B_2} \tag{13}$$
$$+ (r_i^+ - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^+(x)\right|_{x \uparrow B_1}$$

9

Regime	Interval	$\boldsymbol{F}^+(x)$	$\boldsymbol{F}^-(x)$
1	$(0, B_2)$	$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{F}^+(x)(R^+ - C) = \boldsymbol{F}^+(x)Q^+$	0
2	(B_2, B_1)	$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{F}^+(x)(R^+ - C) = \boldsymbol{F}^+(x)Q^+ - \boldsymbol{A}^-$	$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{F}^-(x)(R^- - C) = \boldsymbol{F}^-(x)Q^- + \boldsymbol{A}^-$
3	(B_1, ∞)	$\boldsymbol{F}^+(B_1)$	$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{F}^-(x)(R^- - C) = \boldsymbol{F}^-(x)Q^- + \boldsymbol{A}^- + \boldsymbol{A}^+$

Table 1: *Overview of balance equations in the different regimes*

In order to derive the equations for the other values of i we should note that in Eqn. (12) the term $F_i^-(t, B_2 - h(r_i^- - c))$ appears for all $i \in S_D^-$ and the term $(1 + o(1))\left(F_i^+(t, B_1) - F_i^+(t, B_1 - h(r_i^+ - c))\right)$ for $i \in S_U^+$. Further on in (13) the term $F_i^-(t, B_2 - h(r_i^- - c))$ results in

$$(r_i^- - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x)\right|_{x \downarrow B_2},$$

whereas $(1 + o(1))\left(F_i^+(t, B_1) - F_i^+(t, B_1 - h(r_i^+ - c))\right)$ leads to

$$(r_i^+ - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^+(x)\right|_{x \uparrow B_1}.$$

Therefore, we obtain, for any $i = 1, \ldots, d$,

$$(r_i^- - c)\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x) = \sum_j q_{j,i}^- F_j^-(x) + A_i^- + A_i^+, \tag{14}$$

where A_i^- is given by Eqn. (8) and A_i^+ is

$$A_i^+ := \begin{cases} (r_i^+ - c)\left.\dfrac{\mathrm{d}}{\mathrm{d}x}F_i^+(x)\right|_{x \uparrow B_1} & \text{for } i \in S_U^+; \\ 0 & \text{for } i \in S_D^+. \end{cases} \tag{15}$$

In order to get an overview of the balance equations in the different regimes we have summarized the results so far in matrix form in Table 1, where $\boldsymbol{F}^+(x) \equiv (F_1^+(x), \ldots, F_d^+(x))$; the row vectors $\boldsymbol{F}^-(x)$, $\boldsymbol{A}^-$ and $\boldsymbol{A}^+$ are defined similarly. R^+ is the diagonal matrix $\mathrm{diag}\{r_1^+, \ldots, r_d^+\}$ and R^- the diagonal matrix $\mathrm{diag}\{r_1^-, \ldots, r_d^-\}$. We also introduce $C := cI_d$, where I_d is the identity matrix of dimension d.

3.2 Solution to the balance equations

In the previous subsection we have derived the balance equations for both $\boldsymbol{F}^-(x)$ and $\boldsymbol{F}^+(x)$. In this subsection we provide the solutions to these equations, using the spectral expansion method used in [3] and [15]. The solution can then be presented in a simple form, but it involves a number of unknown coefficients.

1^+ : The balance equation for $\boldsymbol{F}^+(x)$ in Regime 1 is $\mathrm{d}\boldsymbol{F}^+(x)/\mathrm{d}x \cdot (R^+ - C) = \boldsymbol{F}^+(x)Q^+$. The spectral expansion of the solution to this equation is given by

$$\boldsymbol{F}^+(x) = \sum_{j=1}^{d} a_j^{1+} \boldsymbol{v}_j^+ \exp[z_j^+ x]$$

where $(z_j^+, \boldsymbol{v}_j^+)$ is an eigenvalue-eigenvector pair statisfying $z_j^+ \boldsymbol{v}_j^+(R^+ - C) = \boldsymbol{v}_j^+ Q^+$, and the a_j^{1+} are coefficients.

In the above solution we tacitly assumed that the matrix $Q^+(R^+ - C)^{-1}$ has full eigenspace, in that all eigenvalues are simple (i.e., have multiplicity 1). Two remarks are in place here. (A) In the first place, we mention that it is known that if the Q^+ matrix has a specific structure, the eigenvalues are indeed simple (and real); most notably, as shown in [6], if Q^+ corresponds to a birth-death Markov process this property indeed applies. (B) Secondly, eigenvalues with multiplicity k larger than 1 do not lead to any conceptual problems. Standard theory on linear differential equations entails that then the density of the stationary queue content has terms proportional to $x^j \exp(-z_j^+ x)$, with $j = 0, \ldots, k - 1$. We decided to assume in our analysis that the eigenvalues of $Q^+(R^+ - C)^{-1}$ (and later also those of $Q^-(R^- - C)^{-1}$) are simple as the corresponding formulas for the 'non-simple case' do not add much extra insight, and are notationally cumbersome.

2^+ : We now consider $\boldsymbol{F}^+(x)$ in the interval $B_2 < x < B_1$ for which the balance equation is

$$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{F}^+(x)(R^+ - C) = \boldsymbol{F}^+(x)Q^+ - \boldsymbol{A}^-.$$

This equation has an inhomogeneous term because of which we cannot write the solution as for $\boldsymbol{F}^+(x)$ in Regime 1. Since A_i^- is a constant (see Eqn. (8)), differentiation of the above equation with respect to x gives us a homogeneous equation in $\boldsymbol{f}^+(x) \equiv \mathrm{d}\boldsymbol{F}^+(x)/\mathrm{d}x$ as below

$$\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{f}^+(x)(R^+ - C) = \boldsymbol{f}^+(x)Q^+.$$

Now that we have a homogeneous equation its solution is given as

$$\boldsymbol{f}^+(x) = \sum_{j=1}^{d} \tilde{a}_j^{2+} \boldsymbol{v}_j^+ \exp[z_j^+ x]$$

where $(z_j^+, \boldsymbol{v}_j^+)$ is the same eigenvalue-eigenvector pair as before and $\tilde{a}_j^{2+}$ are coefficients. As Q^+ is the generator, it has eigenvalue 0, and hence one of the eigenvalues z_j^+ is zero, say $z_{j*}^+ = 0$. With this in mind, integration immediately yields that

$$\boldsymbol{F}^+(x) = \sum_{j=1, j\neq j^*}^{d} a_j^{2+} \boldsymbol{v}_j^+ \exp[z_j^+ x] + a_{j*}^{2+} \boldsymbol{v}_{j*}^+ x + \boldsymbol{w}^{2+}$$

where $a_j^{2+} = \tilde{a}_j^{2+}/z_j^+$ for $j \neq j^*, a_{j*}^{2+} = \tilde{a}_{j*}^{2+}$, and the components w_i^{2+} of $\boldsymbol{w}^{2+}$ are integration constants.

11

Regime	Interval	$\boldsymbol{F}^+(x)$	$\boldsymbol{F}^-(x)$
1	$(0, B_2)$	$\sum_{j=1}^{d} a_j^{1+} \boldsymbol{v}_j^+ \exp[z_j^+ x]$	0
2	(B_2, B_1)	$\sum_{j=1, j\neq j^\star}^{d} a_j^{2+} \boldsymbol{v}_j^+ \exp[z_j^+ x] + a_{j^\star}^{2+} \boldsymbol{v}_{j^\star}^+ x + \boldsymbol{w}^{2+}$	$\sum_{j=1, j\neq j^\star}^{d} a_j^{2-} \boldsymbol{v}_j^- \exp[z_j^- x] + a_{j^\star}^{2-} \boldsymbol{v}_{j^\star}^- x + \boldsymbol{w}^{2-}$
3	(B_1, ∞)	$F^+(B_1)$	$\sum_{j=1, j\neq j^\star}^{d} a_j^{3-} \boldsymbol{v}_j^- \exp[z_j^- x] + a_{j^\star}^{3-} \boldsymbol{v}_{j^\star}^- x + \boldsymbol{w}^{3-}$

Table 2: *Overview of solution*

2^- : For $B_2 < x < B_1$, the balance equation for $\boldsymbol{F}^-(x)$ is $\mathrm{d}\boldsymbol{F}^-(x)/\mathrm{d}x \cdot (R^- - C) = \boldsymbol{F}^-(x)Q^- + \boldsymbol{A}^-$. This is again a inhomogeneous equation and the spectral method cannot be used directly. Therefore, we follow the same procedure as for $\boldsymbol{F}^+(x)$ in Regime 2. We differentiate the equation on both sides to get a homogeneous equation in $\boldsymbol{f}^-(x)$, which we integrate to obtain

$$\boldsymbol{F}^-(x) = \sum_{j=1, j\neq j^*}^{d} a_j^{2-} \boldsymbol{v}_j^- \exp[z_j^- x] + a_{j^*}^{2-} \boldsymbol{v}_{j^*}^- x + \boldsymbol{w}^{2-}$$

where $(z_j^-, \boldsymbol{v}_j^-)$ is an eigenvalue-eigenvector pair satisfying $z_j^- \boldsymbol{v}_j^- (R^- - C) = \boldsymbol{v}_j^- Q^-$, and the a_j^{2-} are coefficients. The components w_i^{2-} of $\boldsymbol{w}^{2-}$, are integration constants and the coefficient $a_{j^*}^{2-}$ corresponds to the eigenvalue $z_{j^*}^- = 0$ of Q^-.

3^- : The balance equation for $\boldsymbol{F}^-(x)$ in Regime 3 is $\mathrm{d}\boldsymbol{F}^-(x)/\mathrm{d}x \cdot (R^- - C) = \boldsymbol{F}^-(x)Q^- + \boldsymbol{A}^- + \boldsymbol{A}^+$. In this equation we have two inhomogeneous terms as opposed to one in the previous cases. Nevertheless, we can still apply the same method as for $\boldsymbol{F}^+(x)$ in Regime 2 and $\boldsymbol{F}^-(x)$ in Regime 2. This is because both the inhomogeneous terms in the equation above consist of constant elements $\lim_{x \downarrow B_2} \mathrm{d}F_i^-(x)/\mathrm{d}x$ and $\lim_{x \uparrow B_1} \mathrm{d}F_i^+(x)/\mathrm{d}x$ or zero (see Eqns. (8) and (15)) which disappear after differentiating with respect to x. Therefore, after differentiation and then integration we get the solution for $\boldsymbol{F}^-(x)$ in Regime 3 as

$$\boldsymbol{F}^-(x) = \sum_{j=1, j\neq j^\star}^{d} a_j^{3-} \boldsymbol{v}_j^- \exp[z_j^- x] + a_{j^\star}^{3-} \boldsymbol{v}_{j^\star}^- x + \boldsymbol{w}^{3-}$$

where $(z_j^-, \boldsymbol{v}_j^-)$ is again the eigenvalue-eigenvector pair that satisfies $z_j^- \boldsymbol{v}_j^- (R^- - C) = \boldsymbol{v}_j^- Q^-$, and the a_j^{3-} are coefficients. The components w_i^{3-} of $\boldsymbol{w}^{3-}$ are integration constants and the coefficient $a_{j^\star}^{3-}$ corresponds to $z_{j^\star}^- = 0$.

We summarize the solutions found in the various intervals in Table 2.

3.3 Derivation of conditions for finding the unknowns in the solution

In Section 3.2, we provided the solution to $\boldsymbol{F}^-(x)$ and $\boldsymbol{F}^+(x)$ using the spectral expansion method. However, the solution includes many unknowns which need to be found with additional conditions. Table 2 presents an overview of the solution where the vectors $\boldsymbol{a}^{1+}, \boldsymbol{a}^{2+}, \boldsymbol{a}^{2-}, \boldsymbol{a}^{3-}, \boldsymbol{w}^{2+}, \boldsymbol{w}^{2-}$ and $\boldsymbol{w}^{3-}$ are unknown. Table 3 enumerates all unknowns giving a total of $8d$. In this section our goal is to find $8d$ conditions so as to solve the system.

A. *Boundary conditions* at $x = 0$ and $x = \infty$ are as in [3] and [11]:

- $F_i^+(0) = 0$, for i in S_{U}^+. This is because it cannot be that simultaneously the buffer is empty and the background process is in an up-state. This gives us d_{U}^+ conditions.

- For $x \to \infty$, $F_i^-(x)$ should remain bounded, and therefore for all z_j^- with a non-negative real part, the corresponding a_j^{3-} has to be zero. Notice that this also entails that the equilibrium distribution of $W^-(t)$ is given by $\boldsymbol{w}^{3-}$. This gives us d_{D}^- conditions.

B. *Continuity conditions.* $F_i^+(x)$ and $F_i^-(x)$ are both continuous at the thresholds B_1 and B_2. This gives us the following $4d$ equations:

- $\lim_{x \uparrow B_2} F_i^+(x) = \lim_{x \downarrow B_2} F_i^+(x)$.
- $\lim_{x \uparrow B_1} F_i^+(x) = F_i^+(B_1)$.
- $\lim_{x \downarrow B_2} F_i^-(x) = 0$.
- $\lim_{x \downarrow B_1} F_i^-(x) = \lim_{x \downarrow B_1} F_i^-(x)$.

C. *Substitution conditions.*

As in [15] we need to substitute the solutions given in Section 3.2 into the inhomogeneous balance equations of Section 3.1. We have 3 inhomogeneous systems. Potentially each of these can lead to d conditions. Therefore, in total we would get $3d$ equations from the substitution.

Boundary conditions, continuity conditions and substitution conditions together were sufficient to solve the model in [15], but as we have a more complicated model in which the fluid source alternates between two modes, this is not the case here. Therefore, we introduce and prove the following:

Regime	Interval	$\boldsymbol{F}^+(x)$	$\boldsymbol{F}^-(x)$
1	$(0, B_2)$	$a_j^{1+} \quad j = 1, ..., d$	0
2	(B_2, B_1)	$a_j^{2+}, w_j^{2+} \quad j = 1, ..., d$	$a_j^{2-}, w_j^{2-} \quad j = 1, ..., d$
3	(B_1, ∞)	$F_j^+(B_1)$, for $j = 1, ..., d$	$a_j^{3-}, w_j^{3-} \quad j = 1, ..., d$
		Total number of unknowns: $8d$	

Table 3: *Overview of the unknowns in the different regimes*

D. *Additional conditions.*

In the first place, suppose that the buffer is filling up in the '+' state. At some point it will reach B_1 and then switch to the '−' process. Since there is no density beyond B_1 (and the phase being '+') it is highly unlikely that the buffer level is just below B_1 while the background process is in a down-state. Similarly, when the buffer is filling up ('−' phase) it is unlikely that the buffer content is just above B_2 while the background process is in an up-state.

The lemma below states these additional conditions more precisely, and is proved by deriving the balance equations at $x = B_1$ and $x = B_2$. Note that these were not addressed in Section 3.1.

Lemma 3.1. *(i) For all $i \in S_D^+$,*

$$\frac{d}{dx} F_i^+(x) \bigg|_{x \uparrow B_1} = 0.$$

(ii) For all $i \in S_U^-$

$$\frac{d}{dx} F_i^-(x) \bigg|_{x \downarrow B_2} = 0.$$

Proof: (i) We first consider the case for i in $S_D^- \cap S_U^+$. In this case we have

$$
\begin{aligned}
F_i^+(t+h, B_1) &= (1 - h \sum_{j \neq i} q_{i,j}^+) F_i^+(t, B_1 - h(r_i^+ - c)) + h \sum_{j \neq i} q_{j,i}^+ F_j^+(t, B_1) \\
&\quad + (1 + o(1)) \left(F_i^-(t, B_2 - h(r_i^- - c)) - F_i^-(t, B_2) \right) + o(h).
\end{aligned}
\tag{16}
$$

By rearranging, dividing by h, taking $h \downarrow 0$ and assuming stationarity we obtain for i in $S_D^- \cap S_U^+$,

$$(r_i^+ - c) \frac{d}{dx} F_i^+(x) \bigg|_{x \uparrow B_1} = \sum_j q_{j,i}^+(F_j^+(B_1)) - (r_i^- - c) \frac{d}{dx} F_i^-(x) \bigg|_{x \downarrow B_2}.$$

Comparison with (8) and (15) shows that

$$\sum_j q_{j,i}^+ F_j^+(B_1) = A_i^+ + A_i^-,
\tag{17}$$

and it is in fact easy to see that this holds for any i. Let us compare this to Eqn. (7) for $F_j^+(x)$ in the interval (B_2, B_1), letting $x \uparrow B_1$, which gives

$$\sum_j q_{j,i}^+ F_j^+(B_1) = (r_i^+ - c) \frac{d}{dx} F_i^+(x) \bigg|_{x \uparrow B_1} + A_i^-.
\tag{18}$$

From this comparison we conclude that for i in S_D^+

$$(r_i^+ - c) \frac{d}{dx} F_i^+(x) \bigg|_{x \uparrow B_1} = A_i^+ = 0.
\tag{19}$$

from which the first claim immediately follows.

(ii) The proof of this part is similar, comparing the two equations for $F_i^-(x)$ at B_2, which are $F_i^-(B_2) = 0$ and

$$(r_i^- - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x)\right|_{x\downarrow B_2} = \sum_j q_{j,i}^+ F_j^-(B_2) + A_i^-.$$

$\square$

The following lemma entails that the number of substitution conditions is really only $2d$ (not $3d$).

Lemma 3.2. *The equations from the substitution condition for $\boldsymbol{F}^-(x)$ in the interval (B_2, B_1) are implied by the continuity condition for $\boldsymbol{F}^-(x)$ at B_2.*

Proof: We start with the substitution for $F_j^-(x)$ in the interval (B_2, B_1). The balance equation in this case is, see (11),

$$(r_i^- - c)\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x) = \sum_k q_{k,i}^- F_k^-(x) + A_i^-.$$

On substituting the solution back into the equation and using the fact that $z_{j^\star}^- = 0$ and $\sum_k q_{k,i}^- v_{j,k}^- = z_j^- v_{j,i}^-(r_i^- - c)$, where $v_{j,i}^-$ refers to the ith component of vector $\boldsymbol{v}_j^-$, we find

$$(r_i^- - c)(a_{j^\star}^{2-} v_{j^\star,i}^-) = \sum_j q_{k,i}^- w_k^{2-} + A_i^-. \tag{20}$$

By the definition of A_i^- and using the first part of Lemma 3.1 we can simply substitute

$$A_i^- = (r_i^- - c)\left.\frac{\mathrm{d}}{\mathrm{d}x}F_i^-(x)\right|_{x\downarrow B_2} = (r_i^- - c)\left(\sum_{j=1, j\neq j^\star}^d a_j^{2-} v_{j,i}^- z_j^- \exp[z_j^- B_2] + a_{j^\star}^{2-} v_{j^\star,i}^-\right)$$

for all $i = 1, ..., d$ in the equation above to give us

$$\sum_k q_{k,i}^- w_k^{2-} + (r_i^- - c)\left(\sum_{j=1, j\neq j^\star}^d a_j^{2-} v_{j,i}^- z_j^- \exp[z_j^- B_2]\right) = 0, \qquad \text{or,}$$

$$\sum_k q_{k,i}^- w_k^{2-} + \sum_k q_{k,i}^-\left(\sum_{j=1, j\neq j^\star}^d a_j^{2-} v_{j,k}^- \exp[z_j^- B_2]\right) = 0$$

as the substitution conditions should hold. However, since we may freely add $\sum_k q_{k,i}^- a_{j^\star}^{2-} v_{j^\star,k}^- B_2$ to the left-hand side of the above (since it is zero), and then combine the sums into one, the conditions turn out to be equivalent to $\sum_k q_{k,i}^- F_k^-(B_2) = 0$, Hence they are indeed implied by the continuity equations at B_2, which say that $F_k^-(B_2) = 0$ for all k. $\square$

Let us now explain the intuition behind Lemma 3.2. If we look at Table 1, we could suspect that an overlap could arise for $F_i^-(x)$ in Regime 2. This is because this is the only inhomogeneous equation which involves a single indicator state (being the '$-$' state). The continuity equations would also have the same characteristics. As for the other inhomogeneous equations, these involve terms with both

15

the '+' and the '−' states, whereas the continuity equations still involve terms with a single indicator state. Hence, for these cases there is a clear difference between the characteristics of the substitution and the continuity equations, and the substitution equations do give additional information.

Let us now count the number of conditions we have at our disposal. The boundary conditions give us d_U^+ and d_D^- equations and the four continuity conditions give us another $4d$ conditions. The substitution conditions could have potentially given us another $3d$ conditions. Adding these to the $d_D^+ + d_U^-$ conditions from Lemma 3.1, we would together have $9d$ conditions, d more than we need. With Lemma 3.2 we proved that an overlap of d conditions exists between the continuity and the substitution conditions, eventually adding up to exactly $8d$ conditions equal to the total number of unknowns in the solution (see Table 3), which we need to solve the system.

However, it is important to note that all the $8d$ equations are linear in the different unknowns enlisted in Table 3, with a rank of $8d − 1$. This is easy to see since the linear system can be solved upto a multiplicative constant. The redundancy in the equations can be removed by replacing any one of the equations in the linear system by the normalization equation

$$\sum_{i=1}^{d} \left(F_i^+(\infty) + F_i^-(\infty) \right) = 1. \tag{21}$$

Thus, we arrive at the following result.

Proposition 3.3. *We have $8d$ conditions on the coefficients, matching the number of unknowns.*

In the next section we illustrate, by means of a numerical example, how one can identify the $8d$ unknowns.

4   Numerical example

In this section we provide numerical results aimed at demonstrating the computation of the stationary distribution of the buffer content and the other performance measures. We consider a two state numerical example, i.e., with $d = 2$. We consider the following generator matrices and rate vectors,

$$Q^+ = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}, \quad Q^- = \begin{pmatrix} -0.8 & 0.8 \\ 5 & -5 \end{pmatrix}, \quad \boldsymbol{r}^- = \begin{pmatrix} 16 \\ 0 \end{pmatrix}, \quad \boldsymbol{r}^+ = \begin{pmatrix} 25 \\ 0 \end{pmatrix}.$$

The other parameters are $c = 15$, $B_1 = 15$ and $B_2 = 10$. The diagonal matrices R^-, R^+, and C then equal $\text{diag}\{16, 0\}$, $\text{diag}\{25, 0\}$, and $\text{diag}\{15, 15\}$, respectively. After the numerical determination of the eigensystems of the matrices $Q^+(R^+ − C)^{-1}$ and $Q^-(R^- − C)^{-1}$ we apply the conditions as listed in Section 3.3. We then solve the resulting linear system of equations for the $8d = 16$ unknowns. This gives us the complete and unique solution for the stationary distribution of the buffer content, $\boldsymbol{F}^+(x)$ and $\boldsymbol{F}^-(x)$. The graphical representation of $F_i^+(x)$ and $F_i^-(x)$ for $i = 1, 2$ is shown in Figure 2. The total throughput of the system can be calculated from Eqn. (1) as

$$\vartheta = r_1^+ F_1^+(\infty) + r_1^- F_1^-(\infty) = 14.1924.$$
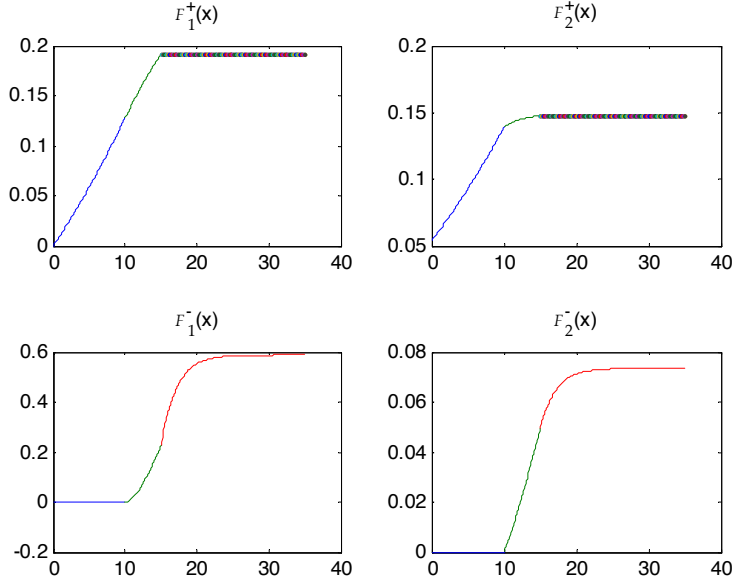
Figure 2: Probability distribution functions of the buffer content with $d = 2$, $B_1 = 15$, $B_2 = 10$.

We can compute the expected buffer content by first computing $F(x) = \sum_{i,j} F_i^j(x)$, then computing the combined probability density as $f(x) = \mathrm{d}F(x)/\mathrm{d}x$. The expected buffer content is then given by

$$\mathbb{E}W = \int_0^\infty x f(x)\mathrm{d}x = 12.2040.$$

In a second experiment we consider the effect of having two thresholds on the signaling overhead. The expected number of phase-transitions per unit time equals

$$\sum_{i \in S_{\mathrm{U}}^+} f_i^+(B_1)(r_i^+ - c) + \sum_{i \in S_{\mathrm{D}}^-} f_i^-(B_2)(c - r_i^-);$$

this (plausible) formula is derived in [14]. The effect of varying B_2, for a given value of B_1, is shown in Figure 3. The other parameters are as above. We observe that indeed the signaling frequency is reduced by choosing B_2 smaller than B_1.

5 Concluding remarks

We have analyzed a feedback fluid queueing system in which the traffic source rates adapt to congestion. The system has two thresholds: the higher threshold B_1 aims at signaling the beginning of a congestion period, whereas the lower threshold B_2 serves to signal the end of the congestion period. This idea is modelled by letting the input process alternate between two Markov fluid processes: the
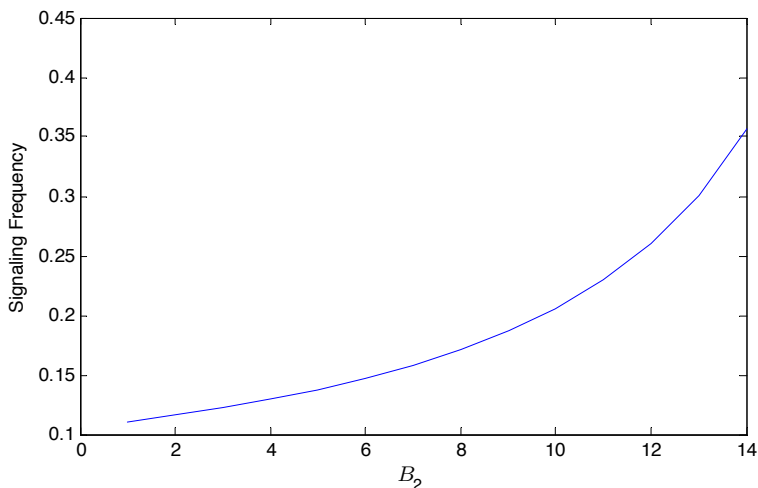
Figure 3: Signaling frequency as B_2 grows to $B_1 = 15$.

first applies as long as the upper threshold B_1 has not been hit from below. As soon as that happens, the traffic source switches to the other process, until B_2 is hit from above.

The resulting model falls in the class of Markovian feedback fluid queues. The numerical complexity of the methodology used to solve for the buffer content distribution and throughput boils down to solving two d-dimensional eigensystems, as well as a (fairly standard) linear system of $8d$ equations; here d denotes the cardinality of the state space on which the two Markov processes are defined.

A central design problem which can be investigated using our model, and is addressed in [14], is the optimization of the threshold positions while considering the trade-off between throughput and delay. Challenging extensions to the present analysis are the systematic assessment of the signaling frequency in the model (that is, what is the rate at which the input process alternates between the two Markov fluid processes), as a function of the model parameters; this is a relevant issue, as the signaling overhead needs to be controlled. Yet another important direction for future research is to incorporate delay in the reception of feedback signals and in the adaptation of the source traffic rate.

References

[1] I. ADAN, E. VAN DOORN, J. RESING, and W. SCHEINHARDT. 'Analysis of a single server queue interacting with a fluid reservoir. *Queueing Systems*, 29:313-336, 1998.

[2] S. AHN and V. RAMASWAMI. Steady state analysis of finite fluid flow models using finite QBDs. *Queueing Systems*, 49: 223-259, 2005.

[3] D. ANICK, D. MITRA, and M. SONDHI. Stochastic theory of a data-handling system with multiple sources. *Bell System Technical Journal*, 61:1871-1894, 1982.

[4] A. DA SILVA SOARES and G. LATOUCHE. A matrix-analytic approach to fluid queues with feed-back control. *International Journal of Simulation: Systems, Science & Technology*, 6: 4-12, 2005.

[5] A. DA SILVA SOARES and G. LATOUCHE. Matrix-analytic methods for fluid queues with finite buffers. *Performance Evaluation*, 63: 295-314, 2006.

[6] E. VAN DOORN, A. JAGERS, and J. DE WIT. A fluid reservoir regulated by a birth death-process. *Stochastic Models,* 4: 457 – 472, 1988.

[7] S. FLOYD. TCP and Explicit Congestion Notification. *ACM Computer Communication Review,* 24: 10-23, 1994.

[8] S. FLOYD and V. JACOBSON. Congestion gateways for packet neworks. *IEEE/ACM Transactions on Networking,* 1: 397-413, 1993.

[9] M. GRIBAUDO and M. TELEK. Stationary analysis of fluid level dependent bounded fluid models. *Performance Evaluation,* 65: 241-261, 2007.

[10] IEEE STANDARD 802.3. Carrier Sense Multiple Access with Collision Detection (CSMA/CD) access method and physical layer specification, Annex 31B, 1998 Edition.

[11] L. KOSTEN. Stochastic theory of a data handling systems with groups of multiple sources. In: *Performance of Computer Communication Systems*, eds. H. Rudin and W. Bux, Elsevier, Amsterdam, 321-331, 1984.

[12] R. MALHOTRA, R. VAN HAALEN, R. DE MAN, and M. VAN EVERDINGEN. Managing SLAs for metropolitan Ethernet networks. *Bell Labs Technical Journal*, 8: 83-95, 2002.

[13] R. MALHOTRA, R. VAN HAALEN, M. MANDJES, and R. NÚÑEZ-QUEIJA. Modeling the inter-action of IEEE 802.3x hop-by-hop flow control and TCP End-to-end Flow Control, *Proc. Next Generation Internet Networks*, 260-267, 2005.

[14] R. MALHOTRA, M. MANDJES, W. SCHEINHARDT, and J.L. VAN DEN BERG. Design issues of a back-pressure-based congestion control mechanism. *In preparation.*

[15] M. MANDJES, D. MITRA, and W. SCHEINHARDT. Models of network access sing feedback fluid queues. *Queueing Systems*, 44: 365-398, 2003.

[16] W. NOUREDDINE and F. TOBAGI. Selective back-pressure in switched Ethernet LANs. *Global Telecommunications Conference* 2, 1256-1263, 1999.

[17] K. RAMANAN and A. WEISS. Sharing bandwidth in ATM. *Proc. Allerton Conference*, 732-740, 1997.

[18] L. ROGERS. Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Annals of Applied Probability*, 4: 390–413, 1994.

[19] W. SCHEINHARDT, N. VAN FOREEST, and M. MANDJES. Continuous feedback fluid queues. *Operations Research Letters* 33: 551-559, 2005.