

INFINITE- AND FINITE-BUFFER MARKOV FLUID QUEUES: A UNIFIED ANALYSIS

NAIL AKAR,* *Bilkent University*

KHOSROW SOHRABY,** *University of Missouri–Kansas City*

Abstract

In this paper, we study Markov fluid queues where the net fluid rate to a single-buffer system varies with respect to the state of an underlying continuous-time Markov chain. We present a novel algorithmic approach to solve numerically for the steady-state solution of such queues. Using this approach, both infinite- and finite-buffer cases are studied. We show that the solution of the infinite-buffer case is reduced to the solution of a generalized spectral divide-and-conquer (SDC) problem applied on a certain matrix pencil. Moreover, this SDC problem does not require the individual computation of any eigenvalues and eigenvectors. Via the solution for the SDC problem, a matrix-exponential representation for the steady-state queue-length distribution is obtained. The finite-buffer case, on the other hand, requires a similar but different decomposition, the so-called additive decomposition (AD). Using the AD, we obtain a modified matrix-exponential representation for the steady-state queue-length distribution. The proposed approach for the finite-buffer case is shown not to have the numerical stability problems reported in the literature.

Keywords: Stochastic fluid model; Markov fluid queue; performance analysis; computer network; spectral divide-and-conquer problem; generalized Newton iteration

2000 Mathematics Subject Classification: Primary 60K25

Secondary 65C40; 65F15; 68M20; 47A15

1. Introduction

Queueing models in which some continuous flow, called the fluid, takes on the role of the traditional ‘customer’ have attracted a lot of attention in the computer networks performance research community [2], [12], [17]. In infinite- or finite-buffer Markov fluid queues, the queue-length cumulative distribution function (CDF) in the steady state obeys an ordinary differential-algebraic equation [2], [21] in which the boundary conditions are not known *a priori*. Traditionally, the eigenvalues and eigenvectors of the system are first computed, leading to a spectral expansion for the CDF. Then the coefficients of this expansion are computed by solving a linear equation using the constraints associated with the boundaries of the system. Anick *et al.* [2] studied the infinite-buffer case in which the arrival rate process is a superposition of a finite number of independent homogeneous continuous fluid sources. In their model, each source is governed by a two-state Markov chain and the service capacity is held constant. In [2], closed-form expressions are available for the eigenvalues and eigenvectors and the authors also provided useful asymptotic approximations for the survivor function. Kosten showed in [12] how to extend the results of [2] to the case of a superposition of a number of different groups,

Received 12 September 2002.

* Postal address: Electrical and Electronic Engineering Department, Bilkent University, 06800 Bilkent, Ankara, Turkey.

** Postal address: School of Interdisciplinary Computing and Engineering, University of Missouri–Kansas City, Kansas City, MO 64110, USA. Email address: sohrabyk@umkc.edu

each group comprised of a superposition of identical Markov fluid sources. Mitra [14] studied the case when the service rate is also modulated by a continuous-time Markov chain. These stochastic fluid flow models have since been extensively used for the performance analysis of packet voice and video multiplexing with or without background data traffic (see [17] for a collection of references). For the finite-buffer Markov fluid queues, Tucker [21] extended the results of [2] to the finite-buffer case in the context of performance analysis of a finite-capacity packet voice multiplexing system. However, the linear equations to solve for the coefficients of the spectral expansion are known to be ill conditioned, and numerical instability associated with large-scale finite queues are reported in [9]. Among the alternative techniques for solving Markov fluid queues, we note the probabilistic analysis given by Asmussen [3] using an iterative technique and the Wiener–Hopf factorization method of Rogers [16].

In this paper, we seek a numerically efficient and stable method for solving large-scale infinite- and finite-buffer Markov fluid queues. However, we do not impose a special structure on the underlying continuous-time Markov chain, i.e. the eigenvalues and eigenvectors cannot therefore be written in closed form. On the other hand, finding the eigenvalues and eigenvectors (in case a closed-form expression does not exist) of a nonsymmetric matrix is generally known to be ill conditioned and methods that rely on such computations cannot scale to large-scale queues.

In the infinite-buffer case, we propose an alternative method that relies on the solution of an associated generalized spectral divide-and-conquer (SDC) problem [5] without any need for computing eigenvalues and eigenvectors. Through the solution of the SDC, it is shown in the current paper that $q(\cdot)$, the queue-length survivor function, takes the matrix-exponential form

$$q(x) = fe^{Gx}h, \quad x \geq 0, \quad (1)$$

in the infinite-buffer case for a row vector f , a column vector h , and a stable square matrix G . Using this expression, we can easily calculate the associated moments using simple matrix–vector operations. The finite-buffer case, on the other hand, requires a similar but different decomposition, the so-called additive decomposition (AD) [11]. AD decomposes a transfer function into two subsystems, one being the forward (or stable) subsystem, and the other being the backward (antistable) subsystem. Using AD, the finite-buffer case with buffer size B leads to a modified matrix-exponential representation for the queue-length survivor function:

$$q(x) = a + f_1e^{G_1x}h_1 + f_2e^{-G_2(B-x)}h_2, \quad 0 \leq x < B, \quad (2)$$

for a scalar a , row vectors f_1 and f_2 , column vectors h_1 and h_2 , and two stable matrices G_1 and $-G_2$, the sizes of which are to be determined within the solution procedure. The first term in this expression is the constant term, the second and third terms are associated with the forward and backward subsystems of the finite fluid queue. The intent of this paper is to provide a numerically efficient and stable method for computing the factors of the representations (1) and (2).

The remainder of the paper is organized as follows. Section 2 is devoted to the notation and the preliminaries used in this paper. Section 3 addresses the Markov fluid queue model and summarizes the existing results. We present our solution for the infinite- and finite-buffer cases in Sections 4 and 5 respectively. Numerical results and conclusions are presented in Sections 6 and 7 respectively.

2. Notation

We use the following notation in this paper. We denote by e and I , respectively, a column vector of ones and an identity matrix of suitable sizes. When the size of the matrix is to be emphasized, we use the notation I_m to denote an $m \times m$ identity matrix. By $\text{rank}(A)$, $\mathcal{R}(A)$,

and $\mathcal{N}(A)$ we denote the rank, range space, and null space of a real-valued matrix A respectively. As far as the current paper is concerned, it is important to find orthonormal basis vectors for the range space of a rank deficient square matrix A of size n . A rank-revealing QR decomposition [6] of A is of the form $A\Pi = QR$, where Π is a permutation matrix, R is upper triangular, and Q is orthonormal. The permutation matrix is chosen so that the absolute values of the diagonal elements of R are decreasing. In case $n_1 = \text{rank}(A)$ is known *a priori*, the first n_1 columns of Q form an orthonormal basis for $\mathcal{R}(A)$. Sometimes it is also important to find orthonormal basis vectors for the null space of the matrix A . Then a rank revealing QR decomposition of A^\top is of the form $A^\top\Pi = QR$ in which the last $n_2 = n - n_1$ columns of Q form an orthonormal basis for $\mathcal{N}(A)$ [19]. We note that matrix algorithms related to finding orthonormal basis vectors for the range or null spaces of a matrix are readily available in linear algebra software packages like MATLAB[®] and LAPACK.

If $a = [a_1, a_2, \dots, a_M]$ is a vector of size M and $1 \leq n \leq m \leq M$, then $a(n : m)$ denotes the subvector $[a_n, a_{n+1}, \dots, a_m]$. We denote by \mathbb{C} the set of complex numbers and by $\text{Re}(z)$ the real part of a complex number z . By \mathbb{C}^- , \mathbb{C}^+ , and \mathbb{C}^0 we denote the open left half, open right half, and the imaginary axis respectively of the complex plane. We use the notation \mathbb{C}^{+0} to denote $\mathbb{C}^+ \cup \mathbb{C}^0$. Given a matrix pair (E, A) , where E and A are square real $n \times n$ matrices, the generalized nonsymmetric eigenvalue problem is to find the eigenvalues λ and corresponding eigenvectors ϕ_r such that $A\phi_r = \lambda E\phi_r$ and $\phi_r \neq 0$. The vector ϕ_r is then called a right eigenvector. A row vector ϕ_l is a left eigenvector of the ordered pair (E, A) if $\phi_l A = \lambda\phi_l E$ and $\phi_l \neq 0$. Equivalent notation used to refer to the generalized eigenproblem for the pair (E, A) is through the pencil $\lambda E - A$, where λ is an indeterminate. So we can also refer to the generalized eigenvalues and eigenvectors of the pencil $\lambda E - A$. The set of all generalized eigenvalues of the ordered pair (E, A) , or the pencil $\lambda E - A$, is denoted by $\lambda(E, A)$. When $\lambda(I, A) \subset \mathbb{C}^-$, A is called a stable matrix and A is called antistable if $\lambda(I, A) \subset \mathbb{C}^+$.

The following is based on [11]. Any subspace \mathcal{X} satisfying

$$\mathcal{Y} = E\mathcal{X} + A\mathcal{X} \quad \text{and} \quad \dim \mathcal{Y} = \dim \mathcal{X} \tag{3}$$

is called a deflating subspace of the pencil $\lambda E - A$. The name stems from the decomposition

$$Q^\top(\lambda E - A)Z = \begin{bmatrix} \lambda E_{11} - A_{11} & \lambda E_{12} - A_{12} \\ 0 & \lambda E_{22} - A_{22} \end{bmatrix}, \tag{4}$$

which is obtained from (3) by choosing orthonormal matrices Q and Z such that the first n_1 (the size of E_{11} and of A_{11}) columns of Z form a basis for the subspace \mathcal{X} and the first n_1 columns of Q form a basis for \mathcal{Y} . For obvious reasons, \mathcal{X} and \mathcal{Y} are sometimes called the right and left deflating subspaces respectively. Given real matrices E and A and a region \mathcal{D} of the complex plane, finding orthonormal matrices Q and Z such that (4) holds with $\lambda(E_{11}, A_{11}) \subset \mathcal{D}$ and $\lambda(E_{22}, A_{22}) \cap \mathcal{D} = \emptyset$ is called the generalized spectral divide-and-conquer (SDC) problem [5] (when $E = I$, it reduces to the ordinary SDC problem). In this case, the first n_1 columns of Z and Q are known to span, respectively, the right and left deflating subspaces of the matrix pencil $\lambda E - A$ associated with the region \mathcal{D} . There is also a unique deflating subspace \mathcal{X}_c of the pencil $\lambda E - A$ associated with the region $\mathbb{C} \setminus \mathcal{D}$. In [7], it was shown that \mathcal{X} and \mathcal{X}_c are complementary, i.e. they are linearly independent and add up to the whole space. The subspaces \mathcal{Y} and \mathcal{Y}_c are also complementary, where $\mathcal{Y} = E\mathcal{X} + A\mathcal{X}$ and $\mathcal{Y}_c = E\mathcal{X}_c + A\mathcal{X}_c$. Furthermore, there exists a decomposition, the so-called additive decomposition (AD) [11],

$$U^{-1}(\lambda E - A)V = \begin{bmatrix} \lambda E_{11} - A_{11} & 0 \\ 0 & \lambda E_{22} - A_{22} \end{bmatrix},$$

where V and U are no longer orthonormal, but their first n_1 columns span \mathcal{X} and \mathcal{Y} respectively, and their last n_2 columns form bases for \mathcal{X}_c and \mathcal{Y}_c respectively.

Of particular interest to the current paper is the case when \mathcal{D} is \mathbb{C}^+ or \mathbb{C}^- . There are several techniques proposed in the literature for the solution of the associated generalized SDC problem. Among these, we list the generalized matrix-sign-function iterations [10], the generalized Schur-decomposition method [8], and the inverse-free SDC method [5]. Given that the matrix pencil $\lambda E - A$ has no eigenvalues on the imaginary axis, we now outline the generalized matrix-sign-function-based algorithm of [10] for finding the right deflating subspaces of the pencil $\lambda E - A$ associated with the regions \mathbb{C}^- and \mathbb{C}^+ . The generalized matrix-sign-function iteration for the matrix pencil $\lambda E - A$ is given by

$$Z_0 = A, \quad Z_{k+1} = \frac{1}{2}(Z_k + EZ_k^{-1}E) \tag{5}$$

with the stopping criterion

$$\|Z_{k+1} - Z_k\|_1 < \varepsilon \|Z_{k+1}\|_1$$

(see [10]). When E is nonsingular (as is the case for the current paper), this iteration converges with quadratic rate to Z_∞ . Furthermore, $\mathcal{N}(E - Z_\infty)$ and $\mathcal{N}(E + Z_\infty)$ give the right deflating subspaces of the matrix pencil $\lambda E - A$ associated with the regions \mathbb{C}^+ and \mathbb{C}^- respectively [19]. Orthonormal basis vectors for the right and left deflating subspaces associated with either of the two regions can then be obtained using rank-revealing QR decompositions as described above and using the definition (3). We note that the SDC problem requires one generalized matrix-sign iteration plus two rank-revealing QR decompositions whereas AD requires two additional rank-revealing QR decompositions. Finally, the generalized Newton iteration requires the computation of the inverse of a matrix explicitly, leading to numerical instability for ill-conditioned matrices [4]. Inverse-free methods are the preferred methods if such numerical instability is detected, at the expense of more storage and arithmetic [5]. A comparison of such methods is beyond the scope of the current paper and we will use the approach outlined above with $\varepsilon = 10^{-7}$ in this paper for our numerical experiments.

3. Model

In this section, we will describe the general fluid-queue model based on [13] where fluid enters and leaves a single buffer system. The input and output rates of the fluid depend on an external environment which is described by a continuous-time Markov chain (CTMC). Let $S(t), t \geq 0$, be the governing irreducible CTMC on the state space $S = \{1, 2, \dots, M\}$ and with generator $A = [A_{ij}]$. Let $X(t)$ denote the amount of fluid in the buffer at time t and let the drift function $d(S(\cdot))$ denote the net rate of change to the buffer (arrival rate minus departure rate). Note that the drift function in this model only depends on the state of the underlying CTMC, so the drift function can be written as $d(\cdot)$, where $d(s)$ is called the drift in state $s \in S$. Without any loss of generality, we will assume throughout this paper that $d(s) \neq 0$ for all s since otherwise the associated algebraic equations can appropriately be removed.

When the buffer capacity is infinite, the dynamics of the buffer-content process $\{X(t), t \geq 0\}$ are given by

$$\frac{dX(t)}{dt} = \begin{cases} d(s) & \text{if } X(t) > 0 \text{ and } S(t) = s, \\ d(s)^+ & \text{if } X(t) = 0 \text{ and } S(t) = s, \end{cases}$$

where $x^+ = \max(x, 0)$. The relationship at the boundary $X(t) = 0$ ensures that the process X stays nonnegative for all t .

When the buffer capacity B is finite, the dynamics of the queue are governed by the following relationship:

$$\frac{dX(t)}{dt} = \begin{cases} d(s)^+ & \text{if } X(t) = 0 \text{ and } S(t) = s, \\ d(s) & \text{if } 0 < X(t) < B \text{ and } S(t) = s, \\ d(s)^- & \text{if } X(t) = B \text{ and } S(t) = s, \end{cases}$$

where $x^- = \min(0, x)$. The relationship at the boundary $X(t) = B$ ensures that the buffer does not exceed the buffer capacity B .

We are interested in finding the steady-state queue-length probabilities in both the infinite- and finite-buffer cases. We first define $F(t, s, x) := \Pr\{S(t) = s, X(t) \leq x\}$. Since the modulating Markov process is finite and irreducible, its equilibrium probabilities, $\pi_s := \lim_{t \rightarrow \infty} \Pr\{S(t) = s\}$, exist. We also define the invariant row vector $\pi = [\pi_1, \pi_2, \dots, \pi_M]$, which is the unique solution to $\pi A = 0, \pi e = 1$. The mean drift rate d of the buffer is expressed as $d := \sum_{s \in S} \pi_s d(s)$. A necessary and sufficient condition for the existence of equilibrium probabilities for the joint process (S, X) in the infinite-buffer case is $d < 0$. We note that no such condition is necessary for the existence of equilibrium probabilities in the finite-buffer case. Assuming the existence of the limiting probabilities $F(s, x) := \lim_{t \rightarrow \infty} F(t, s, x)$, we define the row vector $F(x) := [F(1, x), F(2, x), \dots, F(M, x)]$ and the nonsingular drift matrix $E := \text{diag}\{d(1), d(2), \dots, d(M)\}$. Once $F(x)$ is known, we can easily compute the queue-length survivor function (i.e. the probability of the queue length exceeding a certain value x), $q(x) := \lim_{t \rightarrow \infty} \Pr\{X(t) > x\} = 1 - F(x)e$.

We are now ready to state the following well-known result from [21] and [13].

Theorem 1. *In the infinite-buffer case, $F(x)$ satisfies*

$$\frac{dF(x)}{dx} E = F(x)A, \quad x \geq 0, \tag{6}$$

with the boundary conditions $F(s, 0) = 0$ if $d(s) > 0$. When the buffer capacity is finite, the differential equation (6) is valid when $0 \leq x < B$ and we use the following additional boundary condition: $F(s, B^-) = F(s, B) = \pi_s$ if $d(s) < 0$. The probability $P_{\text{full}}(s, B)$ that the buffer is full in state s when $d(s) > 0$ is then given using the discontinuity of $F(s, x)$ at B as $P_{\text{full}}(s, B) = \pi_s - F(s, B^-)$.

4. Infinite-buffer case

We first define S_- and S_+ to be the set of states such that $d(s) < 0$ and $d(s) > 0$ respectively. Also let M_- and M_+ be the cardinality of the corresponding sets. It is well known that (see [2]), when the stability condition $d < 0$ is satisfied, the generalized matrix-differential equation (6) has $M = M_+ + M_-$ generalized eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_M$, one of them, λ_{M_++1} , being at the origin, M_+ of them, $\lambda_1, \dots, \lambda_{M_+}$, having negative real parts, and $M_- - 1$ of them, $\lambda_{M_++2}, \dots, \lambda_M$, having positive real parts.

The boundary conditions can easily be formed as follows.

Boundary condition 1. *The limiting probability $F(s, x)$ is bounded for all x and $F(s, x) \rightarrow \pi_s$ as $x \rightarrow \infty$.*

Boundary condition 2. *For $s \in S_+$, the queue length is always increasing, so the queue length cannot be zero. Therefore, $F(s, 0) = 0$ for $s \in S_+$.*

In the traditional approach outlined in [13], $\phi_i = [\phi_{i1}, \phi_{i2}, \dots, \phi_{iM}]$ is defined as the left generalized eigenvector corresponding to the generalized eigenvalue λ_i of the matrix pair (E, A) such that the pair (λ_i, ϕ_i) satisfies $\phi_i(\lambda_i E - A) = 0$. Assume that the λ_i are distinct and they are numbered according to ascending real parts:

$$\operatorname{Re}(\lambda_1) \leq \dots \leq \operatorname{Re}(\lambda_{M_+}) < \operatorname{Re}(\lambda_{M_++1}) = 0 < \operatorname{Re}(\lambda_{M_++2}) \leq \dots \leq \operatorname{Re}(\lambda_M)$$

and normalize $\phi_{M_++1} = \pi$. The solution $F(\cdot)$ to (6) is given by the spectral representation

$$F(x) = \sum_{i=1}^M a_i e^{\lambda_i x} \phi_i, \tag{7}$$

where the a_i are unknown scalar coefficients to be determined from Boundary conditions 1 and 2. In particular, the coefficients satisfy the following condition: $a_i = 0$ if $\operatorname{Re}(\lambda_i) > 0$, $a_{M_++1} = 1$ (Boundary condition 1) and $\sum_{i=1}^{M_++1} a_i \phi_{ij}$ should be zero if $j \in S_+$ (Boundary condition 2).

In the following, we propose an alternative algorithmic method. We first define

$$W(s, x) := F(s, x) - \pi_s \quad \text{for } s = 1, \dots, M$$

and

$$W(x) := [W(1, x), W(2, x), \dots, W(M, x)],$$

so that $W(\cdot)$ satisfies the same generalized matrix-differential equation as in (6):

$$\frac{dW(x)}{dx} E = W(x) A.$$

For the matrices E and A , we can find the decomposition (4) using orthonormal matrices Q and Z such that $\lambda(E_{11}, A_{11}) \subset \mathbb{C}^{+0}$ and $\lambda(E_{22}, A_{22}) \subset \mathbb{C}^-$. Such a decomposition is possible using the traditional spectral decomposition techniques including the generalized Newton iterations (see Appendix A for the numerical algorithm). We also note that the dimensions of the matrices E_{11} and A_{11} , and of E_{22} and A_{22} are known *a priori* and they are $M_- \times M_-$ and $M_+ \times M_+$ respectively.

Given the decomposition (4), we define $[u(x), v(x)] := W(x)Q$, where $u(x)$ is a $1 \times M_-$ row vector and $v(x)$ a $1 \times M_+$ row vector, or, equivalently,

$$W(x) = u(x)L_1 + v(x)L_2, \quad \text{where } Q^\top = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}. \tag{8}$$

Then it is not difficult to show by direct substitution that

$$\frac{du(x)}{dx} E_{11} = u(x) A_{11}.$$

However, since $\lambda(E_{11}, A_{11}) \subset \mathbb{C}^{+0}$, $u(0)$ should be the zero vector since otherwise Boundary condition 1 would not be satisfied. Therefore, $u(x) = 0$ for all x and $v(x)$ satisfies a similar differential equation,

$$\frac{dv(x)}{dx} E_{22} = v(x) A_{22},$$

whose solution is

$$v(x) = v(0)e^{Gx}, \quad G = A_{22}E_{22}^{-1}. \tag{9}$$

Invertibility of E_{11} and E_{22} results from the nonsingularity of E . From (9) and (8), we conclude that

$$W(x) = v(x)L_2 = v(0)e^{Gx}L_2,$$

which, upon computation of $v(0)$, will give the desired solution. In order to find $v(0)$, we make use of Boundary condition 2, which indicates that $W(s, 0) = -\pi_s$ for $s \in S_+$. Let $s_1^0, s_2^0, \dots, s_{M_+}^0$ be the overload states of the underlying CTMC. Then we construct an $M \times M_+$ matrix T whose j th column consists of all zeros except for an entry 1 at the s_j^0 th position. Then Boundary condition 2 implies that $-\pi T = v(0)L_2T$. Solving for $v(0)$ from above, we can write the queue-length survivor function $q(\cdot)$ in a simple matrix-exponential form,

$$q(x) = -v(0)e^{Gx}L_2e,$$

where $-v(0)$ corresponds to f in (1) and L_2e corresponds to h .

5. Finite-buffer case

In the finite-buffer case, we still have the same generalized linear matrix-differential equation (6) which is only valid when $0 \leq x < B$. In this case, the condition $d < 0$ is not required for the existence of steady-state probabilities. However, the number of generalized eigenvalues of the pair (E, A) in each half of the complex plane changes with respect to the sign of the average drift d . As noted before, when $d < 0$, we have $M_+ + M_-$ generalized eigenvalues, one of them being at the origin, M_+ of them having negative real parts, and $M_- - 1$ of them having positive real parts. However, when $d > 0$, we have one generalized eigenvalue at the origin, $M_+ - 1$ with negative real parts, and M_- with negative real parts. This can be shown easily by checking the generalized eigenvalues of the pair $(-E, A)$. In this paper, we will not study the $d = 0$ case. Therefore, the number of generalized eigenvalues of the matrix pair (E, A) with negative real parts, say n , and with positive real parts, say p , are known *a priori* by simply checking the sign of the mean drift of the fluid queue. It is also noted that the system size is $M = n + p + 1$.

The boundary conditions for the finite-buffer problem are given as follows [13].

Boundary condition 1'. For $s \in S_-$, $F(s, B^-) = \pi_s$.

Boundary condition 2'. For $s \in S_+$, $F(s, 0) = 0$.

The solution $F(\cdot)$ to (6) is given by the spectral representation (7), where the a_i are unknown scalar coefficients to be determined from Boundary conditions 1' and 2'. In particular, the coefficients a_i satisfy the following conditions:

$$\sum_{i=1}^M a_i \phi_{ij} = 0, \quad j \in S_+,$$

$$\sum_{i=1}^M a_i \phi_{ij} e^{\lambda_i B} = \pi_j, \quad j \in S_-.$$

The above set of linear equations are known to be ill conditioned due to the coexistence of both exponentially dying and growing terms (with respect to the buffer size B) in the linear equation. Numerical instability of the above linear equations are reported and numerical methods are proposed to overcome this instability in [15] and [9].

We present our alternative approach here that avoids the individual computation of the generalized eigenvalues and eigenvectors and which is also numerically stable. This approach

is based on the following additive decomposition using matrices U and V :

$$U^{-1}EV = \begin{bmatrix} b & 0 & 0 \\ 0 & E_{11} & 0 \\ 0 & 0 & E_{22} \end{bmatrix}, \quad U^{-1}AV = \begin{bmatrix} 0 & 0 & 0 \\ 0 & A_{11} & 0 \\ 0 & 0 & A_{22} \end{bmatrix}, \quad (10)$$

where b is a nonzero number, $\lambda(E_{11}, A_{11}) \subset \mathbb{C}^-$ and $\lambda(E_{22}, A_{22}) \subset \mathbb{C}^+$. With (10), the differential system (6) is decomposed into three subsystems. These subsystems are associated with the matrix pairs $(b, 0)$, (E_{11}, A_{11}) , and (E_{22}, A_{22}) , and are referred to as the constant, stable, and antistable subsystems respectively. Such a decomposition is possible using the generalized Newton iterations (see Appendix B for the numerical algorithm). We also note that the dimensions of the matrices E_{11} and E_{22} are $n \times n$ and $p \times p$ respectively. Given the decomposition (10), we define $[t(x), u(x), v(x)] := F(x)U$, with $t(x)$ a scalar, $u(x)$ a $1 \times n$ row vector and $v(x)$ a $1 \times p$ row vector, or, equivalently,

$$F(x) = t(x)L_1 + u(x)L_2 + v(x)L_3, \quad \text{where } U^{-1} = \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}.$$

It is then easy to show that

$$\frac{dt(x)}{dx} = 0,$$

which yields that $t(x) = c$ (a constant). It can also be shown by direct substitution that

$$\frac{du(x)}{dx}E_{11} = u(x)A_{11}, \quad 0 \leq x < B.$$

Then $u(x)$, $0 \leq x < B$, can be written as

$$u(x) = u(0)e^{G_1x}, \quad G_1 = A_{11}E_{11}^{-1},$$

for a stable matrix G_1 . This subsystem is called the stable subsystem since G_1 is stable, but is also referred to as a forward subsystem since the solution of the differential equation is written in terms of its initial condition.

Similarly, we can write $v(x)$ in two mathematically equivalent forms, one being dependent on $v(0)$, the other dependent on $v(B)$:

$$v(x) = v(0)e^{G_2x}, \quad G_2 = A_{22}E_{22}^{-1}, \quad 0 \leq x < B,$$

or

$$v(x) = v(B)e^{-G_2(B-x)}, \quad 0 \leq x < B.$$

We note that $-G_2$ is a stable matrix, i.e. all eigenvalues of $-G_2$ have negative real parts, and $e^{-G_2x} \rightarrow 0$ as $x \rightarrow \infty$. We can then write $F(x)$, $0 \leq x < B$, as

$$F(x) = t(x)L_1 + u(x)L_2 + v(x)L_3 = cL_1 + u(0)e^{G_1x}L_2 + v(0)e^{G_2x}L_3 \quad (11)$$

$$= cL_1 + u(0)e^{G_1x}L_2 + v(B)e^{-G_2(B-x)}L_3. \quad (12)$$

We call this third subsystem the antistable subsystem since G_2 is antistable but, since in (12) the solution of the differential equation can be written in terms of its final condition, we also refer it to as the backward subsystem. The identity (12) dictates the form of the expression for $F(x)$; however, the coefficients in this expression, namely c , $u(0)$, and $v(B)$ (or $v(0)$ in (11)), should first be obtained. We will now use Boundary conditions 1' and 2' to find these coefficients.

Similar to the construction of the matrix T , we construct an $M \times M_-$ matrix N such that its j th column consists of all zeros except for an entry 1 at the s_j^u th position, where $s_1^u, s_2^u, \dots, s_{M_-}^u$ are the underload states of the underlying CTMC. With this definition of the matrices T and N , Boundary condition 2' implies that

$$\begin{aligned} 0 &= c\bar{L}_1 + u(0)\bar{L}_2 + v(0)\bar{L}_3, \\ &= c\bar{L}_1 + u(0)\bar{L}_2 + v(B)e^{-G_2B}\bar{L}_3, \end{aligned} \tag{13}$$

where $\bar{L}_i = L_iT$ for $i = 1, 2, 3$. On the other hand, Boundary condition 1' leads to the following linear equations:

$$\begin{aligned} \pi N &= c\tilde{L}_1 + u(0)e^{G_1B}\tilde{L}_2 + v(0)e^{G_2B}\tilde{L}_3, \\ &= c\tilde{L}_1 + u(0)e^{G_1B}\tilde{L}_2 + v(B)\tilde{L}_3, \end{aligned} \tag{14}$$

where $\tilde{L}_i = L_iN$ for $i = 1, 2, 3$. When (13) and (14) are combined, we have the following two equivalent linear systems:

$$[0 \ \pi N] = [c \ u(0) \ v(0)] \begin{bmatrix} \bar{L}_1 & \tilde{L}_1 \\ \bar{L}_2 & e^{G_1B}\tilde{L}_2 \\ \bar{L}_3 & e^{G_2B}\tilde{L}_3 \end{bmatrix} \tag{15}$$

and

$$[0 \ \pi N] = [c \ u(0) \ v(B)] \begin{bmatrix} \bar{L}_1 & \tilde{L}_1 \\ \bar{L}_2 & e^{G_1B}\tilde{L}_2 \\ e^{-G_2B}\bar{L}_3 & \tilde{L}_3 \end{bmatrix} \tag{16}$$

The system in (16) is better conditioned than the system in (15) since it does not have exponentially growing terms due to the stability of G_1 and $-G_2$. Therefore, we propose to solve for c , $u(0)$ and $v(B)$ using the linear system (16) as opposed to (15). We can then write the survivor function $q(\cdot)$ in the following modified matrix exponential form:

$$q(x) = 1 - cL_1e + (-u(0)e^{G_1x}L_2e) + (-v(B)e^{-G_2(B-x)}L_3e), \quad 0 \leq x < B,$$

where $1 - cL_1e$ corresponds to a in (2), $-u(0)$ to f_1 , L_2e to h_1 , $-v(B)$ to f_2 , and L_3e to h_2 .

6. Numerical results

In the first example, numerical stability of the proposed algorithm is investigated. In this example, we study the case of K statistically identical traffic sources that are multiplexed into a single buffer. Each source is modelled as a three-state Markov fluid source with one off state and two on states. We note that the aggregation of K such sources produces a state space of dimension $(K + 1)(K + 2)/2$. In the two on states, the rate of each source is 1. The on time is assumed to have a hyperexponential distribution with mean T_{on} and coefficient of variation $c_{on} > 1$ and with balanced means as described in [20]. The marginal probability of the source being in the on state, or the activity factor of each source, is denoted by ρ_s . The drain rate of the fluid queue is determined so as to achieve a desired overall use of ρ . For the infinite-buffer case, we measure the numerical stability of the algorithm by monitoring the value ϵ_{inf} defined as

$$\epsilon_{inf} := \|[F(0)T, F'(0)E - F(0)A]\|_1,$$

where the first term, $F(0)T$, is indicative of how well the boundary condition at zero is satisfied whereas the second term is a measure of how well the differential equation is satisfied at zero.

TABLE 1: Values of ε_{inf} and of ε_{fin} using the FB and FO schemes for varying use and buffer sizes.

	B	Use, ρ			
		0.9	0.9999	1.0001	1.1
ε_{inf}		6.0×10^{-16}	7.4×10^{-14}	x	x
ε_{fin} (FB)	100	6.2×10^{-16}	5.2×10^{-12}	3.6×10^{-12}	1.0×10^{-15}
ε_{fin} (FO)	100	7.9×10^{-16}	—	3.3×10^{-12}	5.6×10^{-16}
ε_{fin} (FB)	1000	8.5×10^{-16}	4.0×10^{-12}	2.3×10^{-12}	6.1×10^{-16}
ε_{fin} (FO)	1000	1.2×10^{-9}	—	1.9×10^{-11}	6.5×10^{-12}
ε_{fin} (FB)	10000	2.9×10^{-16}	1.4×10^{-12}	8.7×10^{-13}	2.5×10^{-16}
ε_{fin} (FO)	10000	—	—	—	—

For the finite-buffer case, we monitor a similar quantity, ε_{fin} , defined by

$$\varepsilon_{\text{fin}} := \|[F(0)T, F'(0)E - F(0)A, F(B^-)N - \pi N, F'(B^-)E - F(B^-)A]\|_1,$$

to take into consideration whether the boundary conditions as well as the differential equations are satisfied at the two boundaries. In this paper, we have proposed a scheme which is based on a forward-backward decomposition and uses $v(B)$ as the unknown along with the equations (12) and (16). We call this the FB scheme. The second scheme, which is to be compared with the FB scheme, uses the forward equations only through (11) and (15). We call this the FO scheme (forward only). In the FO scheme, $v(0)$ is used as the unknown, similar to the work reported in [21]. We fix $K = 10$, $T_{\text{on}} = 1000$, $c_{\text{on}} = \sqrt{2}$, and $\rho_s = 0.4$, and vary the parameters B and ρ in this numerical example. We use the generalized matrix-sign-function iterations as described in Appendices A and B for the decompositions (4) and (10). We report our results on the performance metrics ε_{inf} and ε_{fin} in Table 1. We observe that ε_{inf} is close to the machine zero u (that is $u \approx 2.2 \times 10^{-16}$) but we observe deteriorating stability as use gets closer to 1. However, this numerical stability problem is not specific to the Markov fluid model but is observed in more general queueing problems [1]. In such numerically challenging cases, we suggest the use of more stable decomposition techniques of linear algebra; see [5] for an extensive discussion of such techniques. It is numerically shown in Table 1 that ε_{fin} does not change with respect to the buffer size B when using the FB scheme. However, ε_{fin} grows without bound with increasing buffer sizes for all values of ρ tried using the FO scheme. We use ‘—’ to denote those cases to show that an acceptable solution cannot be obtained because the associated problem is very ill conditioned. These cases correspond to either probability density functions with negative values or ε_{fin} being too high to have any reliance on the result obtained. The notation ‘x’ is used to show that, for those values of $\rho > 1$, the corresponding infinite-buffer system is unstable and therefore does not have a stationary solution. In conclusion, rewriting the linear equation using the unknown $v(B)$ as in (16) eliminates the inherent numerical stability problem of finite Markov fluid queues reported in the literature [15], [9].

In Table 2, we study the numerical stability properties of the FB scheme when the number of the three state sources is increased. We fix $\rho = 0.9$, $T_{\text{on}} = 2$, $c_{\text{on}} = \sqrt{2}$, and $\rho_s = 0.4$, and vary the parameter K , or, equivalently, vary the state space dimension M . The results clearly demonstrate that the proposed FB scheme is robust in terms of numerical stability with respect to the dimension of the state space.

It is generally believed that the asymptotic behaviour of the infinite-buffer system can be used to approximate the corresponding finite-buffer system. However, in some finite-buffer scenarios studied, the authors observed that the queue-length PDF may not be dominated by the tail

TABLE 2: Values of ε_{fin} using the FB scheme with respect to increasing number of users.

K	M	ε_{fin}
25	351	6.3×10^{-14}
30	496	1.1×10^{-13}
35	666	1.6×10^{-13}
40	861	1.6×10^{-13}

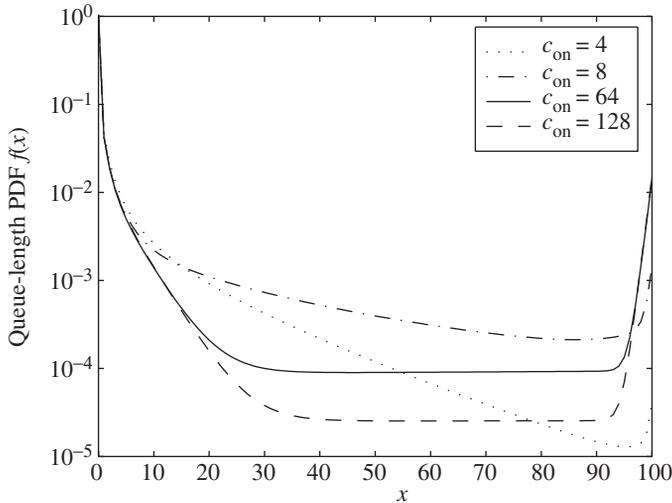


FIGURE 1: Queue-length PDF for varying c_{on} .

behaviour of the associated infinite-buffer system but instead it is the modes of G_2 that determine the PDF behaviour especially in the region close to $x = B$. As an example, we plot in Figure 1 the queue-length PDF $f(\cdot)$ for different values of c_{on} in the scenario $\rho = 0.7$, $T_{\text{on}} = 2$, $\rho_s = 0.4$, and $B = 100$. For convenience, the probability masses are not plotted at the two boundaries in Figure 1. This example clearly demonstrates that, for larger values of c_{on} , the behaviour of the PDF is dominated by the stable and the antistable subsystems for small and large values of x respectively. For values of x distant from the boundaries, it is the constant term that dominates.

7. Conclusions

In this paper, we present a novel method for solving for the steady-state solution of infinite- or finite-buffer Markov fluid queues. This method is based on the spectral decomposition of a matrix pair (E, A) and can be carried out using the powerful techniques of numerical linear algebra including generalized Newton iteration. With this method in place, we show that the finite-buffer solution does not have the numerical stability problems reported in the literature. In cases when numerical instability is encountered using the approach in this paper, due to either the use ρ being close to 1 or the drift matrix E being very close to singularity, we suggest the use of the more stable (but slower) techniques of linear algebra [5]. We also studied several numerical examples demonstrating that finite-buffer systems have their peculiar characteristics and in some scenarios they cannot be approximated well by their infinite-buffer counterparts.

Appendix A. Generalized Newton iteration for (4)

Given the pair of matrices E and A in (6), our goal is to find orthonormal matrices Q and Z such that the decomposition (4) holds, where $\lambda(E_{11}, A_{11}) \subset \mathbb{C}^{+0}$ and $\lambda(E_{22}, A_{22}) \subset \mathbb{C}^-$. We define the rank-one update $\bar{A} := A + Ee\pi E/\pi Ee$ of the matrix A . With this definition, the generalized eigenstructure of the matrix pair (E, \bar{A}) is the same as that of the pair (E, A) except for one generalized eigenvalue at 1 which is obtained by moving the generalized eigenvalue of (E, A) at the origin to the right by one unit. Moreover, \bar{A} is nonsingular. We note that, when the matrices E and \bar{A} are nonsingular, as is the case in this paper, the generalized Newton iteration given in (5) applied to the matrix pencil $\lambda E - \bar{A}$ converges to Z_∞ . Following Section 2, we let $(E - Z_\infty)^\top \Pi = [Z_1, Z_2]R$ be a rank-revealing QR factorization of $(E - Z_\infty)^\top$, where the orthonormal matrix Z_2 has M_- columns. With the additional rank-revealing QR factorization $[\bar{A}Z_2, EZ_2]\Pi = QR$, and defining $Z = [Z_2, Z_1]$, we obtain the orthonormal matrices Q and Z such that (4) holds with the desired spectra of the two subblocks.

Appendix B. Generalized Newton iteration for (10)

We propose finding the matrices Y and X so that

$$Y^{-1}EX = \begin{bmatrix} b & 0 \\ 0 & E_2 \end{bmatrix}, \quad Y^{-1}AX = \begin{bmatrix} 0 & 0 \\ 0 & A_2 \end{bmatrix} \tag{17}$$

as the first step towards the decomposition (10). Then, following [18], we define an orthogonal matrix Z such that $Ze_1 = e$, where e_1 is a column vector whose first element is 1 and other entries all zero. Similarly, the orthogonal matrix Q is defined through $Qe_1 = Ee =: \bar{e}$. The orthonormal matrices Z and Q can easily be obtained via the QR decomposition of the matrices

$$\begin{bmatrix} 1 & 0 \\ e & I_{M-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \bar{e}(1) & 0 \\ \bar{e}(2 : M) & I_{M-1} \end{bmatrix}$$

respectively, where $\bar{e}(1)$ is the first entry of \bar{e} . It is then well known (see [18]) that

$$Q^\top EZ = \begin{bmatrix} E_{11} & E_{12} \\ 0 & E_{22} \end{bmatrix}, \quad Q^\top AZ = \begin{bmatrix} 0 & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

where the square matrices E_{22} and A_{22} are both of dimension $(M - 1) \times (M - 1)$. Solving for X_1 in $X_1A_{22} = A_{12}$ and solving for Y_1 in $-E_{11}Y_1 = E_{12} - X_1E_{22}$, we define the matrices X and Y as follows:

$$X = Z \begin{bmatrix} 1 & Y_1 \\ 0 & I \end{bmatrix} \quad \text{and} \quad Y = Q \begin{bmatrix} 1 & X_1 \\ 0 & I \end{bmatrix}.$$

This choice of the matrices X and Y can be shown by direct substitution to satisfy the decomposition in (17).

In the second step, we use the generalized Newton iteration applied to the matrix pencil $\lambda E_2 - A_2$ to find Z_∞ . We recall that the pair (E_2, A_2) has n generalized eigenvalues with negative real part and p with positive real part. Following Section 2, we use the rank-revealing QR decomposition $(E_2 + Z_\infty)^\top \Pi = [Z_1, Z_2]R$ of $(E_2 + Z_\infty)^\top$ with Z_2 having n columns. We then let $[A_2Z_2, E_2Z_2]\Pi = [Q_1, Q_2]R$ be a rank-revealing QR factorization of $[A_2Z_2, E_2Z_2]$ with Q_1 having n columns. As described in Section 2, the orthonormal matrices Z_2 and Q_1 span, respectively, the right and left deflating subspaces of the matrix pencil $\lambda E_2 - A_2$ associated with the region \mathbb{C}^- . Similarly, we use the rank-revealing QR decomposition $(E_2 - Z_\infty)^\top \Pi = [\bar{Z}_1, \bar{Z}_2]R$ of $(E_2 - Z_\infty)^\top$ with \bar{Z}_2 having p columns. We

then let $[A_2\bar{Z}_2, E_2\bar{Z}_2]\Pi = [\bar{Q}_1, \bar{Q}_2]R$ be a rank-revealing QR factorization of $[A_2\bar{Z}_2, E_2\bar{Z}_2]$ with \bar{Q}_1 having p columns. As described in Section 2, the orthonormal matrices \bar{Z}_2 and \bar{Q}_1 span, respectively, the right and left deflating subspaces of the matrix pencil $\lambda E_2 - A_2$ associated with the region \mathbb{C}^+ . It is not then difficult to show by referring to Section 2 that the choice of the matrices

$$U := Y \begin{bmatrix} 1 & 0 \\ 0 & [Q_1, \bar{Q}_1] \end{bmatrix} \quad \text{and} \quad V := X \begin{bmatrix} 1 & 0 \\ 0 & [Z_2, \bar{Z}_2] \end{bmatrix}$$

leads to the desired decomposition in (10).

Acknowledgements

This work was supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK) under projects EEEAG-101E025 and EEEAG-101E048, and NSF Award INT-011577.

References

- [1] AKAR, N. AND SOHRABY, K. (1997). An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Stoch. Models* **13**, 381–416.
- [2] ANICK, D., MITRA, D. AND SONDHI, M. M. (1982). Stochastic theory of a data handling system with multiple sources. *Bell Syst. Tech. J.* **61**, 1871–1894.
- [3] ASMUSSEN, S. (1995). Stationary distributions for fluid flow models with or without Brownian motion. *Stoch. Models* **11**, 21–49.
- [4] BAI, Z. AND DEMMEL, J. W. (1993). Design of a parallel nonsymmetric eigenroutine toolbox, Part I. In *Proc. 6th SIAM Conf. Parallel Processing for Scientific Computing, Vol. 1*, eds R. F. Sincovec *et al.*, Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 391–398.
- [5] BAI, Z., DEMMEL, J. AND GU, M. (1997). Inverse free parallel spectral divide and conquer algorithms for nonsymmetric eigenproblems. *Numer. Math.* **76**, 389–396.
- [6] CHAN, T. F. (1987). Rank revealing QR factorizations. *Linear Algebra Appl.* **88/89**, 67–82.
- [7] DEMMEL, J. AND KÄGSTRÖM, B. (1987). Computing stable eigendecompositions of matrix pencils. *Linear Algebra Appl.* **88/89**, 139–186.
- [8] DEMMEL, J. AND KÄGSTRÖM, B. (1993). The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. I. Theory and algorithms. *ACM Trans. Math. Software* **19**, 160–174.
- [9] FIEDLER, M. AND VOOS, H. (2000). New results on the numerical stability of the stochastic fluid flow model analysis. In *Proc. NETWORKING 2000, Broadband Communications, High Performance Networking, and Performance of Communication Networks*, Springer, Paris, pp. 446–457.
- [10] GARDINER, J. D. AND LAUB, A. J. (1986). A generalization of the matrix-sign-function solution for algebraic Riccati equations. *Internat. J. Control* **44**, 823–832.
- [11] KÄGSTRÖM, B. AND DOOREN, P. V. (1992). A generalized state-space approach for the additive decomposition of a transfer matrix. *J. Numer. Linear Algebra Appl.* **1**, 165–181.
- [12] KOSTEN, L. (1984). Stochastic theory of data-handling systems with groups of multiple sources. In *Performance of Computer Communication Systems*, eds H. Ruding and W. Bux, Elsevier, Amsterdam, pp. 321–331.
- [13] KULKARNI, V. G. (1997). Fluid models for single buffer systems. In *Frontiers in Queuing: Models and Applications in Science and Engineering*, ed. J. H. Dshalalow, CRC Press, Boca Raton, FL, pp. 321–338.
- [14] MITRA, D. (1988). Stochastic theory of a fluid model of producers and consumers coupled by a buffer. *Adv. Appl. Prob.* **20**, 646–676.
- [15] NAGARAJAN, R., KUROSE, J. F. AND TOWSLEY, D. (1991). Approximation techniques for computing packet loss in finite-buffered voice multiplexers. *IEEE J. Selected Areas Commun.* **9**, 368–377.
- [16] ROGERS, L. C. G. (1994). Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Ann. App. Prob.* **4**, 390–413.
- [17] SCHWARTZ, M. (1994). *Broadband Integrated Networks*. Prentice-Hall, Upper Saddle River, NJ.
- [18] SUN, X. AND QUINTANA-ORTI, E. S. (2004). Spectral division methods for block generalized Schur decompositions. To appear in *Math. Computation*.
- [19] SUN, X. AND QUINTANA-ORTI, E. S. (2003). The generalized Newton iteration for the matrix sign function. *SIAM J. Sci. Comput.* **24**, 669–683.
- [20] TJMS, H. C. (1986). *Stochastic Modelling and Analysis: A Computational Approach*. John Wiley, New York.
- [21] TUCKER, R. (1988). Accurate method for analysis of a packet speech multiplexer with limited delay. *IEEE Trans. Commun.* **36**, 479–483.