STOCHASTIC MODELS
Vol. 19, No. 3, pp. 325–348, 2003

Fluid Flow Models and Queues—A Connection by Stochastic Coupling

Soohan Ahn* and V. Ramaswami*

AT&T Labs, Florham Park, New Jersey, USA

ABSTRACT

We establish in a direct manner that the steady state distribution of Markovian fluid flow models can be obtained from a quasi birth and death queue. This is accomplished through the construction of the processes on a common probability space and the demonstration of a distributional coupling relation between them. The results here provide an interpretation for the quasi-birth-and-death processes in the matrix-geometric approach of Ramaswami and subsequent results based on them obtained by Soares and Latouche.

Key Words: Fluid flows; Queues; Stochastic coupling; Matrix geometric method.

1. INTRODUCTION

The subject of this paper is the steady state distribution of the content of a fluid flow buffer modulated by a finite state, continuous time, irreducible Markov chain of environmental "phases." In a classic paper, S. Asmussen[4] characterized that distribution to be of phase type. Later, V. Ramaswami[17] demonstrated that a representation of that phase type distribution can be obtained from the *G*-matrix of a discrete time, discrete state Quasi Birth and Death Process[14,18], thereby simplifying the analysis of the fluid flow model. The connection to a QBD was a substantially new result that reduced the continuous time, continuous state space

*Correspondence: Soohan Ahn and V. Ramaswami, AT&T Labs, Research, 180 Park Avenue, Florham Park, NJ 07932, USA; E-mail: ahn@research.att.com; vramaswami@research.att.com.

DOI: 10.1081/STM-120023564
1532-6349 (Print); 1532-4214 (Online)

problem of the fluid model to the analysis of a discrete time, discrete state space QBD for which well tested stable algorithms that avoid the computational difficulties arising in the spectral methods (see Ref.[3] and other references in Ref.[17]) exist. The derivation in Ref.[17] was based on a level crossing argument for the fluid flow process, a duality resulting from time reversal,[16,5] and uniformization. Building on those results, Soares and Latouche[19] obtained a representation in terms of a QBD which was arrived at without using time reversal in the analysis, thereby simplifying the approach; a probabilistic interpretation was also attempted by them of their QBD.

The QBDs arising in the matrix-geometric approach have not yet received a satisfactory interpretation. Our main purpose in this paper is to show that the approach via QBDs can be interpreted as being rooted in a distributional coupling of the fluid model to a queue described by the QBD.

Our construction of the queue may have the flavor of the approach in Ref.[6,13], but unlike them. we take into account the detailed structure within the on–off periods, and that is what leads to the consideration of a simple process, namely a discrete time QBD, that provides not only the distribution of the fluid level but also the joint distribution of the fluid level and the environment. While other embedded epochs may still yield a matrix-geometric structure, it is well-known that among matrix-geometric models, the QBD is the simplest, both conceptually and algorithmically. (We hasten to note, however, that the paper[13] covers very general models besides fluid models driven by finite state Markov chains considered here.)

One may also consider other types of queueing models, one such being a queue with interarrival dependent service times. However, the consideration of models with independent service times is what leads us to a quasi birth and death process. Unlike this, the model with interarrival dependent service times does not possess a convenient Markov (renewal) structure at both arrival and departure epochs. The full power of our approach will be seen in a subsequent paper[2] where we demonstrate some strong coupling results that yield the transient analysis of the fluid model as well. The steady state analysis requires only the weaker results which are much simpler, and are the only ones considered here.

In the sequel, we assume that the Markov chain of phases is irreducible and that its state space $S_1 \cup S_2 \cup S_3$ is such that: during sojourn of the Markov chain in state i of S_1, the fluid level changes at rate $c_i > 0$; during sojourn of the Markov chain in state j of S_2, the fluid level changes at rate $c_j < 0$; and, during sojourn in S_3, the fluid level remains constant. We partition the states of the Markov chain in conformity with the three sets identified above and denote the infinitesimal generator of the underlying Markov chain by

$$Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix}. \tag{1}$$

Here Q_{ij} is a matrix of order $|S_i| \times |S_j|$.

Let C denote a diagonal matrix formed by the absolute values $|c_i|$, $i \in S_1 \cup S_2$ and an adequate number of 1's as needed to make C to be of the same order as Q. We define $T = C^{-1}Q$ and note that T is the infinitesimal generator of a Markov chain. It is well-known that the distribution of the general fluid flow model considered above can be determined easily from that of the fluid flow governed by the Markov chain with generator T in which the fluid level increases at rate 1 in S_1, decreases at rate 1 in S_2, and remains constant in S_3. In the sequel, we shall refer to the latter model as the "homogenized flow model." We shall derive the necessary conversion formulae at the end of the paper; these can also be found in Ref.[19] and are included here only for completeness.

2. ASSOCIATED QUEUE

2.1. Motivation

We will construct a queueing model represented by a QBD and a process $\mathcal{Y}$ constructed from its work process such that: (a) the process $\mathcal{Y}$ is distributionally coupled to the homogenized fluid flow model at the epochs of a Poisson process; and (b) the steady state distribution of $\mathcal{Y}$ coincides with that of the fluid flow model. To that end, it is convenient to assume that the phase process governed by T has been uniformized by a Poisson process. That is, we choose a (fixed) number $\lambda \geqq max_i(-T_{ii})$ and view the Markov process of phases as a process that makes changes of states only at epochs of a Poisson process with rate λ in such a way that successive states visited form a discrete time Markov chain, independent of the Poisson process, with transition matrix $P = \lambda^{-1}T + I$. Here and throughout the rest of the paper, I denotes an identity matrix of appropriate order. In the remainder of the paper, we shall consider the matrices T and P also as partitioned according as Q, with their submatrices denoted by the symbols T_{ij} and P_{ij} respectively.

To effect the stochastic coupling arguments, first assume the following as given on a *common* probability space $(\Omega, \mathcal{A}, \mathcal{P})$: A pair of independent Poisson processes, say, $\mathcal{M}$ and $\mathcal{N}$, with common rate λ; a discrete time Markov chain $\{L_n : n \geq 0\}$ of phases which has transition matrix P and is independent of the Poisson processes $\mathcal{M}$ and $\mathcal{N}$. Without loss of generality, we shall assume that $L_0 = i$ for some i. With these as building blocks, we will construct for almost all sample points in Ω, (a) a phase process $\mathcal{J} = \{J(t) : t \geq 0\}$ which is a CTMC with generator T; (b) a process $\mathcal{F} = \{F(t) : t \geq 0\}$ such that $F(t)$ increases at rate 1 while $J(t) \in S_1$, decreases at rate 1 while $J(t) \in S_2$ and $F(t)$ is positive, and remains constant while $J(t) \in S_3$—i.e., $(F(t), J(t))$ is a version of the homogenized fluid flow process of interest; and (c) a queueing model $\mathcal{Q} = \{Q(t) : t \geq 0\}$ modulated by the phase process, where $Q(t)$ is the queue length at time t. The queue $\mathcal{Q}$ is to be constructed such that at the epochs $\{s_k : k \geq 0\}$ of the process $\mathcal{M} \oplus \mathcal{N}$, where $\oplus$ denotes superposition of processes, (which will be such that the phase process $\mathcal{J}$ remains constant in each $[s_k, s_{k+1})$), the embedded sequence $\{(F(s_k), J(s_k)) : k \geq 0\}$ has the same probability law as that of the embedded sequence $\{(W(s_k), J(s_k)) : k \geq 0\}$, where $W(s_k)$ is the amount of work in the queue at the epoch s_k+. (Throughout, unless otherwise stated, for any process under consideration "state at a point t" will always denote the state at $t+$.)

Associated with the work in the queue we also construct a process $\mathcal{Y} = \{Y(t) : t \geq 0\}$ as follows. First, we discard from the set $\{s_k : k \geq 0\}$ of the points of $\mathcal{M} \oplus \mathcal{N}$, the points s_k of the process $\mathcal{N}$ for which $J(s_k) \in S_1 \cup S_3$, and call the resulting set of points as

$\{t_k : k \geq 0\}$. Now, for $t \in (t_k, t_{k+1})$, let

$$Y(t) = W(t_k) + (t - t_k), \quad if \quad J(t_k) \in S_1$$

$$= max(0, W(t_k) - (t - t_k)), \quad if \quad J(t_k) \in S_2$$

$$= W(t_k), \quad if \quad J(t_k) \in S_3.$$

We will note that in each of the intervals $[t_k, t_{k+1})$ the phase process remains constant and the growth of $Y(\cdot)$ mimics that of $F(\cdot)$.

So that we may motivate the details to follow, we state some facts we will establish in the sequel: (a) The amount of work done for each customer in the queue $\mathcal{Q}$ will be distributed as $exp(\lambda)$, and $\{(Y(t_k)\}_{k \geq 0}$ can be considered to be a "stochastic discretization" of the fluid model with respect to the uniformization parameter λ; (b) if the fluid model is stable, then the steady state distribution as $t \to \infty$ of $(Y(t), J(t))$ will exist and equal the steady state distribution of $(F(t), J(t))$ as $t \to \infty$; (c) steady state results for $\mathcal{Y}$ can be obtained using the matrix-geometric method for a QBD.

2.2. The Construction

To avoid pedantry and to save notations, we shall suppress the sample point in the ensuing discussion which is indeed a sample point by sample point construction.

Construction of the Phase Process: Let $0 = a_0$ and let $0 < a_1 < a_2 < \cdots$ denote the successive epochs of the Poisson process $\mathcal{M}$. We define $J(t) = L_n$ in the interval $a_n \leqq t < a_{n+1}$, where L_n is the state visited by the discrete time Markov chain at step n. Clearly $\{J(t) : t \geq 0\}$ is a continuous time Markov chain with infinitesimal generator T, and the epochs $\{a_n\}$ form a set of λ-uniformization epochs for the phase process.

Construction of the Homogeneous Fluid Flow: Without loss of generality, we will assume the initial condition $F(0) = 0$. We let $F(0) = 0$ and define the process $\{F(t)\}$ such that for $t \in [a_n, a_{n+1})$, $F(t) = F(a_n) + (t - a_n)$ if $J(t) \in S_1$, $F(t) = max[0, F(a_n) - (t - a_n)]$ if $J(t) \in S_2$, and finally $F(t) = F(a_n)$ if $J(t) \in S_3$. Defined thus, clearly $F(\cdot)$ increases at rate 1 in S_1, decreases at rate 1 in S_2 while it remains positive, and remains constant in S_3 as required. Clearly, the joint process $\{(F(t), J(t))\}$ is stochastically equivalent to the homogenized fluid model that starts empty and in phase i, and is modulated by the Markov process with generator T.

Construction of the Queue $\mathcal{Q}$: The queue $\mathcal{Q}$ will be defined in terms of the successive embedded epochs $t_0 = 0$, and $\{t_k : k \geq 1\}$ where there is an arrival, departure, or phase transition; we emphasize that some phase transitions may be from a phase to itself, as necessitated for instance in the uniformization process. It will be assumed that all queues are FIFO, and service is rendered by the server (at unit rate per unit time) only when the phase is in S_2; specifically, no service is rendered when the phase is in $S_1 \cup S_3$ Also, the queue size at time 0 will be defined to be 0 to match our initial condition $F(0) = 0$ (other initial conditions can be accommodated with minor changes in the construction.) In the sequel we shall denote by Q_k and J_k the queue length (number of customers in the system $\mathcal{Q}$) and the phase $J(t_k)$ at the epoch t_k+.

(a) Let $t_0 = 0$ and $Q_0 = 0$; this initializes the queue size at time 0 to match our initial state specification $F(0) = 0$ for the fluid model. Note that we have $J_0 = J(0) = i$ from the construction of the phase process.

(b) Having defined t_k and (Q_k, J_k), we first specify the next time point t_{k+1} and then the value of the queue size and phase immediately after that epoch. The queue size is assumed to remain constant over intervals of the form $[t_k, t_{k+1})$; that is, we shall set $Q(t) = Q_k$ for all $t \in [t_k, t_{k+1})$. There are several cases to consider.

Case 1: If $J_k \in S_1$, then t_{k+1} is the first epoch in $\mathcal{M}$ to come after t_k, and the next queue length value Q_{k+1} is set to $1 + Q_k$—that is, in this case, the epoch t_{k+1} is defined to be an arrival epoch to the queue $\mathcal{Q}$. The phase at J_{k+1} is set to $J(t_{k+1})$; note that a phase change occurs at the newly defined epoch iff at the epoch t_{k+1} a different phase is entered in the uniformization scheme; otherwise, that epoch will constitute a self-transition for the phase in the queue $\mathcal{Q}$.

Case 2: If $J_k \in S_3$, then the next epoch t_{k+1} is once again the first epoch in $\mathcal{M}$ to come after t_k, but the queue length value Q_{k+1} is set to the same value as Q_k—that is, a construct is made that makes the queue length remain constant just as the fluid level would remain constant over the interval under consideration (note that we are assuming that no work is being done in S_3.) The phase J_{k+1} is set to $J(t_{k+1})$; note that a phase transition to a different phase occurs at the newly defined epoch iff the new phase entered is indeed different; otherwise, the epoch is to be treated as a self-transition epoch.

Case 3: If $J_k \in S_2$, then the next epoch t_{k+1} is the first epoch in the superposition $\mathcal{M} \oplus \mathcal{N}$ to come after t_k. The queue length at that epoch is set depending on whether that epoch comes from $\mathcal{M}$ or from $\mathcal{N}$. Specifically, the next queue length value Q_{k+1} is set to the same value as Q_k if $t_{k+1} \in \mathcal{M}$; it is changed to $max(0, Q_k - 1)$ if the new epoch $t_{k+1} \in \mathcal{N}$. Thus, the next epoch is just a phase transition epoch (with no effect on queue size) if it is an epoch of $\mathcal{M}$, and a departure epoch (with no phase change) if the epoch is in $\mathcal{N}$ and a departure is indeed possible; note that except when the epoch is in $\mathcal{M}$ and the new phase entered is different, the new epoch is a dummy phase change transition epoch (i.e., with a phase self transition).

Remark 1. If one considers the fluid model at the end of sojourn in each phase of S_1 where the trajectory is upward, one could replace the continuous upward increments by a jump and pretend as though a customer with an exponentially distributed service time has been added. But we need to also generate the departure epochs of those customers if we are to analyze the queue. In our construction, the process $\mathcal{N}$ is used to "insert" the departure epochs.

3. DISTRIBUTIONAL COUPLING

The following result is obvious from the way the construction has been carried out.

Theorem 1. (*a*) *The queue $\mathcal{Q}$ is modulated by the same continuous time phase process $J(\cdot)$ that modulates the fluid model.*

(*b*) *The queue* $\{(Q(t), J(t)) : t \geq 0\}$ *is represented by a QBD* (*in continuous time*), *and the embedded sequence* $\{(Q(s_k), J(s_k)) : k \geq 0\}$ *is a discrete time QBD.*

Note that in trying to match certain behavior in the constructed queue to that of the fluid model, we could specify the queue only in terms of its queue length process. We now show formally that each arrival in the queue brings in a random amount of work distributed as *exp*(λ) independently of the history of the process up to the arrival epoch. This is the last assertion of the next theorem whose other assertions are essentially obvious consequences of our construction.

Theorem 2. *The queueing model* $\mathcal{Q}$ *satisfies the following conditions:*

a. Arrivals to the queue occur only at those epochs t_k *for which* $J(t_k-) \in S_1$; *that is, the epoch is a phase transition epoch in* $\mathcal{M}$ *from* S_1 (*which may very well be a phase self transition.*)

b. Departures to the queue can occur at t_k *only if* $J(t_k-) \in S_2$, $Q_{k-1} > 0$ *and* $t_k \in \mathcal{N}$. *Also, the phase immediately after each departure epoch is the same as that immediately prior to that epoch.*

c. Assume that work gets depleted at unit rate while $J(t) \in S_2$, *and that no service is rendered while* $J(t) \in S_1 \cup S_3$. *Then the amounts of work done between successive departure epochs are iid random variables distributed as exp*(λ).

Proof. We only need to prove (c). The service of a customer begins at an epoch s_k with $J_k \in S_2$. The service completion epoch is the first epoch s_{k+r}, $r \geqq 1$ to come after t_k for which $J(s_{k+r}) \in S_2$ and $s_{k+r} \in \mathcal{N}$. We shall evaluate now the distribution of the work done for the customer, given the phase at the beginning of the service.

First of all, note that if we denote by L_{12} and L_{32} the matrices of first passage probabilities for the phase process into S_2 from S_1 and S_3 respectively, then the matrix $L = P_{22} + P_{21}L_{12} + P_{13}L_{32}$ is the matrix of return probabilities into the set S_2, and that under the assumption that P is irreducible (and hence recurrent non-null), L is stochastic; i.e., $L\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes a column vector of 1's.

Let $\phi_i(s)$ denote the transform of the amount of work done by the server until the next departure epoch, given that service starts in phase $i \in S_2$. Let $\phi(s)$ denote the vector with elements $\phi_i(s), i \in S_2$. A simple conditioning argument gives us the following equation:

$$\phi(s) = \frac{2\lambda}{s + 2\lambda}\frac{1}{2}\mathbf{1} + \frac{2\lambda}{s + 2\lambda}\frac{1}{2}L\phi(s). \tag{2}$$

In the above equation, the first term covers the case when the first step of $\mathcal{M} \oplus \mathcal{N}$ itself marks the service completion; this happens if the next epoch is an epoch of $\mathcal{N}$, and that has probability 1/2. The second term corresponds to the case when the first step corresponds to a point from $\mathcal{M}$ (this marks a phase transition without a departure) and the probability of this is again 1/2. In each case, the amount of service rendered in the first step has an exponential distribution with parameter 2λ, and to this random variable must be added the remaining amounts of service, if any, rendered to the customer. In the second case, the phase could change, and we need to wait until a return into S_2 is made before service commences again; all that is reflected in the term L we have used there.

Solving Eq. (2) for $\phi(s)$ and simplifying the result shows that

$$\phi(s) = \left[I - \frac{\lambda}{s+2\lambda}L\right]^{-1} \frac{\lambda}{s+2\lambda}\mathbf{1}$$

$$= \sum_{n=0}^{\infty} \left\{\frac{\lambda}{s+2\lambda}L\right\}^{n} \frac{\lambda}{s+2\lambda}\mathbf{1}$$

$$= \frac{\lambda}{s+\lambda}\mathbf{1}, \qquad (3)$$

and we have used the fact that $L\mathbf{1} = \mathbf{1}$. Thus, the marginal distribution of the amount of service rendered (irrespective of the starting phase) is exponential with parameter λ, and the proof is complete. □

The next result provides a distributional coupling that plays a key role in our analysis.

Theorem 3. *The process* $\{(Y(s_k), J(s_k)) : k \geq 0\}$ *has the same probability law as the process* $\{(F(s_k), J(s_k)) : k \geq 0\}$.

Proof. We will show by mathematical induction that for all $m \geq 0$, the subsequences up to $k = m$ have the same joint distribution. First of all, note that the phase process is identical for both the queue and the fluid. Also, since $F(0) = Y(0) = 0$, the distributional equality holds trivially for $m = 0$.

To complete the proof by induction, since $\{(Y(s_k), J(s_k)) : k \geq 0\}$, and $\{(F(s_k), J(s_k)) : k \geq 0\}$, are both Markov, it suffices to show that for all (x, j), the conditional distribution of $(Y(s_m + 1), J(s_{m+1}))$ given $(Y(s_m), J(s_m)) = (x, j)$ is the same as that of $(F(s_{m+1}), J(s_{m+1}))$ given $(F(s_m), J(s_m)) = (x, j)$.

We now need to consider several cases: (a) If $J(s_m) \in S_1$, then both $F(s_m)$ and $Y(s_m)$ are each increased by independent random variables that have $exp(2\lambda)$ distributions to yield $F(s_{m+1})$ and $Y(s_{m+1})$; (b) If $J(s_m) \in S_3$, by our construction, the values of neither the work in the queue nor the fluid level changes at the next step; (c) In the case where $J(s_m) \in S_2$, in the interval $[s_m, s_{m+1})$ both the fluid level and the work in the queue are depleted at a unit rate as long as they remain positive and left alone once they hit zero; furthermore, no increment occurs to either values at s_{m+1}. In all the cases, therefore, we get equality for the conditional distributions of the values at the next step given they have the same values at the current step. Hence the proof is complete by mathematical induction.

□

Remark 2. We emphasize that what the theorem above gives is only an equality of distributions and not the equality of values of the random variables $F(s_k)$ and $Y(s_k)$ at the embedded points. In other words, what we have is a distributional coupling only. The main value of the theorem lies in the fact that the work in the queue we have constructed increases, whenever it does, by an amount that is independent of the past history of the queue. Unlike this, the fluid level increases by exactly the length of the sojourn interval preceding the epoch of increase. Note that the paths of the fluid model considered directly as paths of the process of work in some queue does not lead to a Markov renewal structure at the embedded epochs; in such a construction, each service time is precisely equal to some previous interarrival time, and dependencies are not just of one step.

Remark 3. The successive epochs $\{t_k\}$ mark epochs at which, compared to the previous epoch, the fluid level is up or down by $\exp(\lambda)$ distributed amounts or remains constant. Hence, the term "stochastic discretization." Such a procedure has been considered by Adan and Resing in Ref.[1] in an operational manner; we thank Adan for bringing[1] to our attention. Our work here and in Ref.[2] makes the approach rigorous. More importantly, the work here sets the stage for transient results in Ref.[2] which are not obtainable by the approach of[1] base on long run average rewards only.

4. ANALYSIS OF $\mathcal{Q}$

The queue $\mathcal{Q}$ can be analyzed through the embedded QBD at the epochs $\{s_k : k \geq 0\}$. Here, however, we will analyze it through the embedded QBD at the epochs $\{t_k : k > 0\}$ since that will help us to reconcile results obtained here with those in previous papers more easily.

4.1. Embedded QBD

Let us define $Q_k = Q(t_k)$. It is trivial to see that the process $\{(Q_k, J(t_k)) : k \geq 0\}$ is a Markov chain. In the following, the possible values of the queue length Q_n will be called as "levels."

The probabilities of the chain $(Q_n, J(t_n))$ making a transition to the next higher level is clearly governed by the substochastic matrix

$$P_0 = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4}$$

Similarly, the probabilities of going downward in levels is governed by the substochastic matrix

$$P_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2}I & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{5}$$

Finally, in the steps where the process Q_n does not change level, the probabilities are governed by the substochastic matrix

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2}P_{21} & \frac{1}{2}P_{22} & \frac{1}{2}P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}. \tag{6}$$

Indeed, if one considers the successive epochs of "arrivals, service completions and phase changes" as defined above, one gets a discrete time QBD with transition matrix

$$\begin{bmatrix} M & P_0 & 0 & 0 & \cdots \\ P_2 & P_1 & P_0 & 0 & \cdots \\ 0 & P_2 & P_1 & P_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \tag{7}$$

where $M = P_2 + P_1$ governs the movements within level 0.

Also, note that the process $\{(Q_k, J(t_k), t_k) : k \geq 0\}$ is a Semi Markov Process (SMP) with exponential sojourn times; furthermore, the mean sojourn time in state (n, j) is given by $1/\lambda$ if $j \in S_1 \cup S_3$, and by $1/(2\lambda)$ if $j \in S_2$.

Clearly, the evolution of the continuous time process defined by $Q(t)$ can be described by the continuous time QBD

$$\begin{bmatrix} B & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{8}$$

where

$$A_i = [diag(\lambda I, 2\lambda I, \lambda I)][P_i - \delta_{i1} I]$$

with δ_{i1} being 1 or 0 depending on whether $i = 1$ or not, and the three blocks of the diagonal matrix on the right side of the above equation are respectively of sizes $|S_i|$ $i = 1, 2, 3$. Finally, $B = A_2 + A_1$. This is clear by considering the instantaneous rates of transitions along with the probabilities of successive states.

We summarize the above discussion as a theorem.

Theorem 4. *The process* $\{(Q_k, J(t_k), t_k) : k \geq 0\}$ *is a semi-Markov process with exponential sojourn times. The mean sojourn time in state* (n, j) *is* $1/\lambda$ *if* $j \in S_1 \cup S_3$, *and it is* $1/(2\lambda)$ *if* $j \in S_2$. *The embedded Markov chain* $\{Q_k : k \geq 0\}$ *is a discrete time QBD with transition matrix given in Eq.* (7). *The continuous time process* $\{(Q(t), J(t)) : t \geq 0\}$ *is a continuous time QBD with infinitesimal generator given in Eq.* (8).

Remark 4. Obviously, there are many diflerent *QBD*s that one can associate with the homogeneous fluid model: (i) For instance, in the construction, instead of considering the queue size and phase to the right of the transition epochs, one may consider those to the left of such epochs and consider the resuting *QBD*; one such leads to results in the form obtained by Soares and Latouche[19] using other techniques. See Section 6. (ii) One may also consider the time reversed fluid process and work in terms of the *QBD*s that arise from it. The *QBD* in Ramaswami[17] corresponds to one such construction; we omit the details.

4.2. Steady State Queue Length

Here, we characterize (when it exists) the steady state distribution as $k \to \infty$ of the embedded queue length and phase processes $(Q(t_k), J(t_k))$. This distribution will be used later to derive the steady state distribution of the fluid flow model.

Define π to be the steady state probability vector of the CTMC of phases—i.e., the unique vector satisfying $\pi T = \mathbf{0}$, $\pi \mathbf{1} = 1$, where $\mathbf{1}$ is a column vector of l's, and partition it as $\pi = (\pi_1, \pi_2, \pi_3)$ corresponding to the sets $S_i, i = 1, 2, 3$. The necessary and sufficient condition for stability of the fluid flow is well-known—namely,

$$\pi_1 \mathbf{1} < \pi_2 \mathbf{1}. \tag{10}$$

In the sequel, we assume this to be the case.

Theorem 5. *The fluid model is stable iff the queue* $\mathcal{Q}$ *is stable.*

Proof. Let

$$\eta = (\eta_1, \eta_2, \eta_3) = (\pi_1, 2\pi_2, \pi_3).$$

One can verify easily that $\pi T = \mathbf{0}$ is equivalent to the condition that

$$\eta = \eta(P_0 + P_1 + P_2).$$

Now, the inequality (10) is equivalent to $\eta P_2 \mathbf{1} > \eta P_0 \mathbf{1}$ which is the necessary and sufficient condition for the stability of the QBD given by Eq. (7). Thus, $\mathcal{Q}$ is stable iff $\mathcal{F}$ is stable. □

Theorem 6. *Let* $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots)$, *where the partitioning corresponds to different values* $0, 1, \ldots$ *of the queue length (level), denote the steady state distribution as* $k \to \infty$ *of the QBD of Eq.* (7) *at the embedded epochs* t_k. *Then* $\mathbf{x}_n = \mathbf{x}_0 R^n$, *where* R *is the minimal nonnegative solution of the quadratic equation*

$$R = P_0 + RP_1 + R^2 P_2. \tag{11}$$

The matrix R *has the structure*

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

Proof. The matrix-geometric structure of $\mathbf{x}$ as asserted is well-known; see Refs.[11,14]. It is trivial to prove that if any row of P_0 is zero, then the corresponding row of R is also zero; examine the simple iterative scheme for R using successive substitutions in Eq. (11) starting with P_0 and note that each iterate has a zero row corresponding to each zero row of P_0. □

Methods to compute R and $\mathbf{x}_0$ are well-known.[11] Since we will not need the value of $\mathbf{x}_0$ explicitly, we shall not discuss the details of computing it. In the matrix-geometric literature, there are two ways to compute the R matrix of a QBD. Both express R in terms of a G-matrix whose computation has greater advantages over iteratively computing R directly (see Ref.[11], Chapter 8). These are considered in the next two theorems.

Theorem 7. *Let* Δ *denote the diagonal matrix formed by the steady state vector of the stochastic matrix* $P_0 + P_1 + P_2$, *and prime denote matrix transposition. The matrix* R *which is the minimal nonnegative solution of Eq.* (11) *is given by* $R = \Delta^{-1}\tilde{G}'\Delta$, *where* $\tilde{G}$ *is the minimal nonnegative solution of the equation*

$$\tilde{G} = \tilde{P}_2 + \tilde{P}_1\tilde{G} + \tilde{P}_0\tilde{G}^2, \tag{13}$$

where $\tilde{P}_i = \Delta^{-1}P'_{2-i}\Delta$ *are nonnegative and sum to a stochastic matrix. The matrix* $\tilde{G}$ *is substochastic and has the structure*

$$\tilde{G} = \begin{bmatrix} \tilde{G}_{11} & 0 & 0 \\ \tilde{G}_{21} & 0 & 0 \\ \tilde{G}_{31} & 0 & 0 \end{bmatrix}. \tag{14}$$

Proof. The results follow from the duality theory (see Refs.[5,16]) and amount to a computation of R in terms of the time reversed duals of the processes under consideration. The substochasticity of $\tilde{G}$ follows from the fact that since the given model is stable, the dual is unstable. The structure of $\tilde{G}$ follows since if any column of $\tilde{P}_2$ is zero, then so is the corresponding column of $\tilde{G}$ as seen from successive substitutions in Eq. (13) starting with $\tilde{P}_2$. □

Remark 5. The above theorem provides a way to compute R via the G-matrix of the time reversed dual which is denoted here by the symbol $\tilde{G}$. The computation of $\tilde{G}$ can be done efficiently using the logarithmic reduction algorithm.[12] This approach is the analog of the one developed in Ref.[17], but using the *QBD* considered in this paper. The matrix $\tilde{G}$ can be related to the busy period of the time reversed fluid flow corresponding to the homogenized fluid model under consideration. The details are routine and omitted; see Ref.[17] for similar results.

Another approach for computing R is given by the following result which leads to an approach similar to that adopted in Ref.[19], but using the *QBD* constructed in this paper. It exploits the connection between the R matrix and the G matrix of a *QBD* and avoids time reversal arguments entirely. One can interpret the matrix G below as related to the busy period of the given homogenized fluid model itself; we omit the details which are routine; see Ref.[19] for similar results.

Theorem 8. *The matrix R which is the minimal solution of Eq.* (11) *is given by*

$$R = P_0[I - P_1 - P_0\hat{G}]^{-1}, \tag{15}$$

where the stochastic matrix $\hat{G}$ *is the minimal nonnegative solution of the equation*

$$\hat{G} = P_2 + P_1\hat{G} + P_0\hat{G}^2 \tag{16}$$

and has the structure

$$\hat{G} = \begin{bmatrix} 0 & \hat{G}_{12} & 0 \\ 0 & \hat{G}_{22} & 0 \\ 0 & \hat{G}_{32} & 0 \end{bmatrix}. \tag{17}$$

Proof. The relationship between R and $\hat{G}$ is a familiar relation obtained by G. Latouche.[10] The stochasticity of $\hat{G}$ is due to the fact that the discrete time *QBD* is stable, and its special structure is a result of the structure of the matrix P_2. □

Remark 6. The relationship between R and G exploited here is special to the *QBD* and does not extend to more general models. However, the duality results presented earlier do extend in a natural manner.[16] Thus, the approach using the dual allows one to analyze much more general fluid flow models (for instance those with jumps) than the ones considered here. We omit the details as it would constitute a major digression.

Before we conclude this section, we present one more result which is needed in the next section. It is obtained by invoking a well-known result[7] that expresses the mean return time of a state in a Markov renewal process in terms of the steady state probabilities of the embedded Markov chain and the "fundamental mean" c of the process. The quantity c is defined as the steady state expected length of a sojourn interval. One can obtain c as the weighted average of the mean sojourn times of individual states weighted by their steady state probabilities in $\mathbf{x}$, but we will have no need for its explicit value.

Theorem 9. *Let $m(n,j)$ denote the mean return time of the state (n,j) in the SMP considered in Theorem 4. We have $m(n,j) = c/x(n,j)$, where c is the fundamental mean of the semi-Markov process.*

5. STEADY STATE RESULTS

We first prove the following important result which explains our interest in the queueing model constructed by us.

Theorem 10. *The process $\{(F(t), J(t)) : t \geq 0\}$ has a steady state distribution as $t \longrightarrow \infty$ iff the process $\{(Y(t), J(t)) : t \geq 0\}$ has a steady state distribution as $t \longrightarrow \infty$, and when they exist both distributions are the same.*

Proof. That the existence or otherwise of the stationary distributions is simultaneous for both processes is immediate from Theorem 5. It is obvious from the construction that the Poisson process $\mathcal{M}\oplus\mathcal{N}$ is such that $\{[M(t+s)+N(t+s)] - [M(t)+N(t)] : s \geq 0\}$ is independent of $\{(F(u-), J(u-)) : 0 \leq u \leq t\}$ as well as the process $\{(Y(u-), J(u-)) : 0 \leq u \leq t\}$; the latter depend only on the history of the superposition in the interval $[0, t)$. Thus, one can use the PASTA (Poisson Arrivals see Time Averages) theorem[9,20] to assert the following two equations:

$$lim_{t\to\infty} P[F(t) \leq x,\ \ J(t) = j] = lim_{k\to\infty} P[F(s_k-) \leq x,\ \ J(s_k-) = j];$$

$$lim_{t\to\infty} P[Y(t) \leq x,\ \ J(t) = j] = lim_{k\to\infty} P[Y(s_k-) \leq x,\ \ J(s_k-) = j].$$

The equality of the stationary distribution now follows from the equidistribution of the pairs $(F(s_k-), J(s_k-))$ and $(Y(s_k-), J(s_k-))$ for each k. □

Now, using the steady state queue length distribution obtained in the previous section, we can obtain the steady state joint distribution of $(Y(t), J(t))$ as $t \longrightarrow \infty$, thereby obtaining the steady state distribution of the homogenized fluid model.

5.1. Steady State Results for $(Y(t), J(t))$

Let $V_j(x)$ denote the steady state probability as $t \longrightarrow \infty$ that $Y^{(1)}(t) \leq x$ and the phase $J(t)$ is j. For $x > 0$, $v_j(x) = \frac{d}{dx} V_j(x)$ is the steady state joint density; its existence is established inter alia in our derivation. In the following, we let $\mathbf{V}(x)$ and $\mathbf{v}(x)$ denote respectively the vectors with elements $V_j(x), j \in S$ and $v_j(x), j \in S$. Also, assumed is that in accordance with the partition of the state space, these have also been partitioned; that is, $\mathbf{V}(x) = (\mathbf{V}_1(x),\ \mathbf{V}_2(x),\ \mathbf{V}_3(x))$, and $\mathbf{v}(x) = (\mathbf{v}_1(x), \mathbf{v}_2(x), \mathbf{v}_3(x))$. Likewise, we also partition the vector of steady state probabilities $\mathbf{x}_n$, corresponding to queue length n as $(\mathbf{x}_{n1,}\mathbf{x}_{n2,}\mathbf{x}_{n3,})$.

5.1.1. Steady State Density in the Set S_1

Let us define $\rho_{(0,i)}^{(n,j)}(du)$ to be the element in the Markov renewal kernel (see Ref.[7]) associated with the semi-Markov process $(Q(t), J(t))$ that is usually interpreted as the elementary probability of a transition by the SMP into the state (n, j) at u, given that it starts in $(0, i)$ at time 0. Conditioning on the last epoch of jump of the SMP t_k before time t, we can write the joint transform given by

$$E_{(0,i)}[e^{-sY(t)}I(J(t) = j)] \tag{18}$$

in terms of the Markov renewal kernel. Indeed, we note the following facts:

(a) For the phase at t to be j, the phase entered at t_k must be j, and no additional transition from j should occur before time t.
(b) If $j \in S_1$, then $Y(t) = Y(t_k) + (t - t_k)$.
(c) If $Q_k = n$, i.e., the queue size at t_k is n, then $Y(t_k) = W(t_k)$ is distributed as the sum of n iid $exp(\lambda)$ random variables.

Piecing together the facts noted above, we can write for $j \in S_1$ the transform in Eq. (18) as given by

$$\sum_{n=0}^{\infty} \int_0^t \rho_{(0,i)}^{(n,j)}(du) \left(\frac{\lambda}{s+\lambda}\right)^n e^{-s(t-u)} P(t_{k+1} - t_k > t - u) \tag{19}$$

We note that the probability appearing in the above integral is given by $exp(-\lambda(t - u))$.

Now, for $j \in S_1$, taking the limit as $t \to \infty$ in Eq. (19) using the Key Renewal Theorem (see Ref.[7]), we get the limit of that transform as

$$\sum_{n=0}^{\infty} c^{-1} x(n, j) \int_0^{\infty} \left(\frac{\lambda}{s+\lambda}\right)^n e^{-st} e^{-\lambda t} dt = \sum_{n=0}^{\infty} c^{-1} x(n, j) \left(\frac{\lambda}{s+\lambda}\right)^n \frac{1}{s+\lambda} \tag{20}$$

Using the matrix-geometric result for $\mathbf{x}_n$ and the structure of R given in Theorem 6 and Eq. (12), which imply that $\mathbf{x}_{n1} = \mathbf{x}_{01} R_{11}^n$, we can write the quantities in Eq. (20) for $j \in S_1$ together in vector form as

$$c^{-1} \mathbf{x}_{01} \sum_{n=0}^{\infty} R_{11}^n \left(\frac{\lambda}{s+\lambda}\right)^n \frac{1}{s+\lambda}. \tag{21}$$

The above expression can be simplified to

$$c^{-1} \mathbf{x}_{01} [sI - \lambda(R_{11} - I)]^{-1}, \tag{22}$$

where the existence of the inverse follows from the fact that R_{11} has spectral radius less than unity.

That is, we can write the joint density

$$\mathbf{v}_1(x) = c^{-1}\mathbf{x}_{01}e^{Kx}, \tag{23}$$

where the matrix

$$K = \lambda[R_{11} - I]. \tag{24}$$

Integrating this over x should, of course, give us the vector π_1 which is comprised of the marginal steady state probabilities of the phases for the set S_1. Thus,

$$\pi_1 = -c^{-1}\mathbf{x}_{01}K^{-1}, \tag{25}$$

and we thus have the steady state density vector of the fluid level in the set S_1 to be given by

$$\mathbf{v}_1(x) = -\pi_1 K e^{Kx}, \quad x > 0, \tag{26}$$

which is in the form obtained earlier.[17]

Incidentally, our discussion also shows that

$$\lim_{t\to\infty} P[Y(t) = 0, J(t) = j] = 0 \quad \text{for } j \in S_1. \tag{27}$$

We now invoke Theorem (10) and summarize the discussion as a theorem.

Theorem 11. *For $j \in S_1$, let $[\mathbf{V}_1(x)]j$ denote the steady state probability*

$$lim_{t\to\infty} P_{0i}[F(t) \le x,\ \ J(t) = j],$$

and let $v_j(x)$ denote its density. The vector $\mathbf{v}_1(x)$ with these elements is given by

$$\mathbf{v}_1(x) = -\pi_1 K e^{Kx}, \quad x > 0, \tag{28}$$

where $K = \lambda(R_{11} - I)$. We also have

$$\mathbf{V}_1(0) = \mathbf{0}, \tag{29}$$

so that this distribution has no mass at zero.

5.1.2. Steady State Density in the Set S_2

Let us now obtain the steady state distribution of $F(\cdot)$ when the phase is in S_2.

Theorem 12. *The steady state density of $F(\cdot)$ with phase in S_2* is given by

$$\mathbf{v}_2(x) = \mathbf{v}_1(x)\Psi, \quad x > 0, \tag{30}$$

where $\Psi = \frac{1}{2}R_{12}$.

Proof. We invoke Theorem (10) and evaluate the density for the process $\mathcal{Y}$. By a routine argument conditioning on the last epoch t_k before time t, we can write for $i \in S_1, j \in S_2$,

$$[\mathbf{v}_2(x)]_j = \lim_{t\to\infty}\sum_{n=1}^{\infty}\int_0^t \rho_{(0,i)}^{(n,j)}(du)e^{-2\lambda(t-u)}e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}\lambda. \tag{31}$$

The above is obtained by noting that for $Y(t)$ to be in $(x, x+dx)$ and the phase at t to be $j \in S_2$, at the last epoch of transition u before time t, there should be some $n > 0$ customers in the system, there should be no further transitions in $(u, t]$, and the total remaining work at t of those n customers (which is distributed as a sum of n exponential random variables) should be in $(x, x+dx)$.

Taking the limit as $t \to \infty$ in the above and writing in vector form, we get

$$\mathbf{v}_2(x) = c^{-1}\sum_{n=1}^{\infty}\mathbf{x}_{n2}\int_0^{\infty}e^{-2\lambda t}dt\, e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}\lambda$$

$$= c^{-1}\sum_{n=1}^{\infty}\mathbf{x}_{01}R_{11}^{n-1}R_{12}\frac{1}{2}e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!} = c^{-1}\mathbf{x}_{01}e^{Kx}\frac{1}{2}R_{12} = \mathbf{v}_1(x)\Psi. \tag{32}$$

In the above, the second equality is obtained using the matrix-geometric structure of $\mathbf{x}_n$ and the structure of R by which $\mathbf{x}_{n2} = \mathbf{x}_{01}R_{11}^{n-1}R_{12}$, and the last equality is obtained by using Eqs. (24) and (23). □

Corollary 1. *The probability that the fluid buffer is empty and the phase is in S_2 is given by the vector*

$$\mathbf{V}_2(0) = \pi_2 - \pi_1\Psi. \tag{33}$$

Proof. Note that π_2 is the steady state probability for the set S_2. Subtracting from this the integral of $\mathbf{v}_2(x)$ in $(0, \infty)$ obviously should give the required vector of emptiness probabilities for S_2. Hence the result. □

5.1.3. Steady State Density in the Set S_3

Now, for $j \in S_3$, we have

$$P_{(0,i)}[0 < Y(t) \leqq x, J(t) = j]$$

$$= \int_0^t\sum_{n=1}^{\infty}\rho_{(0,i)}^{(n,j)}(du)e^{-\lambda(t-u)}\int_0^x\frac{\lambda^n}{(n-1)!}e^{-\lambda y}y^{n-1}dy. \tag{34}$$

This follows from the fact that for the said event to occur, at the time of the last epoch u of transition of the SMP, the phase should be j, no transition should occur in $(u, t]$ and the fluid level at $u+$ should be at most x; we have used the fact that in S_3, no change occurs to the fluid level.

Now taking the limit as $t \to \infty$ in the above and writing the result in vector notation, we get

$$\mathbf{V}_3(x) - \mathbf{V}_3(0) = c^{-1}\sum_{n=1}^{\infty} \mathbf{x}_{n3} \int_0^{\infty} e^{-\lambda t} \int_0^x \frac{\lambda^n}{(n-1)!} e^{-\lambda y} y^{n-1} dy\, dt. \tag{35}$$

We now use the matrix-geometric result and the structure of R which imply that $\mathbf{x}_{n3} = \mathbf{x}_{01} R_{11}^{n-1} R_{13}$ and simplify the above formula to

$$\mathbf{V}_3(x) - \mathbf{V}_3(0) = c^{-1}\mathbf{x}_{01} \int_0^x e^{Ky} dy R_{13} \tag{36}$$

differentiating which with respect to x we get for the density function

$$\mathbf{v}_3(x) = \mathbf{v}_1(x) R_{13}, \quad x > 0. \tag{37}$$

We know that the steady state probability of the phase being in S_3 is π_3. Subtracting the integral of $\mathbf{v}_3(x)$ over $(0, \infty)$ from this, we immediately get the steady state emptiness probability in the set S_3 to be

$$\mathbf{V}_3(0) = \pi_3 - \pi_1 R_{13.} \tag{38}$$

Lemma 1. *The matrix R_{13} is given by*

$$R_{13} = (T_{13} + \Psi T_{23})(-T_{33})^{-1}. \tag{39}$$

Proof. We have from Eq. (11),

$$R_{13} = P_{13} + \frac{1}{2} R_{12} P_{23} + R_{13} P_{33} = \left[P_{13} + \frac{1}{2} R_{12} P_{23}\right](I - P_{33})^{-1}. \tag{40}$$

The result follows by writing the matrices P_{ij} appearing on the right in terms of the corresponding T_{ij}, and substituting Ψ in place of $\frac{1}{2}R_{12}$. □

We summarize the above discussion as a theorem.

Theorem 13. *For $x > 0$ the density of the fluid level for the set S_3 is given by*

$$\mathbf{v}_3(x) = \mathbf{v}_1(x)\Theta, \tag{41}$$

where

$$\Theta = (T_{13} + \Psi T_{23})(-T_{33})^{-1}. \tag{42}$$

We also have

$$\mathbf{V}_3(0) = \pi_3 - \pi_1\Theta. \tag{43}$$

6. COMPUTATIONAL SIMPLIFICATION

The key quantities needed in the computation of the steady state fluid flow distribution are the matrices K, Ψ and Θ. We have already noted in Eq. (42) that Θ can be expressed in terms of Ψ and the matrix T. The following simple result shows that K can be written in terms of Ψ and Θ so much so that the only non-trivial computation needed is that of the matrix Ψ. After noting this, we also provide some results that further simplify the computation of Ψ.

Theorem 14. *The matrix K is given by*

$$K = T_{11} + \Psi T_{21} + \Theta T_{31}. \tag{44}$$

Proof. Using the partitioned structures of R and P_i in Eq. (11) and (4)–(6), one can obtain the following:

$$R_{11} = P_{11} + \Psi P_{21} + R_{13}P_{31}.$$

Substituting $K = \lambda(R_{11} - I)$ and $\Theta = R_{13}$ and substituting for P_{ij} their values in terms of the corresponding T_{ij}, the result is immediate. □

In constructing the embedded Markov chain at epochs t_k, we noted that we could have also used the states to the left of t_k. Considering the states to the left of the epochs t_k, $k \geq 1$ leads to the consideration of the discrete time QBD in which the up, within level, and downward in level transitions are governed respectively by the matrices:

$$\breve{B}_0 = \begin{bmatrix} P_{11} & 0 & 0 \\ \frac{1}{2}P_{21} & 0 & 0 \\ P_{31} & 0 & 0 \end{bmatrix}, \quad \breve{B}_1 = \begin{bmatrix} 0 & P_{12} & P_{13} \\ 0 & \frac{1}{2}P_{22} & \frac{1}{2}P_{23} \\ 0 & P_{32} & P_{33} \end{bmatrix}, \quad \breve{B}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2}I & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{45}$$

For that QBD, the matrix $\breve{G}$ representing the downward passage probabilities from level 1 to level 0 is the minimal nonnegative solution of the equation

$$\breve{G} = \breve{B}_2 + \breve{B}_1\breve{G} + \breve{B}_0\breve{G}^2. \tag{46}$$

Because of the structure of $\breve{B}_2$, we also have the partitioned structure

$$\breve{G} = \begin{bmatrix} 0 & G_{12} & 0 \\ 0 & G_{22} & 0 \\ 0 & G_{32} & 0 \end{bmatrix}. \tag{47}$$

We can now deduce the following result which helps us derive a simpler procedure to compute Ψ.

Theorem 15. *The matrix Ψ is also given by $\Psi = G_{12}$, where G_{12} is the submatrix of $\breve{G}$ corresponding to row indices in S_1 and column indices in S_2.*

Proof. By the results in Ref.[15], there exists a kernel $\hat{R}(du)$ such that $R = \hat{R}(\infty)$, and furthermore $[\hat{R}(du)]_{ij}$ is the elementary probability that the SMP $((Q(t_k), J(t_k), t_k : k \geq 0\}$ visits the state $(1, j)$ in $(u, u + du)$ avoiding level 0 given that it starts in state $(0, i)$. Now, the matrix

$$\Psi = \frac{1}{2}R_{12} = \int_0^\infty \hat{R}_{12}(du)\int_0^\infty 2\lambda e^{-2\lambda x}dx\frac{1}{2}I.$$

The integral on the right clearly records the probability of the queue emptying out from various phases in S_2 given that the semi-Markov process starts a busy cycle in state $(0, i)$, $i \in S_1$; the queue must reach the state $(1, j)$ in its first busy period, and then empty out from that phase in one step. Thus, Ψ records the probability that the process $\{(Q(t_k-), J(t_k-)) : k \geq 2\}$ visits $(0, j)$ when it visits level 0 for the first time given that $(Q(t_1-), J(t_1-)) = (0, i)$. Therefore, $\Psi = G_{12}$ and hence the result. □

The next result allows one to work with matrices of smaller orders when iteratively computing Ψ. The resulting formula for Ψ is similar to that obtained by Soares and Latouche.[19]

Theorem 16. *The matrix*

$$G = \begin{bmatrix} 0 & G_{12} \\ 0 & G_{22} \end{bmatrix}$$

is the minimal nonnegative solution to the equation

$$G = B_2 + B_1G + B_0G^2, \tag{48}$$

where

$$B_2 = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2}I \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 & P_{12} + P_{13}(I - P_{33})^{-1}P_{32} \\ 0 & \frac{1}{2}[P_{22} + P_{23}(I - P_{33})^{-1}P_{32}] \end{bmatrix},$$

$$B_0 = \begin{bmatrix} P_{11} + P_{13}(I - P_{33})^{-1}P_{31} & 0 \\ \frac{1}{2}[P_{21} + P_{23}(I - P_{33})^{-1}P_{32}] & 0 \end{bmatrix}.$$

Proof. Considering the QBD of Theorem 15 restricted to the set $S_1 \cup S_2$, we get the QBD governed by B_i, $i = 0, 1, 2$. That the G-matrix of the censored QBD is given by the matrix G defined above is trivial to see as well. □

7. PHASE TYPE REPRESENTATION

So far, we have worked with the homogeneous fluid flow and we need to translate our results to the original model given by Q and C; we shall briefly refer to that model as "the Q-model." In the sequel, we shall assume that the diagonal matrix C is also partitioned according to the state space as $C = diag(C_{11}, C_{22}, I)$ where each block is also diagonal.

The following result concerning steady state probabilities is a trivial consequence of the relation $T = C^{-1}Q$.

Theorem 17. *Let ξ denote the steady state probability vector corresponding to the infinitesimal generator Q. Then*

$$\xi = \frac{1}{\pi C^{-1}\mathbf{1}} \pi C^{-1}, \tag{49}$$

where, π is the steady state probability vector of T.

Now, let $\mathbf{W}_{\mathrm{i}}(x)$ and $\mathbf{w}_{\mathrm{i}}(x)$ denote the vectors in the Q-model that are analogous to $\mathbf{V}_{\mathrm{i}}(x)$ and $\mathbf{v}_{\mathrm{i}}(x)$.

A standard argument considering the epochs t and $t + \Delta t$ leads to a set of partial differential equations for the time dependent state probabilities of the fluid model, from which one gets by letting $t \longrightarrow \infty$ the following equations for the steady state densities $\mathbf{w}_i$:

$$\mathbf{w}_1(x)Q_{11} + \mathbf{w}_2(x)Q_{21} + \mathbf{w}_3(x)Q_{31} = \mathbf{w}'_1(x)C_{11}$$

$$\mathbf{w}_1(x)Q_{12} + \mathbf{w}_2(x)Q_{22} + \mathbf{w}_3(x)Q_{32} = -\mathbf{w}'_2(x)C_{22}$$

$$\mathbf{w}_1(x)Q_{13} + \mathbf{w}_2(x)Q_{23} + \mathbf{w}_3(x)Q_{33} = 0$$

Comparing the above with the analogous equations for the homogeneous fluid flow, we can immediately deduce that

$$\mathbf{w}_i(x) = \eta \mathbf{v}_i(x) C_{ii}^{-1} \quad \text{for } \; i = 1, 2, 3, \tag{50}$$

for some constant η.

Thus, we can write in partitioned form

$$\mathbf{w}(x) = \eta \mathbf{v}_1(x)[C_{11}^{-1} \vdots \Psi C_{22}^{-1} \vdots \Theta C_{33}^{-1}] = -\eta \pi_1 K e^{Kx}[C_{11}^{-1} \vdots \Psi C_{22}^{-1} \vdots \Theta C_{33}^{-1}]$$

$$= -\eta \pi_1 K e^{Kx} C_{11}^{-1}[I \vdots C_{11}\Psi C_{22}^{-1} \vdots C_{11}\Theta C_{33}^{-1}] = -\eta \pi_1 C_{11}^{-1} \tilde{K} e^{\tilde{K}x}[I \vdots \tilde{\Psi} \vdots \tilde{\Theta}],$$

$$= -\eta(\pi C^{-1}\mathbf{1})\xi_1 \tilde{K} e^{\tilde{K}x}[I \vdots \tilde{\Psi} \vdots \tilde{\Theta}],$$

where

$$\tilde{K} = C_{11}KC_{11}^{-1}, \quad \tilde{\Psi} = C_{11}\Psi C_{22}^{-1}, \quad \tilde{\Theta} = C_{11}\Theta C_{33}^{-1}, \tag{51}$$

are the matrices analogous to K, Ψ and Θ for the Q-model.

Now, noting the boundary conditions

$$\int_0^\infty \mathbf{w}_1(x)dx = \xi_1, \quad \int_0^\infty \mathbf{v}_1(x)dx = \pi_1,$$

we immediately get, using Eq. (50), the normalizing constant η as

$$\eta = 1/(\pi C^{-1}\mathbf{1}).$$

Substituting this in the expression obtained for $\mathbf{w}(x)$ gives the following result.

Theorem 18. *The joint density of the steady state buffer content and phase in the fluid flow model governed by the pair* (Q, C) *is given by the vector*

$$\mathbf{w}(x) = -\xi_1\tilde{K}e^{\tilde{K}x}[I\vdots\tilde{\Psi}\vdots\tilde{\Theta}], \tag{52}$$

where the matrices $\tilde{K}$, $\tilde{\Psi}$ *and* $\tilde{\Theta}$ *are given in Eq.* (51).

The above results completely characterize the steady state distribution for the Q-model.

Computational Steps

Although we went through a number of steps in deriving our results, the steps to effect the computation of the quantities needed to determine the steady state fluid flow distribution are themselves fairly straightforward. So that one may not lose sight of this important fact in the myriad set of intermediate results, we list those key steps below:

Step 1: Compute the matrix $T = C^{-1}Q$, where C is the diagonal matrix formed by the vector of absolute values of the rates of change of the fluid flow in S_1 and S_2 padded by a vector of 1's of order $|S_3|$.

Step 2: Let $\lambda = max(-T_{ii})$. Compute the matrix $P = \lambda^{-1}T + I$. This determines all the needed submatrices P_{ij}.

Step 3: Obtain the steady state probability vector π of the irreducible stochastic matrix P. One may use a simple procedure such as the GTH-algorithm (see Refs.[8,11]) to effect this. Also, compute ξ using Eq. (49).

Step 4: Compute the matrix $\Psi = G_{12}$ by computing the matrix G of Eq. (48). One may use the algorithm in Ref.[12] to do this.

Step 5: Compute the matrix Θ using Eq. (42).

Step 6: Compute the matrix K using Eq. (44).

Step 7: Compute $\tilde{K}$, $\tilde{\Psi}$ and $\tilde{\Theta}$ using Eq. (51).

Once the above quantities are determined, we have a complete characterization of the steady state probabilities of the Q-model through Eq. (52).

The actual numerical computation of the steady state probability density or cdf can be accomplished by using the available techniques for phase type distributions (see Ref.[11], Chapter 2) since we can show that the distribution of the fluid flow is indeed a PH-distribution.

Theorem 19. *For the fluid model given by (Q, C), the steady state distribution of the fluid level is a phase type distribution. One representation for that phase type distribution is given by the pair (β, U), where*

$$\beta = (\mathbf{1} + \tilde{\Psi}\mathbf{1} + \tilde{\Theta}\mathbf{1})'\tilde{\Delta}_1,$$

$U = \tilde{\Delta}_1^{-1}\tilde{K}'\tilde{\Delta}_1$, *the matrix $\tilde{\Delta}_1$, is a diagonal matrix with ξ_1 on its diagonal, and $\mathbf{1}$ is a vector of $1'$s of appropriate order determined by the context in which it appears.*

Proof. The proof proceeds by using the duality arguments similar to those in Ref.[16]. We write the steady state fluid density

$$f(x) = \sum_{j=1}^{3} \mathbf{w}_j(x)\mathbf{1}$$

as also given by

$$f(x) = -\xi_1\tilde{K}e^{\tilde{K}x}(\mathbf{1} + \tilde{\Psi}\mathbf{1} + \tilde{\Theta}\mathbf{1}) = (\mathbf{1}' + \mathbf{1}'\tilde{\Theta}' + \mathbf{1}'\tilde{\Psi}')(e^{\tilde{K}x})'\tilde{K}'\xi_1{}' = \beta e^{Ux}(-U\mathbf{1}).$$

The second equality above is obtained by noting that the transpose of a scalar is itself, and the last is obtained by supplying the necessary products $\tilde{\Delta}_1\tilde{\Delta}_1^{-1} = \tilde{\Delta}_1^{-1}\tilde{\Delta}_1 = I$ in between the terms of the previous equation. The fact that U satisfies the condition for a generator can be verified by considering the duality formula for $\tilde{K}$; we omit the details which are routine; like in Ref.[17], it is easy to show that for the time reversed dual flow of the Q-flow, e^{Ux} indeed records the probability of the busy period ending in various phases given the phase at the beginning of that busy period. □

Remark 7. There is a discrepancy between our results here and those reported in Ref.[19] due to a minor error which crept into that paper. Specifically in Ref.[19], a term $\xi\mathbf{1}$ appears superfluously in the Eq. (4) of that paper, and that error then gets propagated to Theorem 5.1 of that paper. We are grateful to Soares and Latouche for confirming the correctness of our findings.

ACKNOWLEDGMENT

The authors thank Ward Whitt for many helpful discussions, and Benny VanHoudt for identifying several typos in an earlier version. Soohan Ahn was supported (partially) by the Post-Doctoral Fellowship Program of Korea Science & Engineering Foundation (KOSEF).

REFERENCES

1. Adan, I.J.B.F.; Resing, J.A.C. A two-level traffic shaper for an on-off source. Perform. Eval. **2000**, *42*, 279–298.
2. Ahn, S.; Ramaswami, V. Transient analysis of fluid flow models via stochastic coupling to a queue, 2003, to be submitted.
3. Anick, D.; Mitra, D.; Sondhi, M.M. Stochastic theory of a data handling system with multiple sources. Bell Syst. Tech. J. **1982**, *61*, 1871–1894.
4. Asmussen, S. Stationary distributions for fluid flow models with or without Brownian noise. Stochast. Models **1995**, *11*, 1–20.
5. Asmussen, S.; Ramaswami, V. Probabilistic interpretations of some duality results for the matrix paradigms in queueing theory. Stochast. Models **1990**, *6*, 715–734.
6. Choudhury, G.L.; Mandelbaum, A.; Reiman, M.I.; Whitt, W. Fluid and diffusion limits for queues in slowly changing environments. Stochast. Models **1997**, *13*, 121–146.
7. Çinlar, E. Markov renewal theory. Adv. Appl. Probab. **1969**, *1*, 123–187.
8. Grassmann, W.K.; Taksar, M.J.; Heyman, D.P. Regenerative analysis and steady state distributions for Markov chains. Oper. Res. **1985**, *33*, 1107–1116.
9. König, D.; Schmidt, V. Imbedded and non-imbedded stationary characteristics of queueing systems with varying service rate and point processes. J. Appl. Probab. **1980**, *17*, 1048–1061.
10. Latouche, G. A note on two matrices occurring in the solution of quasibirth-and-death processes. Stochast. Models **1987**, *3*, 251–257.
11. Latouche, G.; Ramaswami, V. *Introduction to matrix analytic methods in Stochastic Modeling*; SIAM & ASA: Philadelphia, 1999.
12. Latouche, G.; Ramaswami, V. A logarithmic reduction algorithm for Quasi-Birth-and-Death Processes. J. Appl. Probab. **1993**, *30*, 650–674.
13. Kella, O.; Whitt, W. A storage model with a two state random environment. Oper. Res. **1992**, *40* (2), S257–S262.
14. Neuts, M. F. *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, 1981.
15. Ramaswami, V. The busy period of queues which have a matrix-geometric steady state probability vector. Opsearch **1982**, *19*, 238–261.
16. Ramaswami, V. A duality theorem for the matrix paradigms in queueing theory. Stochast. Models **1990**, *6*, 151–161.
17. Ramaswami, V. Matrix analytic methods for stochastic fluid flows. In *Teletrafic Engineering in a Competitive World*; Smith, D., Key, P., Eds.; Proc. of the 15th International Teletraffic Congress; Elsevier, 1999; 1019–1030.
18. Ramaswami, V. Algorithmic analysis of stochastic models—The changing face of mathematics. In *Mathematics for the New Millennium*; Vijayakumar, A., Krishnakumar, S., Eds.; Ramanujan Memorial Lecture; Anna University: Chennai, India, 2000.

19. da Silva Soares, A.; Latouche, G. Further results on the similarities between fluid queues and QBDs. In *Matrix-Analytic Methods, Theory and Applications*; Latouche, G., Taylor, P.G., Eds.; Proc. of the Fourth International Conference on Matrix Analytic Methods; World Scientific Publishing Co.: Singapore, 2002.
20. Wolff, R.W. Poisson arrivals see time averages. Oper. Res. **1982**, *30*, 223–231.

Received May 10, 2002
Revised October 18, 2002
Accepted February 7, 2003