SECOND-ORDER FLUID FLOW MODELS: REFLECTED BROWNIAN MOTION IN A RANDOM ENVIRONMENT

RAJEEVA L. KARANDIKAR
Indian Statistical Institute, New Delhi, India

VIDYADHAR G. KULKARNI
University of North Carolina, Chapel Hill, North Carolina

This paper considers a stochastic fluid model of a buffer content process $\{X(t),\ t \geq 0\}$ that depends on a finite-state, continuous-time Markov process $\{Z(t),\ t \geq 0\}$ as follows: During the time-intervals when $Z(t)$ is in state i, $X(t)$ is a Brownian motion with drift μ_i, variance parameter σ_i^2 and a reflecting boundary at zero. This paper studies the steady-state analysis of the bivariate process $\{(X(t),\ Z(t)),\ t \geq 0\}$ in terms of the eigenvalues and eigenvectors of a nonlinear matrix system. Algorithms are developed to compute the steady-state distributions as well as moments. Numerical work is reported to show that the variance parameter has a dramatic effect on the buffer content process.

Fluid flow models have been used extensively in modeling high-speed communication networks (see Anick, Mitra and Sondhi 1982, Kosten 1986, Mitra 1988a, b, Elwalid and Mitra 1991, Elwalid, Mitra and Stern 1991, etc.) and as approximate models for queues (see Vandergraft 1983, Chen 1988, Stern and Elwalid 1991, Chen and Yao 1992, Kella and Whitt 1992, etc.). Typically, the fluid represents bits or packets of information and the fluid flow model describes the stochastic behavior of the fluid level in the buffer. The arrival process of the fluid into the buffer and the departure process from the buffer are both modulated by a random external environment. Typically, the environment is described by a finite-state, continuous-time Markov chain and the rate at which the fluid enters and exits the buffer is assumed to depend deterministically upon the state of the environment. We call such models the first-order fluid flow models.

In spite of their success as a modeling tool in high-speed networks, the fluid flow models have one drawback—they are "first-order" models. Hence, they work well when the system characteristics are adequately described through first moments. This is why they are so attractive for the $D|D|1$ queues in random environment, or the ATM networks where packets are of constant size and the sources behave in "on-off" fashion. They cannot account for the variability during an on-period, that is, the variance of the amount of fluid coming into or leaving the buffer during one environment state.

It is this limitation of fluid flow models that has prompted us to investigate the "second-order" models—models that take into account the first *and* second moments of system characteristics. Taking a cue from the diffusion approximations in queueing systems (see Heyman and Sobel 1982), we study diffusion processes whose drift and variance parameters are dependent upon an external random environment. Brownian motion has been used to model simple stochastic flow systems (see Harrison 1985). Hence, our model can be thought of as an extension of that work. When the variance parameter is zero in all states, this model reduces to the standard fluid flow model. Thus, the current work extends the first-order fluid flow models by addressing one of their limitations. London et al. (1982) studied first- and second-order fluid models (not quite the same setup as considered here, but related to it) via Wiener–Hopf factorization methods (see Kennedy and Williams 1990). This approach does not seem to be computationally tractable for the second-order case, and hence we do not discuss it here. Recently, we became aware of the independent work of Asmussen (1992) that substantially overlaps and complements ours. However, our approach is different.

The paper is organized as follows: The model and the notation is introduced in the next section. The main equations satisfied by the transient and steady-state probability distributions are derived in Section 2. Building upon the solution methods developed for first-order fluid models, Section 3 describes a spectral solution to these equations by using methods of eigenvalues and eigenvectors. Several important special cases are studied in Section 4. Section 5 contains several analytical and numerical examples. These examples show that using first-order fluid flow models (i.e., ignoring variability) can lead to serious underestimation of congestion. Hence, it is crucial to appropriately handle the variance components. Finally, Section 6 describes two important extensions and variations of this model to handle other practical boundary behaviors.

Subject classifications: Communication: fluid models using diffusion. Probability, diffusion: reflected Brownian motion, random environment.
Area of review: TELECOMMUNICATIONS (SPECIAL ISSUE ON TELECOMMUNICATION SYSTEMS: MODELING, ANALYSIS AND DESIGN).

1. MODELS

Consider an infinite capacity buffer where fluid enters and exits according to a random process. Assume that the behavior of the random process is influenced by an external environment. The dynamics of the environment is modeled by a finite-state, continuous-time Markov chain (CTMC) $\{Z(t),\ t \geqslant 0\}$ taking values in $E = \{1, 2, \ldots, m\}$. We assume that $\{Z(t),\ t \geqslant 0\}$ is an irreducible CTMC with a generator matrix $Q = (q_{ij})$. Let $X(t)$ denote the amount of fluid in the buffer at time t. We describe the stochastic behavior of the $\{X(t),\ t \geqslant 0\}$ process below.

Given that the environment stays in state i during the interval $[t, t + h)$, the increment $X(t + h) - X(t)$ in the $\{X(t),\ t \geqslant 0\}$ process over $[t, t + h)$ is assumed to be normally distributed with mean $\mu_i h$ and variance $\sigma_i^2 h$, and $X(t + h) - X(t)$ is independent of $\{(X(t), Z(t)), 0 \leqslant s \leqslant t\}$. In other words, while $Z(t)$ stays in state i, the $X(t)$ process is a Brownian motion with drift parameter μ_i and variance parameter σ_i^2, $i = 1, 2, \ldots, m$. Typical sample paths of the $\{X(t),\ t \geqslant 0\}$ and $\{Z(t),\ t \geqslant 0\}$ are shown in Figure 1. Since $X(t)$ is the amount of fluid in the buffer at time t, it cannot be negative. We model this by assuming that the state 0 is a reflecting barrier for the Brownian motion in each state of the environment.

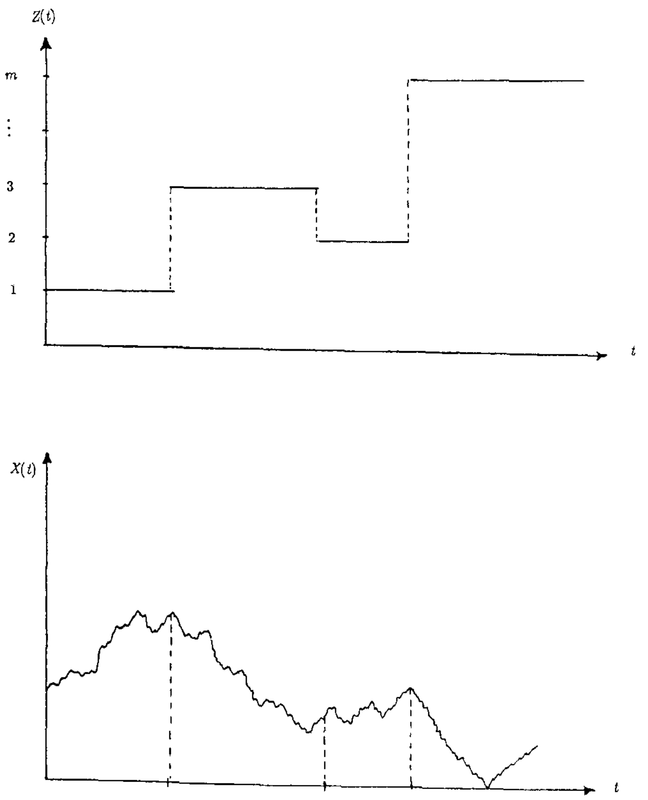

Figure 1. The sample paths $X(t)$ and $Z(t)$, where the drift and variance parameters of $X(t)$ change whenever $Z(t)$ changes.

Let $\pi = [\pi_i]$ be the limiting distribution of the $\{Z(t),\ t \geqslant 0\}$ process. Since we have assumed that $\{Z(t),\ t \geqslant 0\}$ is an irreducible CTMC, it is well known that π is the unique solution to

$$\pi Q = 0,\ \sum_i \pi_i = 1. \tag{1}$$

Let

$$d = \sum_{i=1}^{m} \pi_i \mu_i. \tag{2}$$

The quantity d is the mean drift of the $\{X(t),\ t \geqslant 0\}$ process when the $\{Z(t),\ t \geqslant 0\}$ process is in the steady state. The $\{(X(t), Z(t)),\ t \geqslant 0\}$ process has a limiting distribution if $d < 0$. (This is identical to the stability condition in the first-order models.) Hence, we assume that $d < 0$ when computing the limiting distribution.

2. ANALYSIS

From the description of the model given in the previous section it is clear that $\{(X(t), Z(t)),\ t \geqslant 0\}$ is a Markov process on the state-space $[0, \infty) \times E$. Since the reflecting behavior in environmental state j depends on whether $\sigma_j^2 = 0$, and if so, on the sign of μ_j, we introduce the following partition of E: $E_+ = \{j \in E: \sigma_j^2 > 0\}$, $E_0 = \{j \in E: \sigma_j^2 = 0,\ \mu_j = 0\}$, $E_{0+} = \{j \in E: \sigma_j^2 = 0,\ \mu_j > 0\}$, and $E_{0-} = \{j \in E: \sigma_j^2 = 0,\ \mu_j < 0\}$. Let m_+, m_0, m_{0+}, and m_{0-} be the cardinalities of E_+, E_0, E_{0+}, and E_{0-}, respectively. Finally, let $E_* = E_{0+} \cup E_{0-}$ and $m_* = m_{0+} + m_{0-}$.

Next we describe the generator of the $\{(X(t), Z(t)),\ t \geqslant 0\}$ process. Let f be a real function from $[0, \infty) \times E$ such that $f(\cdot, j)$ is twice continuously differentiable. Define the operator U_t for $t \geqslant 0$ by:

$$(U_t f)(x, j) = E(f(X(t), Z(t)) | X(0) = x, Z(0) = j). \tag{3}$$

The derivative L of U_t with respect to t at $t = 0$ is known as the generator of the Markov process, i.e.,

$$(Lf)(x, j) = \lim_{t \downarrow 0} \frac{1}{t} [(U_t f)(x, j) - f(x, j)]. \tag{4}$$

Following the standard analysis (see Karlin and Taylor 1981), the following expression for L can be derived:

$$\begin{aligned}(Lf)(x, j) &= \tfrac{1}{2}\sigma_j^2 f''(x, j) + \mu_j f'(x, j) \\ &\quad + \sum_{k:k \neq j} q_{jk}(f(x, k) - f(x, j)) \\ &= \tfrac{1}{2}\sigma_j^2 f''(x, j) + \mu_j f'(x, j) \\ &\quad + \sum_{k=1}^{m} q_{jk} f(x, k),\end{aligned} \tag{5}$$

where $f'(x, j)$ and $f''(x, j)$ are the first and second derivatives of $f(x, j)$ with respect to x. The last equality follows from $q_{jj} = -\sum_{k:k \neq j} q_{jk}$. In order that (5) be valid, f needs to be twice continuously differentiable in the x variable, with

$$f'(0, j) = 0 \quad \text{for } j \in E_+ \cup E_0 \cup E_{0-}. \tag{6}$$

The condition (6) arises because the X component of the (X, Z) process has a reflecting barrier at $x = 0$. In all states $j \in E_{0+} \cup E_0 \cup E_{0-}$ of the Z process, the state $x = 0$ acts as a reflecting barrier. We do not need (6) to hold for $j \in E_{0+}$, because in this case the boundary is never reached from the interior.

Using the generator defined by (5) and (6) we can now derive the equations satisfied by the distribution of $(X(t), Z(t))$. Let

$$F(t, x, j; y, i) = P(X(t) \leq x, Z(t) = j | X(0) = y, Z(0) = i) \tag{7}$$

be the joint distribution of $(X(t), Z(t))$. The analog of the Kolmogorov forward equation for F is

$$\frac{d}{dt}\left[\sum_j \int f(x, j) F(t, dx, j; y, i)\right] = \sum_j \int (Lf)(x, j) F(t, x, j; y, i), \tag{8}$$

where the operator L is defined by (5) and (6).

Intuitively, it is clear that $F(t, x, j; y, i)$ has a mass $C(t, j; y, i)$ (which may be zero) at $x = 0$ and a density $p(t, x, j; y, i)$ for $x > 0$. This can also be deduced from (8) with further analysis, which we do not report here. We mention that the first-order models also have this feature.

The next theorem gives the differential equations satisfied by the densities and the boundary conditions (we use p' and p'' to denote the first and second partial derivative of p with respect to x).

Theorem 1. *The densities $\{p(t, x, j; y, i)\}$ satisfy the following partial differential equations*

$$\frac{\partial}{\partial t} p(t, x, j; y, i) = \tfrac{1}{2}\sigma_j^2 p''(t, x, j; y, i) - \mu_j p'(t, x, j; y, i) + \sum_k p(t, x, k; y, i) q_{kj}, \; x > 0, \; t > 0 \tag{9}$$

with the boundary conditions

$$\frac{\partial}{\partial t} c(t, j; y, i) = \tfrac{1}{2}\sigma_j^2 p'(t, 0, j; y, i) - \mu_j p(t, 0, j; y, i) + \sum_k c(t, k; y, i) q_{kj} \quad \text{for all } j \tag{10}$$

and

$$c(t, j; y, i) = 0 \quad \text{for } j \in E_+ \cup E_{0+}. \tag{11}$$

Proof. See the Appendix.

Assuming that the Markov process $\{(X(t), Z(t))\}$ has a limiting distribution, define

$$p_j(x) = \lim_{t \to \infty} p(t, x, j; y, i) \tag{12}$$

and

$$c_j = \lim_{t \to \infty} c(t, j; y, i). \tag{13}$$

(Note that we have implicitly assumed that the limits do not depend upon the initial state (y, i)). The next theorem gives the equations satisfied by $\{p_j(x)\}$, $\{c_j\}$.

Theorem 2. *The limiting densities $\{p_j(x)\}$ satisfy*

$$\tfrac{1}{2}\sigma_j^2 p_j''(x) - \mu_j p_j'(x) + \sum_k p_k(x) q_{kj} = 0 \tag{14}$$

along with the boundary conditions

$$\tfrac{1}{2}\sigma_j^2 p_j'(0) - \mu_j p_j(0) + \sum_k c_k q_{kj} = 0 \tag{15}$$

and

$$c_k = 0 \quad \text{for } k \in E_+ \cup E_{0+}. \tag{16}$$

Proof. This follows by letting $t \to \infty$ in (9), (10), and (11).

In the next section, we examine the spectral representation of the solution to (9) satisfying (10) and (11).

3. LIMITING DISTRIBUTIONS

In this section we use Theorem 2 to solve for $\{p_j(x)\}$ and $\{c_j\}$. It is convenient to introduce the following matrix notation:

$$M = \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & 0 & \\ & 0 & \ddots & \\ & & & \mu_m \end{bmatrix}, \tag{17}$$

$$S = \begin{bmatrix} \tfrac{1}{2}\sigma_1^2 & & & 0 \\ & \tfrac{1}{2}\sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \tfrac{1}{2}\sigma_m^2 \end{bmatrix}, \tag{18}$$

$$p(x) = [p_1(x), p_2(x), \ldots, p_m(x)], \tag{19}$$

and

$$c = [c_1, c_2, \ldots, c_m]. \tag{20}$$

With this notation, we can write (14) and (15) in matrix form as:

$$p''(x)S - p'(x)M + p(x)Q = 0 \tag{21}$$

$$p'(0)S - p(0)M + cQ = 0, \tag{22}$$

where $p'(x)$ and $p''(x)$ are vectors of first and second derivatives of $p(x)$, respectively.

From the general theory of linear differential equations we know that a solution to (21) is a linear combination of functions of the type

$$p(x) = e^{\lambda x}\phi, \tag{23}$$

where λ is a scalar and $\phi = [\phi_1, \phi_2, \ldots, \phi_m]$. Substituting in (21) we get

$$\phi[\lambda^2 S - \lambda M + Q] = 0. \tag{24}$$

Thus, λ must be a solution to

$$\det[\lambda^2 S - \lambda M + Q] = 0 \tag{25}$$

and ϕ satisfies (24). We say that λ is an eigenvalue and ϕ is an eigenvector of the system (24). The quadratic matrix polynomial (24) is similar to the one in Elwalid, Mitra and Stern.

We next show how one can compute the solutions λ to (25) and the corresponding vector ϕ in (24) by using standard matrix eigenvalues and eigenvectors. We need the following matrix notation for the results.

Partition the Q, M, and S matrices over E_+, E_0, and E_* as follows:

$$Q = \begin{bmatrix} Q_{++} & Q_{+0} & Q_{+*} \\ Q_{0+} & Q_{00} & Q_{0*} \\ Q_{*+} & Q_{*0} & Q_{**} \end{bmatrix}, \tag{26}$$

$$M = \begin{bmatrix} M_+ & & 0 \\ & M_0 & \\ 0 & & M_* \end{bmatrix}, \tag{27}$$

$$S = \begin{bmatrix} S_+ & & 0 \\ & S_0 & \\ 0 & & S_* \end{bmatrix}. \tag{28}$$

Note that if $E_0 = E$, then $X(t) = X(0)$ for all $t \geq 0$, and $Z(t)$ is a CTMC with rate matrix Q. Hence, we assume that E_0 is not the whole set E, so that Q_{00} is invertible. Also, notice that M_* is invertible. Now denote

$$\begin{aligned} R_{+*} &= Q_{+*} - Q_{+0}Q_{00}^{-1}Q_{0*}, \\ R_{++} &= Q_{++} - Q_{+0}Q_{00}^{-1}Q_{0+}, \\ R_{**} &= Q_{**} - Q_{*0}Q_{00}^{-1}Q_{0*}, \\ R_{*+} &= Q_{*+} - Q_{*0}Q_{00}^{-1}Q_{0+}. \end{aligned} \tag{29}$$

Finally, define

$$B = \begin{bmatrix} 0 & R_{+*}M_*^{-1} & -R_{++}S_+^{-1} \\ 0 & R_{**}M_*^{-1} & -R_{*+}S_+^{-1} \\ I_+ & 0 & M_+S_+^{-1} \end{bmatrix} \tag{30}$$

where I_+ is an identity matrix of size m_+.

Theorem 3

i. *Let λ be a solution to (25) if and only if λ is an eigenvalue of B of (30).*
ii. *If the pair (λ, ϕ) with $\phi = (\phi_+, \phi_0, \phi_*)$ satisfies (24), then*

$$\phi_0 = -\phi_+Q_{+0}Q_{00}^{-1} - \phi_*Q_{*0}Q_{00}^{-1} \tag{31}$$

and $(\phi_+, \phi_*, \lambda\phi_+)$ is an eigenvector of B corresponding to the eigenvalue λ.
iii. *Conversely, if (ϕ_+, ϕ_*, Ψ) is an eigenvector of B corresponding to the eigenvalue λ, then $\Psi = \lambda\phi_+$. Furthermore, (λ, ϕ) with $\phi = (\phi_+, \phi_0, \phi_*)$, where ϕ_0 is given by (31), satisfy (24).*

Proof. A pair (λ, ϕ) will satisfy (24), i.e., $\phi[\lambda^2 S - \lambda M + Q] = 0$ if and only if it satisfies

$$(\phi, \lambda\phi)\begin{bmatrix} 0 & -Q \\ I & M \end{bmatrix} = \lambda(\phi, \lambda\phi)\begin{bmatrix} I & 0 \\ 0 & S \end{bmatrix}. \tag{32}$$

Now partitioning over E_+, E_0, and E_* as before, write $\phi = (\phi_+, \phi_0, \phi_*)$. Straightforward algebra (using $M_0 = S_0 = 0$) shows that (λ, ϕ) satisfy (32) if and only if ϕ_0 satisfies (31), and

$$(\phi_+, \phi_*, \lambda\phi_+)B = \lambda(\phi_+, \phi_*, \lambda\phi_+). \tag{33}$$

Thus, if (λ, ϕ) satisfies (24), λ is an eigenvalue of B with eigenvector $(\phi_+, \phi_*, \lambda\phi_+)$. Conversely, if λ is an eigenvalue of B with eigenvector (ϕ_+, ϕ_*, Ψ), from the structure of B we get $\Psi = \lambda\phi_+$. The above argument now shows that (λ, ϕ) with $\phi = (\phi_+, \phi_0, \phi_*)$ satisfies (24).

The next theorem provides information about the placement of the eigenvalues λ of (24) in the complex plane when the drift d of (2) is negative.

Theorem 4. *Equation (25) has $n = 2m_+ + m_*$ number of solutions. Let n_+ and n_- be the number of solutions with positive and negative real parts and n_0 be the number of solutions which are zero. Then, under the condition $d < 0$,*

$$n_+ = m_+ + m_{0-} - 1, \; n_0 = 1 \text{ and } n_- = m_+ + m_{0+}.$$

Proof. See the Appendix.

Recall that we are dealing with the ergodic case, i.e., the case of $d < 0$. Index the $n = 2m_+ + m_*$ solutions $\lambda_1, \lambda_2, \ldots, \lambda_n$ in such a way that

$$\begin{aligned} &Re(\lambda_1) \leq \ldots \leq Re(\lambda_{n_-}) < Re(\lambda_{n_-+1}) \\ &\quad = 0 < Re(\lambda_{n_-+2}) \leq \ldots \leq Re(\lambda_n). \end{aligned} \tag{34}$$

We also suppose for the time being that the λ_i's are distinct. In this case, Theorem 3 implies that there is a unique vector $\phi(i)$ (up to a constant multiplier) that solves (24) for $\lambda = \lambda_i$.

Since (21) is a homogeneous system of equations, its general solution is given by

$$p(x) = \sum_{i=1}^{n} e^{\lambda_i x}\phi(i)a_i, \tag{35}$$

where $a_i (i = 1, 2, \ldots, n)$ are scalars. Since $p(x)$ has to be a density, we must have $a_i = 0$ if $Re(\lambda_i) \geq 0$, i.e., $a_i = 0$ for $i \geq n_- + 1$ (from the indexing of (34)). Hence, (35) reduces to

$$p(x) = \sum_{i=1}^{n_-} e^{\lambda_i x}\phi(i)a_i. \tag{36}$$

The n_- unknown scalars $a_1, a_2, \ldots, a_{n_-}$ are determined by the boundary conditions (22) and the normalizing condition

$$\sum_{j=1}^{m} c_j + \sum_{j=1}^{m} \int_0^{\infty} p_j(x)\, dx = 1. \qquad (37)$$

Alternatively, we can use

$$\lim_{t\to\infty} P(Z(t) = j) = \pi_j = c_j + \int_0^{\infty} p_j(x)\, dx. \qquad (38)$$

Both these methods yield the vector equation:

$$c - \sum_{i=1}^{n-} \frac{a_i}{\lambda_i} \phi(i) = \pi. \qquad (39)$$

These provide m equations for the m unknowns a_1, $a_2, \ldots, a_{n_-}$ and c_j for $j \in E_0 \cup E_{0-}$. (Note again that $c_j = 0$ for $j \notin E_0 \cup E_{0-}$.) The above discussion is summarized in the following theorem.

Theorem 5. *Let $\lambda_1, \lambda_2, \ldots, \lambda_{n_-}$ be the eigenvalues of B with negative real parts and assume that they are distinct. Let $(\phi_+(i), \phi_*(i), \lambda_i\phi_+(i))$ be the eigenvector of B corresponding to the eigenvalue λ_i. Let $\phi(i) = (\phi_+(i), \phi_0(i), \phi_*(i))$ where $\phi_0(i)$ is given by (31). Then, $\{(X(t), Z(t)), t \geq 0\}$ has a limiting distribution with density $p_j(x)$ over $x > 0$, and mass c_j at $(0, j)$, $j \in S$. $\{p_j(x), c_j\}$ are given by*

$$p(x) = \sum_{i=1}^{n-} a_i e^{\lambda_i x} \phi(i) \qquad (40)$$

where the constants $a_1, a_2, \ldots, a_{n_-}$ and $c = (c_1, \ldots, c_m)$ are given by the m equations

$$c - \sum_{i=1}^{n-} \frac{a_i}{\lambda_i} \phi(i) = \pi \qquad (41)$$

with $c_j = 0$ for $j \notin E_0 \cup E_{0-}$ and π satisfies (1).

The condition about the distinctness of the λ's can be relaxed. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_k$ are k distinct eigenvalues of B with negative real part. Suppose that $m_1, m_2, \ldots, m_k$ are their multiplicities. Under the assumption of $d < 0$ we must have $m_1 + m_2 + \ldots + m_k = n_-$. Then it is known that the general solution $\{p_j(x)\}$ has the form:

$$p(x) = \sum_{i=1}^{k} \sum_{r=1}^{m_i} \frac{(\lambda_i x)^{r-1}}{(r-1)!} e^{\lambda_i x} a_{ir} \phi_r(i). \qquad (42)$$

The vectors $\phi_r(i)$ ($i = 1, 2, \ldots, k$, $r = 1, 2, \ldots, m_i$) need to be derived from the Jordan form of B. (This part is computationally hard.) The constants a_{ir} and c_j for $j \in E_0 \cup E_{0-}$ can be obtained in a similar fashion to the distinct eigenvalues case. Another alternative when faced with multiple eigenvalues is to perturb the Q matrix slightly so that the eigenvalues are no longer repeated. However, this is a trial-and-error method and provides a solution to the perturbed problem. Continuity of eigenvalue systems implies that the perturbed solution will be close to the original solution if the perturbation is small.

The next theorem gives the expected values and the variances in steady state of the buffer content process. Let τ, s^2 are the first and second moments of the limiting distribution of $\{X(t), t \geq 0\}$. We have the following theorem.

Theorem 6

$$\tau = \sum_{j=1}^{m} \sum_{i=1}^{n-} \frac{a_i}{\lambda_i^2} \phi_j(i), \qquad (43)$$

$$s^2 = -2 \sum_{j=1}^{m} \sum_{i=1}^{n-} \frac{a_i}{\lambda_i^3} \phi_j(i). \qquad (44)$$

Proof. This follows by direct integrations of the steady-state distributions.

Note that this theorem is a consequence of the spectral representation and appears in various papers, e.g., Anick, Mitra and Sondhi.

4. SPECIAL CASES

In this section we consider several special cases of the model studied in the previous sections.

$\sigma_j^2 = 0$ for all $j \in E$: The Fluid Flow Case

In this case, $E_+ = \phi$, and we get the standard first-order fluid flow model. The S matrix is identically zero and (21) and (22) reduce to

$$p'(x)M = p(x)Q \qquad (45)$$

$$p(0)M = cQ. \qquad (46)$$

Following (23)–(25) we see that the general solution is of the form

$$p(x) = e^{\lambda x}\phi, \qquad (47)$$

where the pair (λ, ϕ) satisfies

$$\phi Q = \lambda \phi M \qquad (48)$$

and

$$\det[Q - \lambda M] = 0. \qquad (49)$$

Theorem 4 then implies that there are m_* solutions (λ, ϕ) satisfying (48) and (49) out of which $m_{0-} - 1$ have $Re(\lambda) > 0$, one is zero, and m_{0+} have $Re(\lambda) < 0$, in case $d < 0$. (The condition of ergodicity still is $d < 0$.)

Numbering the m_* solutions $\{(\lambda_i, \phi(i)), i = 1, 2, \ldots, m_*\}$ such that $\lambda_1, \lambda_2, \ldots, \lambda_{m_{0+}}$ have negative real part, we see from (36) that the general solution is given by

$$p(x) = \sum_{i=1}^{m_{0+}} e^{\lambda_i x} \phi(i) a_i, \qquad (50)$$

where $a_1, a_2, \ldots, a_{m_{0+}}$ are unknown scalars. They can be computed by (from Theorem 5)

$$c_j - \sum_{i=1}^{m_{0+}} \frac{a_i}{\lambda_i} \phi_j(i) = \pi_j, \; j \in E \qquad (51)$$

with $c_j = 0$ for $j \in E_{0+}$. These are m equations for the m unknowns a_j, $1 \leq j \leq m_{0+}$, and c_j for $j \in E_0 \cup E_{0-}$. This completes the analysis and is consistent with the results known in the literature.

$\sigma_j^2 > 0$ for all $j \in E$

In this case $E_+ = E$, and hence, E_0, E_{0+}, E_{0-} and E_* are all null sets. The S matrix is invertible. The partitioning of the matrices is very simple in this case. For example, we get $R_{++} = Q$ and the B matrix is given by

$$B = \begin{bmatrix} 0 & -QS^{-1} \\ I & MS^{-1} \end{bmatrix}. \tag{52}$$

The eigenvectors corresponding to the eigenvalue λ of B are of the form $(\phi, \lambda\phi)$. In the ergodic case ($d < 0$), there are m eigenvalues with negative real parts (Theorem 4). Assuming that these are distinct, and writing them as $\lambda_1, \lambda_2, \ldots, \lambda_m$, and their eigenvectors as $(\phi(1), \lambda_1\phi(1))$, $(\phi(2), \lambda_2\phi(2))$, ..., $(\phi(m), \lambda_m\phi(m))$ we see that the steady-state densities are given by

$$p(x) = \sum_{i=1}^{m} e^{\lambda_i x}\phi(i)a_i, \tag{53}$$

where the unknown scalars a_i are given by the solution to

$$\sum_{i=1}^{m} \frac{a_i}{\lambda_i}\phi(i) = -\pi. \tag{54}$$

Note that there is no mass at zero in any environmental state, i.e., $c_j = 0$ for all $j \in E$.

The contrast between these two special cases is worth noting. In the first case, the signs of the μ_j's play an important role, while in the second case they do not. In the first case if all μ_j's are negative, $X(t)$ will get absorbed in state zero and we will get $c_j = \pi_j$ and $p_j(x) \equiv 0$ for all j. However, in the second case, even if all μ_j's are negative, there is no mass at zero, and $X(t)$ is absolutely continuous. Thus, even a small variance coefficient is likely to have a significant impact on the behavior of the process.

$E_0 = \phi$, $E_* \neq \phi$, $E_+ \neq \phi$

In this case, there are no states j with $\sigma_j^2 = 0$ and $m_j = 0$. However, there is at least one state j with $\sigma_j^2 = 0$ but $m_j \neq 0$. Then in the partitioning of matrices the block corresponding to E_0 is missing. Hence, we have

$$Q = \begin{bmatrix} Q_{++} & Q_{+*} \\ Q_{*+} & Q_{**} \end{bmatrix}, M = \begin{bmatrix} M_+ & 0 \\ 0 & M_* \end{bmatrix}, S = \begin{bmatrix} S_+ & 0 \\ 0 & S_* \end{bmatrix}. \tag{55}$$

Here $S_* = 0$, and M_* is invertible. Also, $R_{++} = Q_{++}$, etc. The matrix B becomes

$$B = \begin{bmatrix} 0 & Q_{+*}M_*^{-1} & -Q_{++}S_+^{-1} \\ 0 & Q_{**}M_*^{-1} & -Q_{*+}S_+^{-1} \\ I_+ & 0 & M_+S_+^{-1} \end{bmatrix}. \tag{56}$$

The rest of the results can be applied directly.

$E_0 \neq \phi$, $E_* = \phi$, $E_+ \neq \phi$

In this case, $\sigma_j^2 = 0 \Rightarrow m_j = 0$. In the partitioning of matrices the blocks corresponding to E_* are missing. Hence, we have

$$Q = \begin{bmatrix} Q_{++} & Q_{+0} \\ Q_{0+} & Q_{00} \end{bmatrix}, M = \begin{bmatrix} M_+ & \\ & M_0 \end{bmatrix}, S = \begin{bmatrix} S_+ & \\ & S_0 \end{bmatrix}. \tag{57}$$

Here $S_0 = M_0 = 0$. The matrix B becomes

$$B = \begin{bmatrix} 0 & -R_{++}S_+^{-1} \\ I_+ & M_+S_+^{-1} \end{bmatrix}. \tag{58}$$

Under the condition of ergodicity ($d < 0$), the above matrix has $2m_+$ eigenvalues of which m_+ have negative real part. Let the eigenvalues with negative real parts be $\lambda_1, \lambda_2, \ldots, \lambda_{m_+}$. The eigenvector corresponding to λ_i is of the form $(\phi_+(i), \lambda_i\phi_+(i))$. The general solution is given by (assuming the eigenvalues to be distinct)

$$p(x) = \sum_{i=1}^{m_+} a_i\phi(i)e^{\lambda_i x}, \tag{59}$$

where

$$\phi(i) = (\phi_+(i), -\phi_+(i)Q_{+0}Q_{00}^{-1}). \tag{60}$$

The scalars a_i satisfy

$$c - \sum_{i=1}^{m_+} \frac{a_i}{\lambda_i}\phi(i) = \pi \tag{61}$$

with $c_j = 0$ for all $j \in E_+$. Note that in this case we do not get any mass at zero for the X process, i.e., $c_j = 0$ for all $j \in E$. Thus, the above equations will automatically produce $c_j = 0$ for $j \in E_0$.

5. EXAMPLES

Example 1

We start this section with a simple two-state environment process. Such a process occurs naturally in applications where the source generating the fluid that enters the buffer behaves in an on-off fashion. Such a source spends exponential times in the on and off states and continually alternates between the two states. Typically, when the source is on, the fluid content in the buffer will have a positive drift and when the source is off it will have a negative drift.

Formally, $\{Z(t), t \geq 0\}$ is a CTMC on $E = \{1, 2\}$ with

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}. \tag{62}$$

The drift matrix is

$$M = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \tag{63}$$

and the variance matrix is

$$S = \begin{bmatrix} \sigma_1^2/2 & 0 \\ 0 & \sigma_2^2/2 \end{bmatrix}. \tag{64}$$

We assume that $\mu_1 > 0$ and $\mu_2 < 0$, i.e., state 1 is on and state 2 is off. The steady-state distribution of $\{Z(t), t \geq 0\}$ is

$$[\pi_1, \pi_2] = \left[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right]. \tag{65}$$

The drift is

$$d = \mu_1 \frac{\beta}{\alpha+\beta} + \mu_2 \frac{\alpha}{\alpha+\beta} \tag{66}$$

and we will assume that $d < 0$. This implies that

$$\lambda = \frac{\alpha}{\mu_1} + \frac{\beta}{\mu_2} > 0. \tag{67}$$

We study several special cases of this example below to illustrate the influence of the variance parameters on the behavior of the $\{X(t), t \geq 0\}$ process.

Case 1. ($\sigma_1^2 = \sigma_2^2 = 0$—The fluid flow case) This is the first case of Section 4. There is no variance component to the system. This simple on-off fluid model has been well studied in literature. We restate the final results for comparison:

$$c_1 = 0; \quad p_1(x) = \frac{\beta}{\alpha+\beta} \lambda e^{-\lambda x}, x > 0; \tag{68}$$

$$c_2 = \frac{\alpha}{\alpha+\beta} \cdot \frac{\lambda\mu_2}{\lambda\mu_2 - \beta};$$

$$p_2(x) = \frac{\alpha}{\alpha+\beta} \cdot \frac{\beta}{\beta - \lambda\mu_2} \cdot \lambda e^{-\lambda x}, x > 0. \tag{69}$$

Case 2. ($\sigma_1^2 > 0, \sigma_2^2 = 0$) This is the third case of Section 4. We have

$$E_+ = \{1\}, E_0 = \phi, E_{0+} = \phi, E_{0-} = \{2\} = E_*.$$

The B matrix of (56) becomes

$$B = \begin{bmatrix} 0 & \alpha/\mu_2 & 2\alpha/\sigma_1^2 \\ 0 & -\beta/\mu_2 & -2\beta/\sigma_1^2 \\ 1 & 0 & 2\mu_1/\sigma_1^2 \end{bmatrix}. \tag{70}$$

The three eigenvalues are:

$$\lambda_1 = \frac{\mu_1}{\sigma_1^2}\left\{1 - \frac{\beta\sigma_1^2}{2\mu_1\mu_2} - \sqrt{\left(1 - \frac{\beta\sigma_1^2}{2\mu_1\mu_2}\right)^2 + \frac{2\sigma_1^2}{\mu_1}\left(\frac{\alpha}{\mu_1} + \frac{\beta}{\mu_2}\right)}\right\}$$

$$\lambda_2 = 0 \tag{71}$$

$$\lambda_3 = \frac{\mu_1}{\sigma_1^2}\left\{1 - \frac{\beta\sigma_1^2}{2\mu_1\mu_2} + \sqrt{\left(1 - \frac{\beta\sigma_1^2}{2\mu_1\mu_2}\right)^2 + \frac{2\sigma_1^2}{\mu_1}\left(\frac{\alpha}{\mu_1} + \frac{\beta}{\mu_2}\right)}\right\}.$$

It is easy to see that $\lambda_1 < 0$ and $\lambda_3 > 0$. Using the left eigenvector of B for λ_1 we get

$$\phi(1) = \left[1, \frac{\alpha}{\beta + \mu_2\lambda_1}\right]. \tag{72}$$

Hence, using Theorem 5, we get

$$[p_1(x), p_2(x)] = \left[1, \frac{\alpha}{\beta + \mu_2\lambda_1}\right] a_1 e^{\lambda_1 x}, \tag{73}$$

where a_1 is a constant. Also, we have $c_1 = 0$. From (41) we get

$$[0, c_2] - \frac{a_1}{\lambda_1}\left[1, \frac{\alpha}{\beta + \mu_2\lambda_1}\right] = \left[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right]. \tag{74}$$

Solving for c_2 and a_1 we get

$$c_2 = \frac{\alpha}{\alpha+\beta} \cdot \frac{\lambda_1\mu_2}{\beta + \lambda_1\mu_2}, \quad a_1 = -\frac{\beta}{\alpha+\beta} \cdot \lambda_1. \tag{75}$$

Combining the above we get

$$c_1 = 0; \quad p_1(x) = \frac{\beta}{\alpha+\beta} \cdot (-\lambda_1) e^{\lambda_1 x} \quad x > 0 \tag{76}$$

$$c_2 = \frac{\alpha}{\alpha+\beta} \cdot \frac{\lambda_1\mu_2}{\beta + \lambda_1\mu_2}, \quad p_2(x) = \frac{\alpha}{\alpha+\beta} \cdot \frac{\beta}{\beta + \lambda_1\mu_2} \cdot (-\lambda_1) e^{\lambda_1 x}, \quad x > 0. \tag{77}$$

We should contrast these results with the case where $\sigma_1^2 = 0$, i.e., the fluid model case. Equations (76)–(77) are identical to (68)–(69) if λ_1 is replaced by $-\lambda$. We can show that λ_1 is an increasing function of σ_1^2, increasing from $-\lambda$ at $\sigma_1^2 = 0$ and approaching 0 as σ_1^2 increases. Thus, the presence of the variance component leads to fatter tails and higher variances. Using flow models ignores the variance component, thus leading to less congestion. One can show that the expected buffer content grows asymptotically linearly with σ_1^2.

Case 3. ($\sigma_1^2 = 0; \sigma_2^2 > 0$) Now we have

$$E_+ = \{2\}, E_* = E_{0+} = \{1\}, E_0 = E_{0-} = \phi.$$

The B matrix of (56) is

$$B = \begin{bmatrix} 0 & \beta/\mu_1 & 2\beta/\sigma_2^2 \\ 0 & -\alpha/\mu_1 & -2\alpha/\sigma_2^2 \\ 1 & 0 & 2\mu_2/\sigma_2^2 \end{bmatrix}. \tag{78}$$

Note that column 1 corresponds to state 2 (E_+) and column 2 corresponds to state 1 (E_*). We now get two negative eigenvalues and one zero. These are:

$$\lambda_1 = \frac{\mu_2}{\sigma_2^2}\left\{\left(1 - \frac{\alpha\sigma_2^2}{2\mu_1\mu_2}\right) - \sqrt{\left(1 - \frac{\alpha\sigma_2^2}{2\mu_1\mu_2}\right)^2 + \frac{2\sigma_2^2}{\mu_2}\left(\frac{\alpha}{\mu_1} + \frac{\beta}{\mu_2}\right)}\right\}$$

$$\lambda_2 = \frac{\mu_2}{\sigma_2^2}\left\{\left(1 - \frac{\alpha\sigma_2^2}{2\mu_1\mu_2}\right) + \sqrt{\left(1 - \frac{\alpha\sigma_2^2}{2\mu_1\mu_2}\right)^2 + \frac{2\sigma_2^2}{\mu_2}\left(\frac{\alpha}{\mu_1} + \frac{\beta}{\mu_2}\right)}\right\}. \tag{79}$$

$$\lambda_3 = 0.$$

It is easy to see that $\lambda_1 < 0$, $\lambda_2 < 0$. We also have

$$\phi(i) = \left[1, \frac{\beta}{\alpha + \mu_1\lambda_i}\right] \quad i = 1, 2,$$

as the vectors corresponding to λ_1, λ_2. In this case there is no mass at zero, and hence $c_1 = c_2 = 0$. The densities are given by

$$[p_2(x), p_1(x)] = a_1\phi(1)e^{\lambda_1 x} + a_2\phi(2)e^{\lambda_2 x},$$

where a_1 and a_2 are constants that satisfy:

$$-\frac{a_1}{\lambda_1}\phi(1) - \frac{a_2}{\lambda_2}\phi(2) = \left[\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\right]. \tag{80}$$

The contrast with the fluid flow model ($\sigma_2^2 = 0$) in this case is even more dramatic. The fluid flow case predicts that $c_2 > 0$, while the presence of the variance component makes $c_2 = 0$. As before, variance component generates fatter tails in the limiting distribution of $\{X(t)\}$. It is interesting to study the limiting behavior as σ_2^2 approaches zero. In this case, λ_1 approaches $-\lambda$ as before, however, λ_2 approaches $-\infty$, thus the λ_2 component produces the mass at zero in the limit as σ_2^2 approaches zero. (Note, however, that $c_2 = 0$ as long as $\sigma_2^2 > 0$). The component of $p_2(x)$ corresponding to $e^{\lambda_2 x}$ generates a spike at $x = 0$ as $\sigma_2^2 \to 0$. The expected buffer content in this case grows linearly with σ_2^2. In fact, one can derive the following expression after much algebra:

$$\tau = \left(\frac{1}{\lambda} - \frac{\mu_1}{\alpha+\beta}\right) - \left(\frac{1}{2\lambda}\cdot\frac{\alpha}{\mu_1\mu_2}\right)\sigma_2^2. \tag{81}$$

In contrast, in case 2, τ increased linearly with σ_1^2 only asymptotically.

Case 4. ($\sigma_1^2 > 0$, $\sigma_2^2 > 0$) In this case we have

$E_+ = \{1, 2\}$; $E_0 = E_{0+} = E_{0-} = E_* = \phi$.

Hence, we are in the second case of Section 4. The B matrix in this case is:

$$B = \begin{bmatrix} 0 & 0 & 2\alpha/\sigma_1^2 & -2\alpha/\sigma_1^2 \\ 0 & 0 & -2\beta/\sigma_2^2 & 2\beta/\sigma_2^2 \\ 1 & 0 & 2\mu_1/\sigma_2^2 & 0 \\ 0 & 1 & 0 & 2\mu_2/\sigma_2^2 \end{bmatrix}. \tag{82}$$

The eigenvalues of B have to be computed numerically. However, we know from Theorem 4 that two eigenvalues have negative real part (call them λ_1, λ_2), one is zero (call it λ_3), and one has positive real part (call it λ_4). It can be shown that the corresponding ϕ vectors are given by $\phi(i) = [1, \phi_2(i)]$, where

$$\phi_2(i) = \frac{\sigma_2^2}{2\beta}\left(\frac{2\alpha}{\sigma_1^2} + \frac{2\mu_1}{\sigma_1^2}\lambda_i - \lambda_i^2\right). \tag{83}$$

From the theory developed in Section 3 we see that

$$c_1 = c_2 = 0 \tag{84}$$

$$[p_1(x), p_2(x)] = a_1\phi(1)e^{\lambda_1 x} + a_2\phi(2)e^{\lambda_2 x}, \tag{85}$$

where the scalars a_1 and a_2 satisfy

$$-\frac{a_1}{\lambda_1}\phi(1) - \frac{a_2}{\lambda_2}\phi(2) = \left[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right]. \tag{86}$$

The numerical results for these four cases are plotted in Figures 2 and 3. Figure 2 shows the limiting cdf of the $\{X(t), t \geq 0\}$ process for the four cases. The fluid flow case produces the topmost curve—stochastically the smallest congestion. The case $\sigma_1^2 = \sigma_2^2 = 1$ is the lowest,

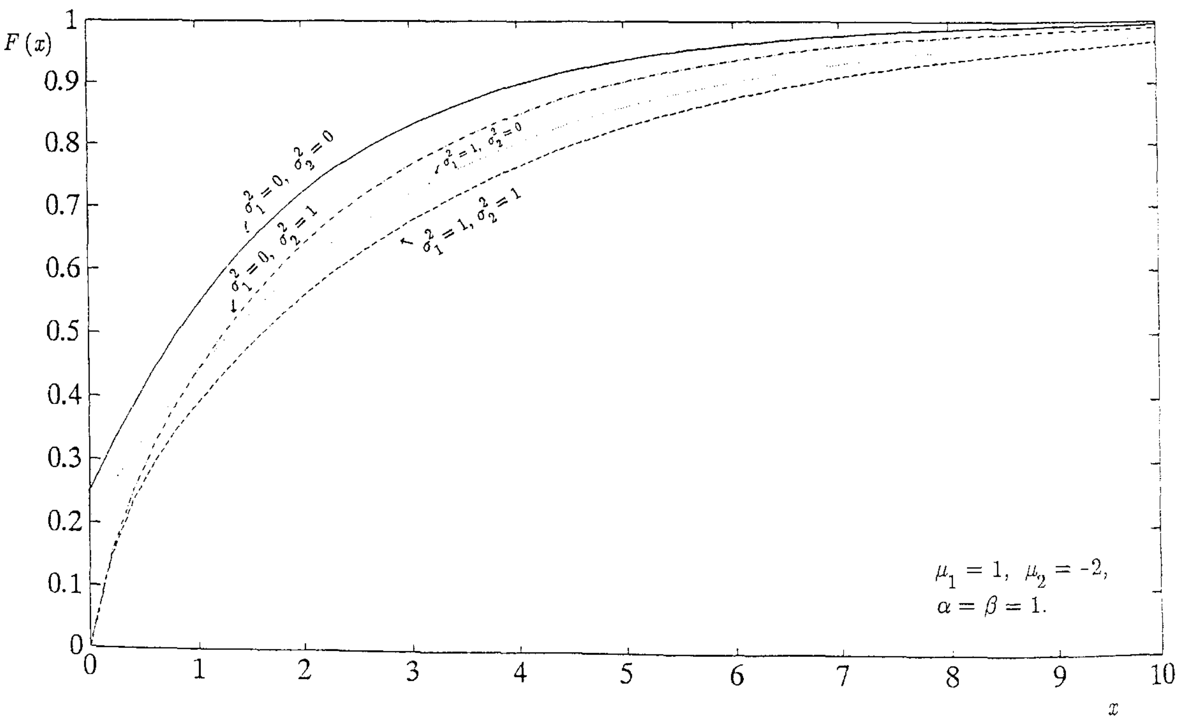

Figure 2. The limiting cdf of the buffer content for a single on-off source.

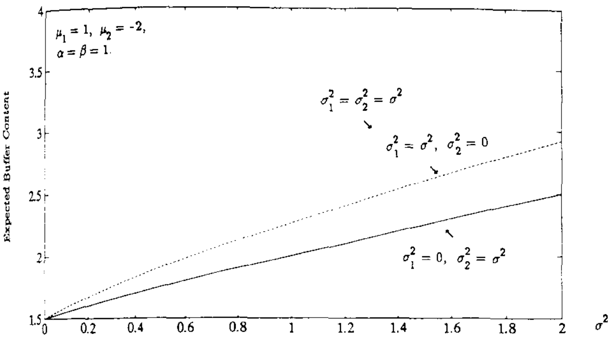

Figure 3. The plot of expected buffer content versus σ^2 for the two-state example.

stochastically largest congestion. The other two curves lie in between these two extremes but they intersect, i.e., the case $\sigma_1^2 = 1$, $\sigma_2^2 = 0$ and $\sigma_1^2 = 0$, $\sigma_2^2 = 1$ are not stochastically comparable.

In many applications, the buffer size is set to be the αth percentile of the steady-state cdf of the X process (typically $\alpha = 1 - 10^{-7}$!). It is clear from the curves that the variance has a dramatic effect on the percentile. Thus, one should be careful when choosing buffer size based on first-order fluid models.

Figure 3 shows the expected buffer content as a function of the variance. The three curves represent the three cases: $2(\sigma_1^2 = \sigma^2, = \sigma_2^2 = 0)$, $3(\sigma_1^2 = 0, \sigma_2^2 = \sigma^2)$, and $4(\sigma_1^2 = \sigma_2^2 = \sigma^2)$. They all start at the same ordinate at $\sigma^2 = 0$, because this point corresponds to the fluid model. Note that the expected value increases asymptotically linearly with σ^2. (In case 3 it is exactly linear as mentioned earlier.) Interestingly, case 2 dominates case 3 in the expected value setting. Thus, variance in a state with positive drift seems to have a more pronounced effect on the expected congestion than the variance in a state with negative drift.

Example 2.

Here we study the effect of multiplexing a fixed number of on-off sources from Example 1. Assume that there are m sources in the system. Each source behaves like an on-off source, as described in Example 1, with drifts μ_i/m and variances $\sigma_i^2/m (i = 1, 2)$. All sources are independent of each other.

The environment $Z(t)$ in this case is the number of sources in on state in the system at time t. From the description above $\{Z(t), t \geq 0\}$ is a birth-and-death process on $\{0, 1, 2, \ldots, m\}$ with birth rates $(m - i)\beta$, in state i, $0 \leq i \leq m$, and death rates $i\alpha$ in state i, $0 \leq i \leq m$. Assuming additivity, when there are i sources in on state, the fluid level in the buffer is a Brownian motion with drift $(i\mu_1 + (m - i)\mu_2)/m$, and variance coefficient $(i\sigma_1^2 + (m - i)\sigma_2^2)/m$.

From the theory of birth-and-death processes we get

$$\pi_i = \lim_{t\to\infty} P\{Z(t) = i\} = \binom{m}{i}\left(\frac{\beta}{\alpha + \beta}\right)^i \left(\frac{\alpha}{\alpha + \beta}\right)^{m-i}. \quad (87)$$

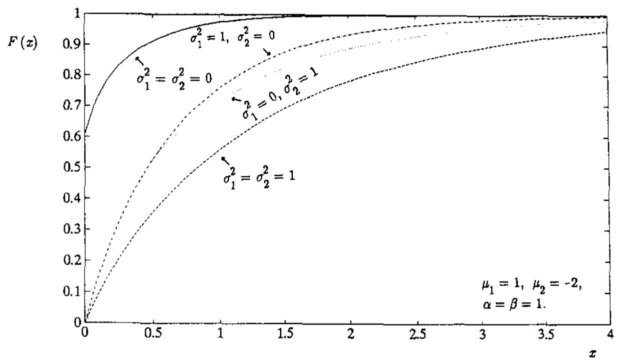

Figure 4. The limiting cdf of buffer content for the case of five multiplexed sources.

The drift is given by

$$d = \left(\mu_1 \frac{\beta}{\alpha + \beta} + \mu_2 \frac{\alpha}{\alpha + \beta}\right) \quad (88)$$

which is the same as in the single source model of Example 1. For ergodicity, we assume that $d < 0$. As before, we study four cases numerically:

1. $\sigma_1^2 = \sigma_2^2 = 0$,
2. $\sigma_1^2 > 0;\quad \sigma_2^2 = 0$,
3. $\sigma_1^2 = 0;\quad \sigma_2^2 > 0$,
4. $\sigma_1^2 > 0;\quad \sigma_2^2 > 0$.

The effect of variance coefficient on congestion is plotted in Figures 4 and 5. In Figure 4 we show the steady cdf of $\{X(t), t \geq 0\}$ for the cases 2, 3, and 4 when $m = 5$. The qualitative nature of Figure 4 is the same as that of Figure 2. In Figure 5, we show the effect of increasing m on the expected congestion. Note that the effective load on the system (as characterized by d) is independent of m. From the figure it seems that multiplexing a large number of (correspondingly) small sources is helpful in reducing congestion. This is the beneficial effect of multiplexing.

It would be useful to have closed-form results for this case along the same lines as in Anick, Mitra and

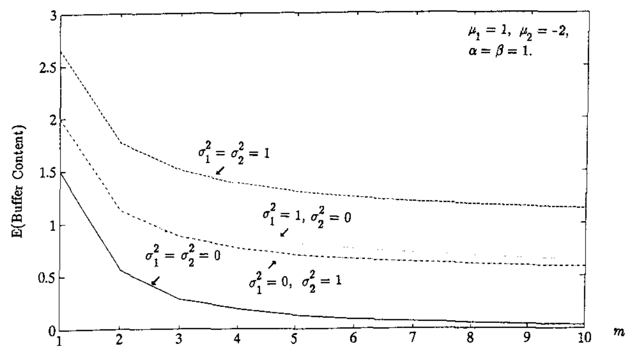

Figure 5. The expected buffer content as a function of m, the number of multiplexed sources.

Sondhi, however, we have been unable to obtain them so far.

6. EXTENSIONS

6.1. Finite Capacity Buffer

An extension suggested by practical application is to consider the buffer to have a finite capacity M. This means $0 \leqslant X(t) \leqslant M$ for all t. This is achieved by making the state M reflecting for the $X(t)$ process.

In this case, the $\{(X(t), Z(t))\}$ process is always ergodic regardless of the drift d. Now it could have point masses at $\{(0, j), (M, j), j \in E\}$. Let c_j be the mass at $(0, j)$, h_j be the mass at (M, j), and $p_j(x)$ be the density over $\{(x, j)\colon 0 \leqslant x \leqslant M, j \in E\}$. The analog of Theorem 2 is as follows.

Theorem 7

$$\tfrac{1}{2}\sigma_j^2 p_j''(x) - \mu_j p_j'(x) + \sum_k p_k(x) q_{kj} = 0 \tag{89}$$

along with boundary conditions

$$\tfrac{1}{2}\sigma_j^2 p_j'(0) - \mu_j p_j(0) + \sum_k c_k q_{kj} = 0 \tag{90}$$

$$\tfrac{1}{2}/\sigma_j^2 p_j'(M) - \mu_j p_j(M) - \sum_k h_k q_{kj} = 0 \tag{91}$$

$$c_k = 0 \quad \textit{for } k \in E_+ \cup E_{0+} \tag{92}$$

and

$$h_k = 0 \quad \textit{for } k \in E_+ \cup E_{0-}. \tag{93}$$

In this case, the general solution $p(x)$ to (89) is still given by (35). We can no longer conclude that $a_i = 0$ if $Re(\lambda_i) > 0$. The $2m_+ + m^*$ scalars $\{a_i\}$ along with c_k: $k \notin (E_+ \cup E_{0+})$ and h_k: $k \notin (E_+ \cup E_{0-})$ form a total of $2m$ unknown constants. These are determined by the $2m$ equations (90) and (91).

6.2. Absorbing States

In certain applications, the state $x = 0$ is an absorbing state (rather than reflecting, as we have assumed so far) for the $\{X(t)\}$ process when the environment is in a set A of given states, e.g., $A = \{j\colon \mu_j < 0\}$. It is reflecting in other states. Here the generator of the $\{X(t), Z(t))\}$ process is still given by (5), however the class of function f for which it is valid is: f twice continuously differentiable in the x variable with

$$f'(x, j) = 0 \quad \text{for } j \in [E_+ \cup E_0 \cup E_{0-}] \cap A^c \tag{94}$$

and

$$\tfrac{1}{2}\sigma_j^2 f''(0, j) + \mu_j f'(0, j) = 0 \quad \text{for } j \in A \cap E_{0+}^c. \tag{95}$$

Theorem 1 continues to hold with the following equations replacing (11).

$$c(t, j; y, i) = 0 \quad \text{for } j \in (E_+ \cap A^c) \cup E_{0+} \tag{96}$$

and

$$p(t, 0, j; y, i) = 0 \quad \text{for } j \in (E_+ \cap A). \tag{97}$$

Similarly, the steady-state distribution is given by (14) with boundary conditions (15) and

$$c_j = 0 \quad \text{for } j \in (E_+ \cap A^c) \cup E_{0+} \tag{98}$$

and

$$p_j(0) = 0 \quad \text{for } j \in E_+ \cap A. \tag{99}$$

The explicit solution for the stationary distribution can be obtained following the development in subsection 3.5.

It is also possible to consider the case of finite buffer, where (M, j) is absorbing for j in a subset B of E. We do not go into the details of this case.

APPENDIX

Proof of Theorem 1. Since the initial condition (y, i) is fixed throughout, we drop it from the notation and write $F(t, x, j; y, i)$ as $F(t, x, j)$, etc. Substituting for L from (5) in (8), we get

$$\begin{aligned}\frac{d}{dt}\left[\sum_j f(0, j)c(t, j) + \sum_j \int_0^\infty f(x, j)p(t, x, j)\, dx\right] &= \sum_j \{\tfrac{1}{2}\sigma_j^2 f''(0, j) + \mu_j f'(0, j) + \sum_{k=1}^m q_{jk} f(0, k)\} \\ &\quad \cdot c(t, j) + \sum_j \int_0^\infty \{\tfrac{1}{2}\sigma_j^2 f''(x, j) + \mu_j f'(x, j) \\ &\quad + \sum_{k=1}^m q_{jk} f(x, k)\} p(t, x, j)\, dx,\end{aligned} \tag{A.1}$$

where $f'(x, j)$ and $f''(x, j)$ are the first and second derivatives of $f(x, j)$ with respect to x.

Restricting f to be such that $f(x, j) = 0$ for $x \geqslant T$ and using integration by parts, we see that

$$\begin{aligned}\int_0^T f''(x, j)p(t, x, j)\, dx &= f'(x, j)p(t, x, j)\Big|_0^T - \int_0^T f'(x, j)p'(t, x, j)\, dx \\ &= -f'(0, j)p(t, 0, j) - f(x, j)p'(t, x, j)\Big|_0^T + \int_0^T f'(x, j)p'(t, x, j)\, dx \\ &= -f'(0, j)p(t, 0, j) + f(0, j)p'(t, 0, j) + \int_0^T f(x, j)p''(t, x, j)\, dx.\end{aligned} \tag{A.2}$$

Similarly

$$\int_0^T f'(x, j)p'(t, x, j)\, dx = -f'(0, j)p(t, 0, j) - \int_0^T f(x, j)p'(t, x, j)\, dx. \tag{A.3}$$

Substituting in (A.1) and collecting terms, we get

$$\sum_j \int_0^\infty \Big\{ \frac{\partial}{\partial t} p(t, x, j) - \tfrac{1}{2}\sigma_j^2 p''(t, x, j) + \mu_j p'(t, x, j) - \sum_k p(t, x, j) q_{kj} \Big\} f(x, j)\, dx$$
$$+ \sum_j f(0, j) \Big\{ \frac{\partial}{\partial t} c(t, j) - \sum_k c(t, k) q_{kj} - \tfrac{1}{2}\sigma_j^2 p'(t, 0, j) + \mu_j p(t, 0, j) \Big\}$$
$$+ \sum_j f'(0, j)\{\tfrac{1}{2}\sigma_j^2 p(t, 0, j) - \mu_j c(t, j)\} - \sum_j f''(0, j) \cdot \{\tfrac{1}{2}\sigma_j^2 c(t, j)\} = 0. \qquad (A.4)$$

In (A.4), $\{f(x, j)\}$ is any smooth function, vanishing for $x \geq T$ and such that $f'(0, j) = 0$ for $j \in E_+ \cup E_0 \cup E_{0-}$. By first taking $\{f(x, k)\}$ such that $f(0, j) = f'(0, j) = f''(0, j) = 0$, for a fixed j and $f(x, k) = 0$ for $k \neq j$, we get

$$\frac{\partial}{\partial t} p(t, x, j) = \tfrac{1}{2}\sigma_j^2 p''(t, x, j) - \mu_j p'(t, x, j) + \sum_k p(t, x, k) q_{kj} \qquad (A.5)$$

which is (9). Arguing similarly, we will get that the coefficients of $f(0, j)$, $f''(0, j)$ are zero and the coefficients of $f'(0, j)$ are zero for $j \in E_{0+}$, thus yielding (10) and (11).

Proof of Theorem 4. The proof closely follows that of Gerschgorin's theorem (see Horn 1990). It is clear that $\det[\lambda^2 S - \lambda M + Q]$ is a polynomial of degree $2m_+ + m_*$, and hence, has $2m_+ + m_*$ solutions.

Now write $Q = Q_d + Q_0$, where $Q_d = \text{diag}(Q)$. For $0 \leq \epsilon \leq 1$, define $Q^\epsilon = Q_d + (1 - \epsilon)Q_0$. Thus $Q^0 = Q$.

We now study the roots of

$$\det[\lambda^2 S - \lambda M + Q^\epsilon] = 0 \qquad (A.6)$$

for $\epsilon > 0$. If λ satisfies (A.6), then there is a nonzero vector ϕ such that

$$\phi[\lambda^2 S - \lambda M + Q^\epsilon] = 0 \qquad (A.7)$$

If the rth component of ϕ has the largest modulus, normalize $\phi = (\phi_1, \phi_2, \ldots, \phi_m)$ such that $\phi_r = 1$ and $|\phi_i| \leq 1$, $1 \leq i \leq m$. The rth component of (A.7) yields

$$\tfrac{1}{2}\sigma_r^2 \lambda^2 - \lambda \mu_r - \sum_i \phi_i [Q^\epsilon]_{ir} = 0$$

i.e.,

$$\tfrac{1}{2}\sigma_r^2 \lambda^2 - \lambda \mu_r - q_r + \sum_{i \neq r} \phi_i (1 - \epsilon) q_{ir} = 0 \qquad (A.8)$$

(where $q_r = -q_{rr}$ is positive). Hence

$$|\tfrac{1}{2}\sigma_r^2 \lambda^2 - \lambda \mu_r - q_r| \leq \sum_{i \neq r} |\phi_i (1 - \epsilon) q_{ir}| \leq (1 - \epsilon) \sum_{i \neq r} q_{ir} = (1 - \epsilon) q_r. \qquad (A.9)$$

It now follows that all solutions λ of (A.6) must belong to $\cup_{r=1}^m A_r^\epsilon$ where,

$$A_r^\epsilon = \{\lambda \in C: |\lambda^2 \sigma_r^2/2 - \lambda \mu_r - q_r| \leq (1 - \epsilon) q_r\}. \qquad (A.10)$$

It is easy to see that for $r \in E_+$, A_r^ϵ is a union of two disjoint regions, one in $C^- = \{\lambda \in C: Re(\lambda) < 0\}$ and one in $C^+ = \{\lambda \in C: Re(\lambda) > 0\}$. Furthermore, for $r \in E_0$, A_r^ϵ is empty, $r \in E_{0+}$, $A_r^\epsilon \subseteq C^-$ and for $r \in E_{0-}$, $A_r^\epsilon \subseteq C^+$. Thus, there are $2m_+ + m^*$ nonempty regions of which $m_+ + m_{0-}$ are in C^+, and $m_+ + m_{0+}$ are in C^-.

Using the fact that the solutions λ of (A.6) follow a continuous trajectory as a function of ϵ and following the proof of Gerschgorin's theorem, we can show that exactly $m_+ + m_{0-}$ solutions are in C^+ and $m_+ + m_{0+}$ solutions are in C^-.

We are interested in limiting case $\epsilon = 0$. At $\epsilon = 0$, $\lambda = 0$ is a solution to (A.6). Let λ^ϵ be a solution of (A.6) with the property that $\lambda^\epsilon \to 0$ as $\epsilon \to 0$. This can be done in view of the continuity of solutions mentioned above.

Arguing as in Horn, Article 6.3.12, it can be deduced that

$$\left.\frac{d\lambda^\epsilon}{d\epsilon}\right|_{\epsilon=0} = -\frac{\sum \pi_i q_i}{\sum \pi_i \mu_i}. \qquad (A.11)$$

Since $d = \sum_{i=1}^m \mu_i \pi_i$ is assumed to be negative, and $\sum \pi_i q_i$ is positive, $d\lambda^\epsilon/d\epsilon$ is positive at $\epsilon = 0$. Thus $\lambda^\epsilon \in C^+$ for ϵ near zero. Hence, in the limit we get $m_+ + m_{0-} - 1$ solutions to (25) in C^+, $m_+ + m_{0+}$ solution in C^-, and one solution $\lambda = 0$.

ACKNOWLEDGMENT

The research of the first author was supported by NSF, AFOSR grant F49620-92-J0154 and ARO grant DAAL0392G0008. The research of the second author was partially supported by IBM contract no. 1374.

REFERENCES

ANICK, D., D. MITRA AND M. M. SONDHI. 1982. Stochastic Theory of Data Handling Systems. *Bell Syst. Tech. J.* **61,** 1871–1894.

ASMUSSEN, S. 1992. Stationary Distribution for Fluid Flow Models and Markov-Modulated Reflected Brownian Motion. TR R92-2015, Institute of Electronic Systems, Department of Mathematics and Computer Science, Aalborg University, Aalborg, Denmark.

CHEN, H. 1988. A Heterogenous Fluid Model and Its Applications in Multiclass Queueing Networks. Working Paper 89 MSC-013, Faculty of Commerce, University of British Columbia, Vancouver, Canada.

CHEN, H., AND D. YAO. 1992. A Fluid Model for Systems With Random Disruptions. *Opns. Res.* **40,** S239–S247.

ELWALID, A. I., AND D. MITRA. 1991. Analysis and Design of Rate-Based Congestion Control of Highspeed Networks I: Stochastic Fluid Models, Access Regulation. *Queue. Syst.* **9,** 29–64.