Large Deviations of Birth Death Markov Fluids

G. De Veciana, C. Olivier*, and J.Walrand
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley
Berkeley CA 94720

July 25, 1995

Abstract

The asymptotic probability of buffer overflow for a queueing system with a Markov fluid input and deterministic service rate is derived by way of large deviation theory. The equations characterizing the deviant behavior are presented and examples are given for which closed form solutions may be obtained. An independence result extends the analysis to cases where the input is an aggregate of independent Markov fluids.

1 Introduction

We will investigate the probability of a buffer overflow for systems in which the input traffic is modeled as a Markov fluid with an underlying chain of the birth death type. This problem has received much attention in the literature, and many approaches have been developed to characterize, not only the statistics of the queue length, but also the manner in which overflows occur Anick et al. [1]. Our results are obtained by way of large deviation theory, and thus are closely related to the work of Weiss [7]. However our approach is different and motivated by the recent work of Kesidis [4]. Our goal is to fully explore this framework for the specific case of the birth death processes.

We consider a buffering system of size B with a deterministic service rate c, and an N-rate Markov fluid source. Let X_t denote the free buffer process, in the sense that it is not constrained to be positive or below B. The evolution of X_t is given by

$$\frac{dX_t}{dt} = r(Y_t) - c \tag{1}$$

where Y_t is a continuous-time birth death process with states $0, \ldots, N-1$ and rate matrix P^0 :

$$\begin{aligned} {P^0}_{i,i+1} &= \lambda_i, \quad i = 0, \ldots, N-2; \\ {P^0}_{i,i-1} &= \mu_i, \quad i = 1, \ldots, N-1. \end{aligned}$$

*Research supported by Direction des Recherches, Etudes, et Techniques, France.

A deterministic traffic rate $r(Y_t)$ is associated with each state of the Markov chain. Note that the traffic rate will be Markov if $r(\cdot)$ is one-to-one, however the queue length process is not.

The goal is to compute the asymptotic probability of a buffer overflow in a busy cycle as $B \to \infty$. If the system is stable this becomes a large deviation, and that theory becomes directly applicable. We refer the interested reader to Bucklew [2], Kesidis [4], and references therein and only provide a heuristic introduction to the results that we will use.

Our starting point is an expression for the relative entropy between two continuous-time Markov chains. Let P and P^0 be the rate matrices of two N state continuous-time Markov chains and denote by π_P the stationary distribution of P. Following Kesidis[5] we define the relative entropy by

$$H(P \parallel P^0) = \sum_{i=1}^{N} \pi_P(i) \left(\sum_{j \neq i} P_{i,j} \log \frac{P_{i,j}}{{P^0}_{i,j}} + {P^0}_{i,j} - P_{i,j} \right).$$

This notion permits us to explore the probability that the Markov chain P^0 behaves as another chain P for an extended period of time. Additionally, one can obtain the action functional for the empirical distributions by considering

$$J_{P^0}(\pi) = \inf_{\{P | \pi P = 0\}} H(P \parallel P^0)$$

from which we extract the exponent for the likelihood of observing a distribution π. We wish to compute $P(X_t > B)$ during a busy cycle. For large B, our heuristic, justifiable by convexity arguments, is that overflows are not due to fluctuations but to a steady buildup in the queue, i.e., "the path to an overflow is a straight line." Thus we evaluate the probability that P^0 behaves like an alternate Markov chain P, which offers a mean traffic rate $M + c$ exceeding the buffer service rate, for a prolonged period of time. This is obtained, once again, by considering a functional of the excess rate M :

$$H^*(M) = \inf_{\{P | E_{\pi_P} r(\cdot) = c + M\}} H(P \parallel P^0).$$

Finally, a bound on the probability of an overflow in a busy cycle of the form $P(X_t > B) \leq \exp(-BE)$ is obtained by considering the most likely path, or slope, leading up to this event :

$$E = \inf_{M > 0} \frac{H^*(M)}{M}. \tag{2}$$

The remainder of this paper is organized as follows. In Section 2 we discuss some properties of the relative entropy when two independent processes are involved. The required minimizations are investigated in Section 3. We establish necessary and sufficient optimality conditions and the uniqueness of the minimum for the constrained relative entropy problem. In addition we discuss the problem of obtaining the overflow exponent directly. This provides some insight into the nature of the solution. Section 4 includes two examples which have analytical solutions. In Section 5, numerical procedures for obtaining the parameters of the deviant Markov chain, as well as the exponent and the effective bandwidth are presented. We conclude with some remarks in Section 6.

2 A preliminary decoupling result

Theorem 1 *Consider the relative entropy $H(Q \parallel P)$ between two rate matrices, Q and P, where P corresponds to the product of two independent Markov chains, $P = P^1 \times P^2$, on two spaces $\mathcal{X}_1$ and $\mathcal{X}_2$. Let Q^1 and Q^2 correspond to the marginal transition rates of the first and second components; thus for example $Q^1_{x_1,y_1} = \sum_{x_2} \pi_Q(x_2 \mid x_1) Q_{(x_1,x_2),(y_1,x_2)}$ where $x_1, y_1 \in \mathcal{X}_1$ and π_Q denotes the invariant distribution of Q. Consider the product-form Markov chain $Q^1 \times Q^2$. Then :*

$$H(Q \parallel P^1 \times P^2) \geq H(Q^1 \parallel P^1) + H(Q^2 \parallel P^2) = H(Q^1 \times Q^2 \parallel P^1 \times P^2). \tag{3}$$

Proof : By definition

$$H(Q \parallel P) = \sum_x \pi_Q(x) \sum_{y \neq x} Q_{x,y} \log \frac{Q_{x,y}}{P_{x,y}} + P_{x,y} - Q_{x,y},$$

where $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Using the fact that P is a product of two Markov chains we can rewrite the entropy as

$$\begin{aligned} &H(Q \parallel P^1 \times P^2) = \\ &\quad \sum_{(x_1,x_2)} \pi_Q(x_1, x_2) \sum_{y_1 \neq x_1} Q_{(x_1,x_2),(y_1,x_2)} \log \frac{Q_{(x_1,x_2),(y_1,x_2)}}{P^1_{x_1,y_1}} + P^1_{x_1,y_1} - Q_{(x_1,x_2),(y_1,x_2)} \\ &\quad + \sum_{(x_1,x_2)} \pi_Q(x_1, x_2) \sum_{y_2 \neq x_2} Q_{(x_1,x_2),(x_1,y_2)} \log \frac{Q_{(x_1,x_2),(x_1,y_2)}}{P^2_{x_2,y_2}} + P^2_{x_2,y_2} - Q_{(x_1,x_2),(x_1,y_2)}. \end{aligned}$$

Using the relationship

$$\pi_Q(x_1, x_2) = \pi_Q(x_1) \pi_Q(x_2 \mid x_1),$$

we can rewrite the terms on the right hand side to obtain

$$\begin{aligned} &H(Q \parallel P^1 \times P^2) = \\ &\quad \sum_{x_1} \sum_{y_1 \neq x_1} \pi_Q(x_1) \sum_{x_2} \pi_Q(x_2 \mid x_1) Q_{(x_1,x_2),(y_1,x_2)} \log \frac{Q_{(x_1,x_2),(y_1,x_2)}}{P^1_{x_1,y_1}} + P^1_{x_1,y_1} - Q_{(x_1,x_2),(y_1,x_2)} \\ &\quad + \text{ the symmetrical term.} \end{aligned}$$

Recall that we have defined $Q^1_{x_1,y_1} = \sum_{x_2} \pi_Q(x_2 \mid x_1) Q_{(x_1,x_2),(y_1,x_2)}$ and $Q^2_{x_2,y_2}$ similarly. Q^1 corresponds to a rate matrix with the average transition rates of Q on the first component. With this definition in mind and using Jensen's inequality on the convex function $x \log x$ we obtain

$$\begin{aligned} H(Q \parallel P^1 \times P^2) \quad \geq \quad & \sum_{x_1} \pi_Q(x_1) \sum_{y_1 \neq x_1} Q^1_{x_1,y_1} \log \frac{Q^1_{x_1,y_1}}{P^1_{x_1,y_1}} + P^1_{x_1,y_1} - Q^1_{x_1,y_1} \\ & + \text{ the symmetrical term.} \end{aligned}$$

Finally, note that in fact

$$Q^1_{x_1,y_1} = \sum_{x_2,y_2} \pi_Q(x_2 \mid x_1) Q_{(x_1,x_2),(y_1,y_2)},$$

so it follows that $\pi_Q(x_1)$ is the invariant distribution for Q^1. Thus,

$$\sum_{x_1} \pi_Q(x_1) \sum_{y_1 \neq x_1} Q^1_{x_1,y_1} \log \frac{Q^1_{x_1,y_1}}{P^1_{x_1,y_1}} + P^1_{x_1,y_1} - Q^1_{x_1,y_1} = H(Q^1 \parallel P^1),$$

and a similar expression is found for the symmetrical term. We have established

$$H(Q \parallel P^1 \times P^2) \geq H(Q^1 \parallel P^1) + H(Q^2 \parallel P^2) = H(Q^1 \times Q^2 \parallel P^1 \times P^2).$$

□

Corollary 1 *Consider the relative entropy $H(Q \parallel P)$ between two rate matrices, Q and P, where P corresponds to the product of two independent Markov chains, $P = P^1 \times P^2$, on two spaces $\mathcal{X}_1$ and $\mathcal{X}_2$. Suppose that the traffic rate corresponding to each product state $r(x_1, x_2)$ is additive* i.e., *there exist functions r_1 and r_2 such that*

$$\forall (x_1, x_2), r(x_1, x_2) = r_1(x_1) + r_2(x_2). \tag{4}$$

Then the minimizer Q^* of $H(Q \parallel P)$ subject to a mean traffic constraint

$$M = E_{\pi_Q} r(X_1, X_2) = E_{\pi_Q} r_1(X_1) + E_{\pi_Q} r_2(X_2) = M_1 + M_2 \tag{5}$$

is of product-form.

Proof : Suppose $Q_{x,y}$ is a minimizer satisfying the constraint. Then by the previous theorem, there is a corresponding product-form Markov chain $Q^1 \times Q^2$ with the same marginal distributions, whence still satisfying the constraints, but with a lesser or equal relative entropy. □

Corollary 2 *Under the assumptions of Corollary 1*

$$\inf_{\{Q | E_{\pi_Q} r(X_1, X_2) = M\}} H(Q \parallel P^1 \times P^2) =$$

$$\inf_{\{(M_1, M_2) | M_1 + M_2 = M\}} \left\{ \inf_{\{Q^1 | E_{\pi_{Q^1}} r^1(X_1) = M_1\}} H(Q^1 \parallel P^1) + \inf_{\{Q^2 | E_{\pi_{Q^2}} r^2(X_2) = M_2\}} H(Q^2 \parallel P^2) \right\}. \tag{6}$$

These results obviously hold in general for N independent Markov fluid sources with additive rates. Intuitively the decoupling result implies that the most likely way for independent Markov chains to deviate from their typical behavior is independently.

This constitutes a justification of the additive property of relative entropy for independent Markov chains assumed in Kesidis [4], leading to the notion of *effective bandwidth.*

3 Solving the optimization problems

In this section, necessary and sufficient optimality equations are derived for the following minimization problems

$$H^*(M) = \inf_{\{P | E_{\pi_P} r(\cdot) = M + c\}} H(P \parallel P^0) \tag{7}$$

where $E_{\pi_P} r(\cdot) = \sum_{i=0}^{N-1} \pi_i r_i$, and

$$E^* = \inf_{\{P | E_{\pi_P} r(\cdot) > c\}} \frac{H(P \parallel P^0)}{E_{\pi_P} r(\cdot) - c}. \tag{8}$$

In computing these infima, we only consider rate matrices P which have the same graph as the initial Markov Chain P^0, i.e., the set of BD processes with rates λ_i, μ_j. We in fact restrain ourselves to structure-preserving parametric changes of measure.

The stationary distribution of such a P is given by Neuts [6] :

$$\pi_i = \pi_0 \frac{\lambda_0 \lambda_1 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i}, \quad i \geq 1, \tag{9}$$

with

$$\pi_0 = \left(1 + \sum_{i=1}^{N-1} \frac{\lambda_0 \lambda_1 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i}\right)^{-1}. \tag{10}$$

The expression for the relative entropy between P and P^0 is now

$$\begin{aligned} H(P \parallel P^0) &= \sum_{i=0}^{N-1} \pi_i \left(\lambda_i \log \frac{\lambda_i}{\lambda_i^0} + \lambda_i^0 - \lambda_i + \mu_i \log \frac{\mu_i}{\mu_i^0} + \mu_i^0 - \mu_i\right) \qquad (11) \\ &= \sum_{i=0}^{N-1} \pi_i \left(\phi_i + \psi_i\right), \end{aligned}$$

with

$$\begin{aligned} \phi_i &= \lambda_i \log \frac{\lambda_i}{\lambda_i^0} + \lambda_i^0 - \lambda_i, \\ \psi_i &= \mu_i \log \frac{\mu_i}{\mu_i^0} + \mu_i^0 - \mu_i, \end{aligned}$$

and the convention that $\lambda_{N-1} = \mu_0 = 0$.

3.1 Minimization of the relative entropy under constraints

We first solve problem (7).

This is done by forming the Lagrangian :

$$L(P, P^0) = H(P \parallel P^0) + K(E_{\pi_P} r(\cdot) - (M + c)), \tag{12}$$

where K is a Lagrange multiplier.

The following two lemmas enable us to derive simple forms for the first order optimality conditions.

Lemma 1 *Define* $S_k = \sum_{j=k}^{N-1} \pi_j$. *Then*

$$\begin{aligned} \frac{\partial \pi_i}{\partial \lambda_k} &= \frac{\pi_i}{\lambda_k}(1_{\{i-1 \geq k\}} - S_{k+1}), \qquad (13) \\ \frac{\partial \pi_i}{\partial \mu_k} &= -\frac{\pi_i}{\mu_k}(1_{\{i \geq k\}} - S_k). \qquad (14) \end{aligned}$$

(15)

Lemma 2 *Define*

$$\alpha_k = \sum_{i=k+1}^{N-1} \pi_i(\phi_i + \psi_i + K r_i). \tag{16}$$

Then the first-order optimality equations for problem (7) are

$$\begin{aligned} -S_{k+1}(H^*(M) + K(M+c)) + \alpha_{k+1} + \pi_k \lambda_k \log \frac{\lambda_k}{\lambda_k^0} &= 0, && (17)\\ S_{k+1}(H^*(M) + K(M+c)) - \alpha_{k+1} + \pi_{k+1}\mu_{k+1} \log \frac{\mu_{k+1}}{\mu_{k+1}^0} &= 0, && (18)\end{aligned}$$

for $k = 0, \ldots, N-2$.

The derivation of these equations as well as the algebra required to establish the following set of optimality conditions have been placed in the appendix.

Proposition 1 *The first-order optimality equations for problem (7) are the following :*

$$\begin{aligned} \lambda_k \mu_{k+1} &= \lambda_k^0 \mu_{k+1}^0, \quad k = 0, \ldots, N-2; && (19)\\ \lambda_k + \mu_k &= -H^*(M) + K(r_k - (M+c)) + \lambda_k^0 + \mu_k^0, \quad k = 0, \ldots, N-1; && (20)\\ \sum_{k=0}^{N-1} \pi_k r_k &= M + c; && (21)\\ H^*(M) &= \sum_{i=0}^{N-2} \Big(\pi_i(\lambda_i^0 - \lambda_i) + \pi_{i+1}(\mu_{i+1}^0 - \mu_{i+1})\Big). && (22)\end{aligned}$$

The last equation defining $H^*(M)$ is in fact redundant, since it can be obtained by adding the second set of equations weighted by coefficients π_k.

Moreover, we have the following proposition.

Proposition 2 *Define* $\{\pi_k\}$ *as the set* $(\pi_0, \ldots, \pi_{N-1})$ *and* $\{\mu_k\}$ *similarly. Consider* $H(P \parallel P^0)$ *as a function of* $\{\pi_k\}$ *and* $\{\mu_k\}$. *Then* $H(P \parallel P^0)$ *is convex in* $\{\mu_k\}$ *and, furthermore, the function* $H^*(\{\pi_k\})$ *defined as* $H^*(\{\pi_k\}) = \min_{\{\mu_k\}} H(P \parallel P^0)$ *is convex in* $\{\pi_k\}$, *so that the optimality equations above actually define a unique minimum.*

The proof is given in the appendix.

This proposition is related to a standard result in information theory regarding the convexity of the relative entropy $D(p \parallel q)$ in the pair of probability distributions (p, q) over a discrete space, this quantity being defined by $D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$ Cover [3]. It implies that the infimum of problem (7) exists and is unique over the class of birth death processes we have considered. This fact shows that among the set of all trajectories with rate $M + c$ over a period T generated by parametric changes of measure, there exists one specific trajectory for P^0 that is asymptotically stricly more likely than the others, assuming that a large deviation principle holds for such a set Bucklew [2]. By Laplace's argument, the probability of P^0 to fire at a mean rate $M + c$ over a period T is then asymptotically equal to the probability of this most likely trajectory when T is large.

3.2 Direct computation of the exponent

Suppose now that we wish to directly compute the exponent, i.e., to solve problem (8) :

$$E = \inf_{\{P | E_{\pi_P} r(.) > c\}} \frac{H(P \parallel P^0)}{E_{\pi_P} r(.) - c}. \tag{23}$$

We will prove successively that this infimum may be characterized by a set of equations similar to those derived before and the minimum is unique.

Proposition 3 *A necessary and sufficient set of optimality equations for problem (8) is*

$$\begin{aligned}
\lambda_k \mu_{k+1} &= \lambda_k^0 \mu_{k+1}^0, \quad k = 0, \ldots, N-2 && (24) \\
H^*(c - r_k) + (\lambda_k^0 - \lambda_k + \mu_k^0 - \mu_k)(\sum_{i=0}^{N-1} \pi_i r_i - c) &= 0, \quad k = 0, \ldots, N-1 && (25) \\
\sum_{i=0}^{N-1} \pi_i r_i &> c && (26) \\
\lambda_k &> 0, \quad k = 0, \ldots, N-2, && (27)
\end{aligned}$$

with, as before, $H^* = \sum_{i=0}^{N-2} \Big(\pi_i(\lambda_i^0 - \lambda_i) + \pi_{i+1}(\mu_{i+1}^0 - \mu_{i+1}) \Big)$.

The proof is similar to that of Proposition 1 with a simplification due to the fact that the Lagrange multiplier is 0 at the optimum because the inequality constraint will not be saturated.

These optimality equations are necessary and sufficient from the existence and uniqueness argument in the previous problem. Indeed, the minimum will be attained in a point P^* such that $E_{\pi_{P^*}} r(\cdot) = c + \epsilon$ for some stricly positive ϵ and, moreover,

$$H(P^* \parallel P^0) = \inf_{\{P | E_{\pi_P} r(\cdot) = c + \epsilon\}} H(P \parallel P^0),$$

which is unique by Proposition 2.

An algorithm is suggested at the end of this paper to compute the solution to these equations.

4 Examples

4.1 Sum of on-off Markov fluid sources

We first consider the case where the input traffic to the buffer is an aggregate of $N-1$ two-state Markov chains. Each of these will contribute a traffic rate of a_0 when off and a_1 when on. A source turns on with intensity λ^0 and off with with intensity μ^0. The aggregate Markov fluid, corresponds to a birth death process with the following parameters :

$$\begin{aligned}
\lambda_i^0 &= (N - 1 - i)\lambda^0; \\
\mu_i^0 &= i\mu^0,
\end{aligned}$$

and with r_k linear in k, i.e. $r_k = \alpha k + \beta$, where $\alpha = a_1 - a_0$ and $\beta = (N-1)a_0$.

Proposition 4 characterizes the solution to the constrained minimization of the relative entropy.

Proposition 4 *Define $a = \alpha(N-1)$. In the on-off case, the optimal solution is :*

$$\begin{aligned} \lambda_i &= (N-1-i)\lambda; \\ \mu_i &= i\mu, \end{aligned}$$

with

$$\begin{aligned} \lambda &= \left(\lambda^0\mu^0 \frac{M+c-\beta}{(a-(M+c)+\beta)}\right)^{\frac{1}{2}}; \\ \mu &= \left(\lambda^0\mu^0 \frac{(a-(M+c)+\beta)}{M+c-\beta}\right)^{\frac{1}{2}}. \end{aligned}$$

The entropy is given by :

$$H^*(M) = (N-1)\left(\frac{\lambda}{\lambda+\mu}(\mu^0-\mu) + \frac{\mu}{\lambda+\mu}(\lambda^0-\lambda)\right). \tag{28}$$

$H^*(M)$ is a sum of $N-1$ equal terms corresponding to each of the $N-1$ two-state Markov fluids that constitute the input traffic. Note that this result could have been obtained directly using Theorem 1.

The exponent is given in the following proposition.

Proposition 5 *Denote the mean offered traffic rate by $\gamma = \frac{\lambda^0}{\lambda^0+\mu^0}a + \beta$. Suppose $\gamma < c$, i.e. the system is stable. Then :*

$$E = (N-1)\frac{((\lambda^0+\mu^0)(c-\beta) - \lambda^0 a)}{(c-\beta)(a+\beta-c)}. \tag{29}$$

This result can be compared to both Weiss [7] and Anick et al. [1]. In the first case let $\beta = 0$, and consider a limit as the number of sources increases, $N \to \infty$, such that the mean offered traffic rate is constant and equal to $\frac{\lambda^0}{\lambda^0+\mu^0} < c < 1$. This corresponds to letting $a = 1$, so $\alpha = 1/N$, and the exponent becomes

$$E = (N-1)(\frac{\mu^0}{1-c} - \frac{\lambda^0}{c}). \tag{30}$$

This is identical to the result obtained by Weiss, in the case of large buffers, and a large number of sources, from a conceptually different point of view [7].

We obtain the asymptotics established by Anick et al. by rewriting our exponent as follows. Note that $\beta = \Gamma_{\min}$, $a + \beta = \Gamma_{\max}$, are respectively the minimum and maximum traffic the aggregate source can offer. One can then rewrite the exponent, in their intuitively pleasing form :

$$E = (N-1)\frac{(\lambda^0+\mu^0)(c-\gamma)}{(c-\Gamma_{\min})(\Gamma_{\max}-c)}. \tag{31}$$

4.2 M/M/∞ Markov Fluid Source

In our second example, we consider an aggregate source in which sources arrive at rate λ^0 and contribute a traffic rate α. They turn off after an exponential period with mean $\frac{1}{\mu^0}$. The corresponding a birth death process has the following parameters :

$$\begin{aligned} \lambda_i^0 &= \lambda^0; \\ \mu_i^0 &= i\mu^0, \end{aligned}$$

and with r_k linear in k, i.e., $r_k = \alpha k$.

Proposition 6 *In this case, the optimal solution is :*

$$\begin{aligned} \lambda_i &= \lambda; \\ \mu_i &= i\mu, \end{aligned}$$

with

$$\begin{aligned} \lambda &= \left(\lambda^0\mu^0\frac{\alpha}{M+c}\right)^{\frac{1}{2}}; \\ \mu &= \left(\lambda^0\mu^0\frac{M+c}{\alpha}\right)^{\frac{1}{2}}. \end{aligned}$$

The entropy is given by :

$$H^* = (\lambda^0 - \lambda) + \frac{\lambda}{\mu}(\mu^0 - \mu). \tag{32}$$

Once again for this case the large deviation exponent can be found and is given by the relatively simple form that follows.

Proposition 7 *Suppose $\rho^0\alpha < c$, i.e. the system is stable. Then*

$$E = \frac{\mu^0}{\alpha} - \frac{\lambda^0}{c}. \tag{33}$$

This result is only true when $N = \infty$ but of course will hold for systems, processing a large number of calls N.

5 Numerical solution

We present herein an algorithm to compute the exponent and the effective bandwidth for the general case. Returning to equations (24), denote $q_k = \lambda_{k-1}^0\mu_k^0$ for $k = 1, \ldots, N-1$, and rewrite an equivalent set of equations :

$$H^*(c - r_0) + (\lambda_0^0 - \lambda_0)(\sum_{i=0}^{N-1}\pi_i r_i - c) = 0; \tag{34}$$

$$(\lambda_k^0 - \lambda_k) + \mu_k^0 - \frac{q_k}{\lambda_{k-1}} = \left(\frac{c - r_k}{c - r_0}\right)(\lambda_0^0 - \lambda_0),\ k = 0, \ldots, N-1. \tag{35}$$

The last series of equations may be rewritten as :

$$\lambda_k + \frac{q_k}{\lambda_{k-1}} = \lambda_k^0 + \mu_k^0 - \left(\frac{c - r_k}{c - r_0}\right)(\lambda_0^0 - \lambda_0) = \Delta_k(\lambda_0), \quad k = 0, \ldots, N-1.$$

By setting $\lambda_k = \frac{u_k}{v_k}$ for $k = 0, \ldots, N-2$, we obtain

$$\frac{u_k}{v_k} + q_k \frac{v_{k-1}}{u_{k-1}} = \Delta_k(\lambda_0), \tag{36}$$

which can be written in a matrix form as

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} \Delta_k(\lambda_0) & -q_k \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix} = A_k(\lambda_0) \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix} \tag{37}$$

where $A_k(\lambda^0)$ is 2×2 dimensional.

Then, by induction

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = A_k(\lambda_0) A_{k-1}(\lambda_0) \ldots A_1(\lambda_0) \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}.$$

For a given choice of λ_0, $\lambda_1, \ldots, \lambda_{N-2}$ may be computed recursively, and $\mu_1, \ldots, \mu_{N-1}$ can be obtained by way of equations (24). However a boundary condition must be satisfied. In particular, we require that $\lambda_{N-1} = 0$, to be consistent with our set-up. Thus, the unknown λ_0 must be a root of the equation

$$u_{N-1} = 0,$$

or

$$\begin{pmatrix} 1 & 0 \end{pmatrix} A_{N-1}(\lambda_0) A_{N-2}(\lambda_0) \ldots A_1(\lambda_0) \begin{pmatrix} \lambda_0 \\ 1 \end{pmatrix} = 0, \tag{38}$$

such that

$$\begin{aligned} \lambda_k &\geq 0, \quad k = 0, \ldots, N-1; \\ \sum_{i=0}^{N-1} \pi_i r_i &> c. \end{aligned}$$

This equation is in fact a polynomial of order N in λ_0, so that the algorithm reduces to the computation of the roots of this polynomial. Since we have established the existence and uniqueness of the solution, there can be only one root satisfying these conditions. Let λ_0^* be this root. The exponent is then obtained from equation (34) :

$$E = \frac{H^*}{\sum_{i=0}^{N-1} \pi_i r_i - c} = \frac{\lambda_0^* - \lambda_0^0}{c - r_0}. \tag{39}$$

We turn now to the computation of the effective bandwidth. This is defined [4] as the function $a(\delta)$ such that :

$$\frac{H^*}{\sum_{i=0}^{N-1} \pi_i r_i - a(\delta)} = \delta. \tag{40}$$

$a(\delta)$ can be interpreted as the service rate required to guarantee that the given source will have an asymptotic probability of overflow less than or equal to $e^{-B\delta}$.

From equations (34), we obtain

$$\delta = \frac{\lambda_0^* - \lambda_0^0}{a(\delta) - r_0}$$

or

$$\lambda_0^* = \delta(a(\delta) - r_0) + \lambda_0^0.$$

Then, a similar calculation to the above gives :

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} \lambda_k^0 + \mu_k^0 + \delta(a(\delta) - r_k) & -q_k \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix} = A_k(\delta, a(\delta)) \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix}, \quad (41)$$

and the equation determining $a(\delta)$ as a function of δ becomes

$$\begin{pmatrix} 1 & 0 \end{pmatrix} A_{N-1}(\delta, a(\delta)) A_{N-2}(\delta, a(\delta)) \ldots A_1(\delta, a(\delta)) \begin{pmatrix} \delta(a(\delta) - r_0) + \lambda_0^0 \\ 1 \end{pmatrix} = 0, \quad (42)$$

with the constraints

$$\begin{aligned} \lambda_k &\geq 0, k = 0, \ldots, N-1; \\ \sum_{i=0}^{N-1} \pi_i r_i &> c. \end{aligned}$$

6 Conclusions

The problem of computing the asymptotic probability of buffer overflow for a queueing system fed by N independent Markov fluids has been addressed. It has been shown that when an overflow occurs as a large deviation, independent sources actually deviate in an independent fashion, resulting in the notion of effective bandwidth discussed in [4]. Thus it suffices to analyze the single source case. Necessary and sufficient optimality conditions have been derived characterizing the deviant behavior, when the underlying Markov chain is of the birth death type. Closed-form solutions have been found, when the input is the aggregate of on/off sources, and when it corresponds to a system in which sources arrive as a Poisson process, and leave independently after an exponential period. In the former case, similar results have been obtained by different approaches. Numerical algorithms are suggested for computing the actual deviant behavior for the general case. These may be of interest for quick simulation. Further work is required on this topic.

References

[1] D. Anick, D. Mitra, and M.M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *Bell System Tech. Journal*, 61, 1982.

[2] J.A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation.* John Wiley and Sons, New York, NY, 1990.

[3] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications, 1991.

[4] G. Kesidis. *Cell Loss Estimation in High-Speed Digital Networks.* Ph.D. Thesis. Dept. of EECS, University of California, Berkeley, 1992.

[5] G. Kesidis and J. Walrand. Relative entropy between markov transition rate matrices. *to appear in IEEE Trans. Info Th.*

[6] M.Neuts. *Matrix-geometric solutions in stochastic models.* John Hopkins University Press, 1981.

[7] Alan Weiss. A new technique for analyzing large traffic systems. *Adv. Appl. Probab.*, 18:506–532, 1986.

Appendix

Proof of Lemma 1 : Define :

$$P_i = \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \ldots \mu_i}.$$

We note that

$$\begin{aligned} \frac{\partial P_i}{\partial \lambda_k} &= \frac{P_i}{\lambda_k} 1_{\{i-1 \geq k\}} \\ \frac{\partial P_i}{\partial \mu_k} &= -\frac{P_i}{\mu_k} 1_{\{i \geq k\}}. \end{aligned}$$

Since $\pi_i = \frac{P_i}{\sum_{j=0}^{N-1} P_j} = \frac{P_i}{P}$,

$$\begin{aligned} \frac{\partial \pi_i}{\partial \lambda_k} &= \frac{\left(\frac{\partial P_i}{\partial \lambda_k} P - P_i \frac{\partial P}{\partial \lambda_k} \right)}{P^2} \\ &= \frac{P_i}{P \lambda_k} 1_{\{i-1 \geq k\}} - \frac{P_i}{P} \sum_{j=0}^{N-1} \frac{P_j}{P \lambda_k} 1_{\{j-1 \geq k\}} \\ &= \frac{\pi_i}{\lambda_k} (1_{\{i-1 \geq k\}} - S_{k+1}). \end{aligned}$$

Derivatives with respect to μ_k follow in a similar fashion. □

Proof of Lemma 2 : We first compute the derivatives of the Lagrangian with respect to $\lambda_k, k = 0, \ldots, N-2$:

$$\begin{aligned} \frac{\partial L}{\partial \lambda_k} &= \sum_{i=0}^{N-1} \frac{\pi_i}{\lambda_k} (1_{i-1 \geq k} - S_{k+1})(\phi_i + \psi_i) + \pi_k \log(\frac{\lambda_k}{\lambda_k^0}) + K \sum_{i=0}^{N-1} \frac{\pi_i}{\lambda_k} (1_{i-1 \geq k} - S_{k+1}) r_i \\ &= -\frac{S_{k+1}}{\lambda_k} \left(\sum_{i=0}^{N-1} \pi_i (\phi_i + \psi_i) + \sum_{i=0}^{N-1} \pi_i r_i \right) + \frac{1}{\lambda_k} \left(\sum_{i=k+1}^{N-1} \pi_i (\phi_i + \psi_i + K r_i) \right) + \pi_k \log(\frac{\lambda_k}{\lambda_k^0}) \\ &= -\frac{S_{k+1}}{\lambda_k} (H^*(M) + K(M+c)) + \frac{\alpha_{k+1}}{\lambda_k} + \pi_k \log(\frac{\lambda_k}{\lambda_k^0}). \end{aligned}$$

In the same vein, for $k = 1, \dots, N-1$:

$$\frac{\partial L}{\partial \mu_k} = \frac{S_k}{\mu_k}(H^*(M) + K(M+c)) - \frac{\alpha_k}{\mu_k} + \pi_k \log(\frac{\mu_k}{\mu_k^0}).$$

So that we finally get the following system for $k = 0, \dots, N-2$ stated in lemma 2 :

$$\begin{aligned} -\frac{S_{k+1}}{\lambda_k}(H^*(M) + K(M+c)) + \frac{\alpha_{k+1}}{\lambda_k} + \pi_k \log(\frac{\lambda_k}{\lambda_k^0}) &= 0; \\ \frac{S_{k+1}}{\mu_{k+1}}(H^*(M) + K(M+c)) - \frac{\alpha_{k+1}}{\mu_{k+1}} + \pi_{k+1} \log(\frac{\mu_{k+1}}{\mu_{k+1}^0}) &= 0. \end{aligned}$$

□

Proof of Proposition 1 : By multiplying the first and second type of equation obtained in Lemma 1 by λ_k and μ_{k+1} respectively, and adding them together, we obtain

$$\pi_k \lambda_k \log \frac{\lambda_k}{\lambda_k^0} + \pi_{k+1}\mu_{k+1} \log \frac{\mu_{k+1}}{\mu_{k+1}^0} = 0.$$

Since

$$\pi_{k+1} = \pi_k \frac{\lambda_k}{\mu_{k+1}},$$

we have

$$\log \frac{\lambda_k}{\lambda_k^0} + \log \frac{\mu_{k+1}}{\mu_{k+1}^0} = 0.$$

In other words :

$$\lambda_k \mu_{k+1} = \lambda_k^0 \mu_{k+1}^0, k = 0, \dots, N-2. \tag{43}$$

We now look for another set of characteristic equations. There are obtained by subtracting two consecutive first equations (7) with indices k and $k+1$, to first get

$$(S_k - S_{k+1})(H^*(M) + K(M+c)) + (\alpha_{k+1} - \alpha_k) + \pi_k \lambda_k \log \frac{\lambda_k}{\lambda_k^0} - \pi_{k-1}\lambda_{k-1} \log \frac{\lambda_{k-1}}{\lambda_{k-1}^0} = 0.$$

Equations (43) yield

$$\log \frac{\lambda_{k-1}}{\lambda_{k-1}^0} = \log \frac{\lambda_{k-1}^0 \mu_k^0}{\lambda_{k-1}^0 \mu_k} = -\log \frac{\mu_k}{\mu_k^0},$$

and on the other hand

$$\pi_k = \pi_{k-1} \frac{\lambda_{k-1}}{\mu_k}.$$

So that the subtraction above reads

$$\pi_k (H^*(M) + K(M+c)) - \pi_k(\phi_k + \psi_k + K r_k) + \pi_k \left(\lambda_k \log \frac{\lambda_k}{\lambda_k^0} + \mu_k \log \frac{\mu_k}{\mu_k^0} \right) = 0.$$

Then, after substitution of ϕ_k and ψ_k by their expressions, we get

$$\lambda_k + \mu_k = -H^*(M) + K(r_k - (M+c)) + \lambda_k^0 + \mu_k^0, \quad k = 0, \dots, N-1, \tag{44}$$

with the usual convention $\lambda_{N-1}^0 = \lambda_{N-1} = \mu_0^0 = \mu_0 = 0$.

Returning to the definition of $H^*(M)$, we obtain

$$\begin{aligned}
H^*(M) &= \sum_{i=0}^{N-1} \pi_i \left(\lambda_i \log \frac{\lambda_i}{\lambda_i^0} + \lambda_i^0 - \lambda_i + \mu_i \log \frac{\mu_i}{\mu_i^0} + \mu_i^0 - \mu_i \right) \\
&= \sum_{i=0}^{N-2} \pi_i \left(\lambda_i \log \frac{\lambda_i}{\lambda_i^0} + \lambda_i - \lambda_i^0 \right) + \sum_{i=0}^{N-2} \pi_{i+1} \left(\mu_{i+1} \log \frac{\mu_{i+1}}{\mu_{i+1}^0} + \mu_{i+1}^0 - \mu_{i+1} \right) \\
&= \sum_{i=0}^{N-2} \left(\pi_i (\lambda_i^0 - \lambda_i) + \pi_{i+1} (\mu_{i+1}^0 - \mu_{i+1}) \right).
\end{aligned}$$

□

Proof of Proposition 2 : the existence and uniqueness of a single minimum :

Consider again the initial constrained minimization problem :

$$H^*(M) = \inf_{\{P | E_P r(\cdot) = M + c\}} H(P \parallel P^0). \tag{45}$$

To prove that the optimality equations derived above actually define a minimum that is moreover unique, change variables is required since there is no direct convexity result on H as a function of the set of parameters (λ_i, μ_j). Define first $\rho_i = \frac{\lambda_i}{\mu_{i+1}}$, for $i = 0, \ldots, N-2$. We see that, since

$$\pi_i = \pi_0 \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \ldots \mu_i} = \pi_0 \rho_0 \rho_1 \ldots \rho_{i-1}, \quad i \geq 1$$

with

$$\pi_0 = \left(1 + \sum_{i=1}^{N-1} \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \ldots \mu_i} \right)^{-1},$$

$$\rho_i = \frac{\pi_{i+1}}{\pi_i}, \quad i = 0, \ldots, N-2. \tag{46}$$

On the other hand

$$\lambda_i = \rho_i \mu_{i+1}, \quad i = 0, \ldots, N-2. \tag{47}$$

Then

$$H(\{\pi_i\}, \{\mu_{i+1}\}) = \sum_{i=0}^{N-1} \pi_i \left(\frac{\pi_{i+1}}{\pi_i} \mu_{i+1} \log \frac{\frac{\pi_{i+1}}{\pi_i} \mu_{i+1}}{\lambda_i^0} + \lambda_i^0 - \frac{\pi_{i+1}}{\pi_i} \mu_{i+1} + \mu_i \log \frac{\mu_i}{\mu_i^0} + \mu_i^0 - \mu_i \right),$$

the constraints being as follows :

$$\begin{aligned}
\sum_{i=0}^{N-1} \pi_i r_i &= M + c; \\
\sum_{i=0}^{N-1} \pi_i &= 1; \\
\pi_i &\geq 0, i = 0, \ldots, N-1,
\end{aligned}$$

defining a constraint set C_M.

The unconstrained minimization over the variables μ_{i+1} is performed without constraint and we have seen that, for a given set of $\{\pi_i\}$, the optimum is given by

$$\mu_{i+1} = \sqrt{\frac{\lambda_i^0 \mu_{i+1}^0}{\lambda_i}} = \sqrt{\frac{\lambda_i^0 \mu_{i+1}^0}{\frac{\pi_{i+1}}{\pi_i}}}, \quad i = 0, \ldots, N-2.$$

We know furthermore that this is the unique minimum in μ_{i+1} since the functions

$$\begin{aligned}
\phi_i(\mu_{i+1}) &= \frac{\pi_{i+1}}{\pi_i}\mu_{i+1} \log \frac{\frac{\pi_{i+1}}{\pi_i}\mu_{i+1}}{\lambda_i^0} + \lambda_i^0 - \frac{\pi_{i+1}}{\pi_i}\mu_{i+1} \\
\psi_{i+1}(\mu_{i+1}) &= \mu_{i+1} \log \frac{\mu_{i+1}}{\mu_{i+1}^0} + \mu_{i+1}^0 - \mu_{i+1}
\end{aligned}$$

are convex.

We replace the μ_{i+1} by their optimal values and turn to a constrained problem in π_i :

$$\begin{aligned}
H^*(\{\pi_i\}) &= \sum_{i=0}^{N-2} \pi_i \left(\sqrt{\lambda_i^0} - \sqrt{\frac{\pi_{i+1}}{\pi_i}\mu_{i+1}^0}\right)^2 \\
&= \sum_{i=0}^{N-2} \left(\sqrt{\lambda_i^0 \pi_i} - \sqrt{\pi_{i+1}\mu_{i+1}^0}\right)^2
\end{aligned}$$

under the same constraints as before.

The next step is to prove that this function is strictly convex in the variables $\{\pi_i\}$ on the set C_M. But in fact, H is a function of N variables that can be written a sum of functions of 2 variables. Each one of these functions is strictly convex in its two variables, so the sum is strictly convex in all the variables. $H^*(\{\pi_i\})$ is then strictly convex, and the constraint set C_M is convex. H^* then admits a unique local minimum that is at the same time the global minimum. And then the optimality equations (19) define a single minimum. □

Proof of Proposition 4 : For simplicity we consider the case $\beta = 0$. One can verify that the proposed solution :

$$\begin{aligned}
\lambda_i &= (N-1-i)\lambda; \\
\mu_i &= i\mu,
\end{aligned}$$

with

$$\begin{aligned}
\lambda &= \left(\lambda^0 \mu^0 \frac{(\alpha(N-1)-M)}{M}\right)^{\frac{1}{2}}; \\
\mu &= \left(\lambda^0 \mu^0 \frac{M}{(\alpha(N-1)-M)}\right)^{\frac{1}{2}},
\end{aligned}$$

satisfies the optimality equations (19). Then, noting that the mean $\overline{m}$ of such a birth death process is equal to

$$\overline{m} = (N-1)\frac{\lambda}{\lambda+\mu},$$

we find that :

$$
\begin{aligned}
H^*(M) &= \sum_{i=0}^{N-2} \pi_i\ (\lambda_i^0 - \lambda_i) + \pi_{i+1}\ (\mu_{i+1}^0 - \mu_{i+1}) \\
&= \sum_{i=0}^{N-2} \pi_i\ (N-1-i)(\lambda^0 - \lambda) + \pi_{i+1}\ (i+1)(\mu_{i+1}^0 - \mu_{i+1}) \\
&= (N-1)(\lambda^0 - \lambda)(1-\pi_{N-1}) - (\lambda^0 - \lambda)(\overline{m} - (N-1)\pi_{N-1}) + (\mu^0 - \mu)\overline{m} \\
&= (N-1)\left((\lambda^0 - \lambda)(1 - \frac{\lambda}{\lambda+\mu}) + (\mu^0 - \mu)\frac{\lambda}{\lambda+\mu}\right) \\
&= (N-1)\left(\frac{\lambda}{\lambda+\mu}(\mu^0 - \mu) + \frac{\mu}{\lambda+\mu}(\lambda^0 - \lambda)\right).
\end{aligned}
$$

□

Proof of Proposition 5 : Given the expression obtained above, and since

$$\lambda = \left(\lambda^0 \mu^0 \frac{M+c}{a-(M+c)}\right)^{\frac{1}{2}}$$

in this specific case, the ratio to be minimized becomes

$$\frac{H^*(M)}{M} = (N-1)\left((\frac{(M+c)\mu^0}{Ma})^{\frac{1}{2}} - (\frac{(a-(M+c))\lambda^0}{Ma})^{\frac{1}{2}} \right)^2 .$$

Cancelling the first derivative of this ratio with respect to M yields

$$\frac{1}{2}(\mu^0 \frac{M+c}{Ma})^{-\frac{1}{2}} \left(\frac{Ma-(M+c)a}{(Ma)^2}\right) = \frac{1}{2}(\lambda^0 \frac{a-(M+c)}{Ma})^{-\frac{1}{2}} \left(\frac{-Ma - a(a-(M+c))}{(Ma)^2}\right). \quad (48)$$

Simplifying on both sides and taking the squares, we find that the optimal M is given by

$$M + c = \frac{a}{1+\rho^0 \frac{(a-c)^2}{c^2}}. \quad (49)$$

By substitution we find the exponent :

$$
\begin{aligned}
E &= \frac{N-1}{(a-c) - \rho^0 \frac{(a-c)^2}{c}} \mu^0 \left(1 - (\rho^0)^{\frac{1}{2}} \frac{a-c}{c}\right)^2 \\
&= (N-1)\frac{((\lambda^0 + \mu^0)(c) - \lambda^0 a)}{c(a-c)}.
\end{aligned}
$$

In the case where β is nonzero one can simply replace c by $c - \beta$. □

Proof of Proposition 6 : One can easily verify that the proposed solution :

$$
\begin{aligned}
\lambda_i &= \lambda, \\
\mu_i &= i\mu,
\end{aligned}
$$

with

$$
\begin{aligned}
\lambda &= \left(\lambda^0 \mu^0 \frac{M+c}{\alpha}\right)^{\frac{1}{2}}, \\
\mu &= \left(\lambda^0 \mu^0 \frac{\alpha}{M+c}\right)^{\frac{1}{2}},
\end{aligned}
$$

satisfies the optimality equations in Proposition 1. Using the fact that the invariant distribution for an $M/M/\infty$ process is Poisson with parameter $\rho = \frac{\lambda}{\mu}$, and hence mean ρ we find that

$$\begin{aligned} H^*(M) &= \sum_{i=0}^{\infty} \pi_i \; (\lambda_i^0 - \lambda_i) + \pi_{i+1} \; (\mu_{i+1}^0 - \mu_{i+1}) \\ &= \sum_{i=0}^{\infty} \pi_i \; (\lambda^0 - \lambda) + \pi_{i+1} \; i(\mu^0 - \mu) \\ &= \pi_i \; (\lambda^0 - \lambda) + \frac{\lambda}{\mu}(\mu^0 - \mu). \end{aligned}$$

□

Proof of proposition 7 : Since

$$H^*(M) = \lambda^0 + \mu^0 \frac{M+c}{\alpha} - 2\sqrt{\frac{\lambda^0 \mu^0 (M+c)}{\alpha}},$$

we must minimize

$$\frac{H^*(M)}{M} = \frac{\left(\sqrt{\lambda^0} - \sqrt{\mu^0 \frac{(M+c)}{\alpha}}\right)^2}{M}.$$

Cancelling the first derivative of this ratio with respect to M yields

$$M + c = \frac{c^2}{\alpha \rho^0}.$$

By substituting one obtains the exponent :

$$E = \inf_{M>0} \frac{H^*(M)}{M} = \frac{\mu^0}{\alpha} - \frac{\lambda^0}{c}.$$

□