Output Analysis of Multiclass Fluid Models with Static Priorities

June 2, 2005

Abstract

We consider a stochastic fluid flow model with a single server and K infinite capacity buffers. The input to the k-th buffer is a Markovian on-off process that transmits fluid at a constant rate p_k while it is on and at rate 0 while it is off. The fluid is emptied from the buffers by a single server at a constant rate μ according to a *static priority service discipline* in which class 1 fluid has the highest and class K fluid has the lowest priority. The output process of class k is defined to be *on* if fluid of class k is leaving the buffer at a positive rate and *off* otherwise. In this paper we derive an exact method for computing the mean on-time and the mean off-time of the output process of class k. We illustrate the techniques by numerical results.

1 Introduction

We study a fluid-flow model with a single server and incoming fluid flows generated by K independent on-off Markovian input sources. The k-th source has $\text{Exp}(\alpha_k)$ on-times and $\text{Exp}(\beta_k)$ off-times, $k = 1, \ldots, K$. It generates fluid at rate p_k while it is on and at rate 0 while it is off. Thus, the input process of class k is completely described by three parameters (α_k, β_k, p_k). The fluid generated by the K on-off sources is stored in K separate infinite capacity buffers from where it is removed according to a *static priority* service discipline under which the highest priority fluid always takes precedence over any of the lower priority fluids. We assume that class 1 fluid has the highest priority and class K has the lowest priority. Then class k enjoys complete priority over fluid of class $j > k$, at all times. The leftover service capacity after serving the fluids of classes $1, 2, \ldots, k$ is available to serve fluids of class $k + 1$ and above.

The output process under the static priority service discipline is rather complex. Several input bursts of a given class may combine in one output burst. Similarly, a single input burst may get split into several output bursts due to interruptions by higher priority fluid coming to the buffer. The output processes of different classes of fluid are neither independent, nor on-off. The rate during an output burst is not constant. The idea is to approximate the class k output process by a three-parameter Markov on-off process with parameters $(\alpha_k^o, \beta_k^o, p_k^o)$ which involves finding the mean on- and off-times. The output process of class k is defined to be *on* if fluid of class k is leaving the buffer at a positive rate and *off* otherwise. We approximate the non-constant output rate of class k by the mean peak rate p_k^o . A related work is [10] of Kulkarni and Glazebrook, which provides the output analysis for a single-buffer multi-class queue with First-Come-First-Serve (FCFS) discipline. Mean on- and off-times of a given class are found using appropriately constructed reward processes. However, in our experience this approach does not readily extend to the case of static priority service discipline.

The motivation behind this analysis arises from the study of telecommunication networks as multi-class fluid networks (MFN) described by a set of nodes and a set of different classes of fluid. Each node has an infinite capacity buffer for storing each class of fluid that enters this node. The input of each node may consist of fluid generated by an external environment process and/or fluid coming from other nodes within the network. Once in a given node the different classes of fluid are served according to a predefined service discipline, in our case static priority. The external fluid inputs to each node in the network are assumed to be generated by independent Markovian on-off sources. We approximate the output processes from a given node as three-parameter independent on-off processes by using the approach developed here. The analysis of the network can then proceed recursively by using these approximated output on-off processes as inputs to other nodes in the same spirit as in Whitt, [15]. Thus, each node acts as a non-linear mapping of the input parameters to the output parameters. In [3] Hirasawa adopts a similar approach where he studies a multi-class fluid network (MFN) with First-Come-First-Serve (FCFS) discipline and develops a MFN algorithm based on the parametric-decomposition method of Kuehn [9]. Hirasawa characterizes the network traffic in terms of four parameters - mean rate, effective peak rate, mean burst length, and mean squared burst length. In this paper we develop the output approximations for the single-node model. However, the network extension will be a topic for future research.

Several authors have studied specially structured fluid networks. In [7] Kella and Whitt study a tandem fluid network with Lévy Input and in which they analyze the mean buffer content of each node. In [5] Kella considers parallel and tandem fluid networks with dependent Lévy inputs. He derives the Laplace Stieltjes Transform of the limiting distribution of the fluid content process. Kaspi and Kella [4] study the stability of feed-forward fluid networks with Lévy input. These results have been further extended by Kella [6] where he studies the stability and non-product form of stochastic fluid networks with Lévy inputs. In [8] Kella and Whitt introduce linear stochastic fluid networks as continuous analogues of open networks of infinite-server queues. As with infinite-server queues, the tractability makes the linear stochastic fluid networks appealing for approximations. Generally, exact analysis of queueing networks is intractable and only approximate results are available, see Kuehn [9], Whitt [15], Reiser and Kobayashi [12], Gelenbe and Mitrani [2], Chandy and Sauer [1].

2 Problem Description

Denote the state of source k at time t by $I_k(t)$, where

$$I_k(t) = \begin{cases} 0, & \text{if source } k \text{ is off at time } t, \\ 1, & \text{if source } k \text{ is on at time } t. \end{cases}$$

The combined state of the K sources at time t (called the environment) is given by

$$I(t) = (I_1(t), \ldots, I_K(t)),\ t \geq 0.$$

Thus, $\{I(t),\ t \geq 0\}$ is an irreducible CTMC on the finite state space

$$S = \{i = (i_1, \ldots, i_K) : i_j = 0, 1,\ j = 1, \ldots, K\}.$$

Define the combined input rate $p(i)$ of all sources in state $i = (i_1, \ldots, i_K) \in S$ as

$$p(i) = \sum_{k=1}^{K} i_k p_k.$$

The server operates at a constant rate μ, also called its capacity. The net input rate $r(i)$ to the buffer when the environment is in state $i = (i_1, \ldots, i_K) \in S$ is given by

$$r(i) = \sum_{k=1}^{K} i_k p_k - \mu = p(i) - \mu.$$

Furthermore, let $R = \text{diag}(r(i),\ i \in S)$ denote the net input rate matrix.

Let $X_k(t), k = 1, \ldots, K$, denote the amount of fluid of class k in the buffer at time t and $X(t) = \sum_{i=1}^{K} X_i(t)$ be the total amount of fluid in the buffer at time t. Then $\{(I(t), X(t)), t \geq 0\}$ is a Markov process. The rate of change of the fluid level in the buffer $\{X(t), t \geq 0\}$ is given by

$$\frac{d}{dt} X(t) = \begin{cases} r(i) & \text{if } I(t) = i,\ X(t) > 0, \\ \\ \max(r(i), 0) & \text{if } I(t) = i,\ X(t) = 0. \end{cases}$$

Now let

$$\pi(i) = \lim_{t \to \infty} P(I(t) = i),\ i \in S$$

be the limiting distribution of the governing CTMC $\{I(t), t \geq 0\}$. The system is stable if and only if the expected net input rate is negative in steady state,

$$\sum_{i \in S} \pi(i) r(i) < 0.$$

We assume that the system is stable so that the limiting distribution of the bivariate process $\{(I(t), X(t)), t \geq 0\}$ exists. Let us denote it by

$$\pi(i, x) = \lim_{t \to \infty} P(I(t) = i, X(t) \leq x),\ i \in S, \tag{2.1}$$

e.g. see Kulkarni [11] for methods of computing $\pi(i, x)$.

In this paper we study the output (service) process of class k defined by

$$S_k(t) = \begin{cases} 1, & \text{if fluid of class } k \text{ is leaving the buffer at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

More precisely, let B_k^K, C_k^K denote the mean sojourn times of the S_k process in states 1, 0, respectively, if there are K input sources in the system, i.e. the mean on-time, off-time, of the output process of class k. Due to the static priority service discipline, the presence of class K fluid in the system will not have any effect on the higher priority service. Hence

$$B_i^K = B_i^i, \; i = 1, \ldots, K.$$

The same reasoning applies to the mean output off-times C_i^K. Hence we shall skip the superscript when the number of sources in the discussed system is clear. In this paper we derive $B_i = B_i^i$, $C_i = C_i^i$. The output process of class k can then be approximated by a Markovian on-off process with $\exp(\alpha_k^o)$ on-times and $\exp(\beta_k^o)$ off-times, where $\alpha_k^o = 1/B_k$ and $\beta_k^o = 1/C_k$, and constant transmission rate p_k^o approximated by the use of the conservation law of fluid, i.e. the mean output rate of class k should equal the mean input rate of class k,

$$\frac{p_k \beta_k}{\alpha_k + \beta_k} = \frac{p_k^o \beta_k^o}{\alpha_k^o + \beta_k^o},$$

and therefore

$$p_k^o = \frac{p_k \beta_k}{\alpha_k + \beta_k} \frac{(\alpha_k^o + \beta_k^o)}{\beta_k^o}, \; k = 1, 2, \ldots, K.$$

The problem of finding the mean out on- and off-times is trivial if $\sum_{k=1}^K p_k \le \mu$ since the output process of class k will be exactly equal to the input process of class k for every $k = 1, \ldots, K$. Therefore we assume that $\sum_{i=1}^{K} p_i > \mu$.

The static priority rule leads to a recursive solution in the case of $K > 2$ classes of fluid. From the *original* K-class system we construct a new 2-class *aggregated* system as follows. Class 1 fluid input process in the aggregated system is identical to the superposition of the input processes of the first $K-1$ on-off sources in the original system. Class 2 fluid input process in the aggregated system is identical to the input process of the K-th on-off source in the original system. Thus, the class 1 input in the aggregated system is modulated by a CTMC $\{I_1^a(t) = (I_1(t), \ldots, I_{K-1}(t)), t \ge 0\}$ that can be in 2^{K-1} different states $\{i = (i_1, \ldots, i_{K-1}), i_j = 0, 1, \; j = 1, \ldots, K-1\}$ and the class 2 input is modulated by the two state CTMC $\{I_2^a(t) = I_K(t), t \ge 0\}$. We continue with the analysis of this two-source priority model. It is convenient to introduce the input rates

$$p_1(i) := \sum_{k=1}^{K-1} i_k p_k, \text{ and } p_2(i) := i_K p_K, \; i \in S,$$

and the net input rate of class 1 fluid to the buffer

$$r_1(i) := p_1(i) - \mu,$$

when the environment is in state $i = (i_1, \dots, i_K) \in S$. Then $r_1(i)$ determines the following partitioning of the state space S,

$$\begin{aligned} S'_- &:= \{i \in S : r_1(i) < 0\},\ N'_- := |S'_-|, \\ S'_+ &:= \{i \in S : r_1(i) \geq 0\},\ N'_+ := |S'_+|. \end{aligned}$$

We refer to the periods of time when there is (is not) positive server capacity and class 2 fluid can (can not) be served, i.e. $i \in S'_-$ ($i \in S'_+$), as *on-periods* (*off-periods*). First, we consider the easier case of $S'_+ = \emptyset$ for which the solution can be found directly, by applying the theory of Alternating Renewal Processes. Then we study the more complicated case of $S'_+ \neq \emptyset$.

3 Output analysis if S'_+ is empty

It is clear that under this condition there is always some leftover capacity of the server at which the lowest priority class of fluid K is served. In other words, the moment source K turns on, fluid of class K starts immediately leaving the buffer with a rate that depends on the state of the environment at that moment. Consider the S_K process as defined in equation (2.2) for $k = K$, describing the output process of class K. Let us assume that it is off at time 0, i.e. $S_K(0) = 0$. It stays off for an $\mathrm{Exp}(\beta_K)$ and then it turns on as soon as source K turns on. Then it stays on for some random amount of time (as long as there is class K fluid passing through the buffer) which depends on the state of the environment, and thus depends on the off time of the S_K process. Therefore, the S_K process is an alternating renewal process and we obtain

$$\lim_{t\to\infty} P(S_K(t) = 0) = \frac{\dfrac{1}{\beta_K}}{\dfrac{1}{\beta_K} + B_K^K},$$

where B_K^K is the mean on-time of the output process of class K. The left-hand side of this equation can be computed as

$$\lim_{t\to\infty} P(S_K(t) = 0) = \sum_{i:\ i_K = 0} \pi(i, 0),$$

where $\pi(i, 0)$ is defined in Eq. (2.1). Hence we obtain

$$B_K^K = \frac{1 - \sum_{i:\ i_K=0} \pi(i,0)}{\beta_K \sum_{i:\ i_K=0} \pi(i,0)}.$$

Clearly we also have

$$C_K^K = \frac{1}{\beta_K}.$$

4 Output Analysis if S'_+ is non-empty

Assume that the fluid in the buffer is generated by only two independent input sources (with the second lower priority source being on-off and the first higher priority source being in any of the 2^{K-1} possible states as described above). Let $\gamma_k,\ k=1,2$, be the long-run fraction of time class k fluid is not being served, i.e. is not leaving the buffer. Consider the following reward structure for a fluid of class $k,\ k=1,2$: a unit reward is earned every time the output process S_k as defined above switches from 0 to 1. Denote by ν_k the long-run average reward of class k (or equivalently, ν_k is the mean number of class k output off-periods per unit time). Then from the classic theory of Markov-Regenerative processes it follows that

$$B_k + C_k = \frac{1}{\nu_k},\ k = 1,2,$$

and

$$\gamma_k = \frac{C_k}{C_k + B_k},\quad 1-\gamma_k = \frac{B_k}{C_k+B_k},\ k=1,2.$$

Hence

$$C_k = \frac{\gamma_k}{\nu_k},$$

and

$$B_k = \frac{1-\gamma_k}{\nu_k}.$$

Thus, to find B_k and C_k we need to determine γ_k and ν_k. We calculate them, first, for $k=1$, and then for $k=2$.

Let $\pi_i^1(x) = \lim_{t\to\infty} P(X_1(t)\le x, I(t)=i)$ for $x\ge 0$. Then we immediately have:

Theorem 4.1

$$\gamma_1 = \sum_{i:p_1(i)=0} \pi_i^1(0)$$

and

$$\nu_1 = \sum_{i:p_1(i)=0} \pi_i^1(0) \sum_{j:p_1(j)>0} q_{ij}.$$

Under the assumption of this section, $S'_+ \neq \emptyset$, there are off-periods alternating with on-periods for class 2 fluid. Next, we evaluate γ_2 and ν_2 by implementing earlier results on the embedded stochastic processes that describe the evolution of the X_2 and I processes during on-periods (skipping the off-periods). More precisely, for a given $t\ge 0$ denote by $\tau(t)$ the time spent in on-periods over $[0,t]$ and define the restricted processes $\{X_2^{on}(t),\ t\ge 0\} := \{X_2(\tau(t)+),\ t\ge 0\}$ and $\{I^{on}(t),\ t\ge 0\} := \{I(\tau(t)+),\ t\ge 0\}$. Clearly, $I^{on}(t)\in S'_-$, for all $t\ge 0$. Then $\{X_2^{on}(t),\ t\ge 0\}$ is a *fluid process with jumps*, as analyzed in Tzenova et al. [13], where the jump sizes correspond to the total amount of class 2 fluid accumulated in the buffer during the skipped off-periods, see Figure 4.1. Let

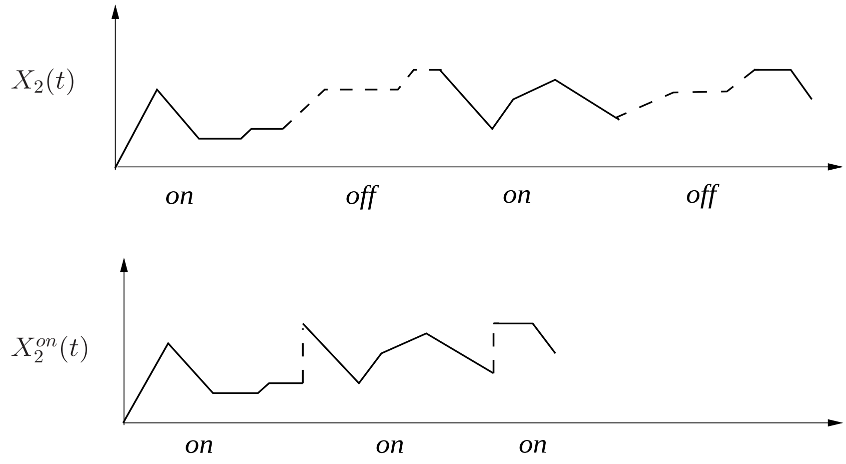

Figure 4.1: Class 2 buffer content process $X_2(t)$ embedded over on-periods and off-periods.

$$F_i(t,x) := P(X_2^{on}(t) \le x,\ I^{on}(t) = i),\ t \ge 0,\ x \ge 0, i \in S'_-,$$

$$F_i(x) := \lim_{t\to\infty} F_i(t,x),\ x \ge 0, i \in S'_-,\ F(x) := [F_i(x),\ i \in S'_-].$$

The analysis of the class 2 output process uses $F_i(0), i \in S'_-$, as computed in Theorem 2.4 of [13] which in return needs the LST of the jump sizes $\tilde{Q}_{ji}(s)$. To this end we study

$$\tilde{\psi}_{ji}(x,s) := E(e^{-sA_2(T)}; I(T) = i | I(0) = j, X_1(0) = x)\ i \in S'_-, \tag{4.3}$$

where

$$T := \inf\{t \ge 0 :\ X_1(t) = 0 \text{ and } I(t) \in S'_-\},$$

and $A_2(T)$ denotes the total amount of class 2 fluid that comes in the buffer during $[0,T]$. Then clearly the LST of the jump sizes can be found as

$$\tilde{Q}_{ji}(s) = \tilde{\psi}_{ji}(0,s) := E(e^{-sA_2(T)}; I(T) = i | I(0) = j, X_1(0) = 0).$$

For the purposes of the following results we also define the net input matrix for class 1 fluid,

$$R_1 := \text{diag}(r_1(i),\ i \in S).$$

Lemma 4.2 *For a fixed* $i \in S'_-$ *the LST* $\tilde{\psi}_i(x,s) := [\tilde{\psi}_{1i}(x,s), \ldots, \tilde{\psi}_{Ni}(x,s)]^t$ *satisfies the following system of differential equations*

$$R_1 \frac{\partial \tilde{\psi}_i}{\partial x}(x,s) + (Q - sP_2)\tilde{\psi}_i(x,s) = 0 \tag{4.4}$$

with boundary conditions

$$\tilde{\psi}_{ji}(0,s) = \delta_{ji},\ \textit{for } j \in S'_-, \tag{4.5}$$

where δ_{ji} is the Kronecker symbol and $P_2 := diag[p_2(1), \ldots, p_2(N)]$.

Let $\lambda_k(s)$, $\phi_k(s)$ denote the eigenvalues and eigenvectors to $(\lambda_k(s)R_1+Q-sP_2)\phi_k(s)=0$. Then the solution to (4.4) with boundary conditions (4.5) is given by

$$\tilde{\psi}_i(x,s) = \sum_{k:Re(\lambda_k(s))\le 0} a_k^i \phi_k(s) e^{\lambda_k(s)x},$$

where the coefficients a_k are determined as the solution to the linear system

$$\tilde{\psi}_i(0,s) = \sum_{k:Re(\lambda_k(s))\le 0} a_k^i \phi_k(s) = \delta_{ji},\ \ j \in S'_-$$

Proof: Let $x>0$, $j\in S$ and $i \in S'_-$. After conditioning on a small time interval of length $h>0$ we have

$$\tilde{\psi}_{ji}(x,s) = \sum_{k\ne j} q_{jk} h E(e^{-s(A_2(T)+p_2(j)h)}; I(T)=i|I(0)=k,\ X_1(0)=x+r_1(j)h)$$

$$+(1+q_{jj}h+o(h))E(e^{-s(A_2(T)+p_2(j)h)}; I(T)=i|I(0)=j,\ X_1(0)=x+r_1(j)h)+o(h).$$

Using the notation of (4) and rearranging the last equation we get

$$\frac{e^{sp_2(j)h}\tilde{\psi}_{ji}(x,s)-\tilde{\psi}_{ji}(x+r_1(j)h,s)}{h} = \sum_{k\in S} q_{jk}\tilde{\psi}_{ki}(x+r_1(j)h,s)+o(1).$$

Next, we substitute $e^{sp_2(j)h}=1+sp_2(j)h+o(h)$ and let $h\to 0$ to get

$$-r_1(j)\frac{\partial \tilde{\psi}_{ji}}{\partial x}(x,s) + sp_2(j)\tilde{\psi}_{ji}(x,s) = \sum_{k\in S} q_{jk}\tilde{\psi}_{ki}(x,s).$$

In vector notation this equation is equivalent to (4.4). The boundary conditions (4.5) follow from the definition of $A_2(T)$. Given $X_1(0)=0$ and $I(0)\in S'_-$ it is clear that the length of the off-period is 0 and therefore $A_2(T)=0$. The solution to (4.4) with boundary conditions (4.5) follows by well known results from the classical theory of linear differential equations. $\Diamond$

In addition, the analysis of class 2 output process involves the result of the following lemma where we compute the probability of positive increase (a jump of the embedded process) of the buffer content of class 2 during an off-period. Clearly, zero jumps are also possible if there is no incoming fluid of class 2 during an off-period. More precisely, we are interested in

$$g_j(x) := P(A_2(T)>0|I(0)=j, X_1(0)=x),\ j\in S,\ x\ge 0.$$

For a given $j\in S$ and $x>0$ it is clear that if $p_2(j)>0$ then $g_j(x)=1$. Therefore, we assume that j is such that $p_2(j)=0$. The following notation will be used,

$$S_{20} := \{i\in S: p_2(i)=0\},\ N_{20} := |S_{20}|, \text{and } S_{2+} := \{i\in S: p_2(i)>0\},\ N_{2+} := |S_{2+}|,$$

$$g(x) := [g_j(x),\ j \in S_{20}]^t, \quad g' := \left[\frac{dg_j(x)}{dx},\ j \in S_{20}\right]^t,$$

where the superscript t denotes transposition of a vector. For a given matrix M and subsets of indices A, B, we denote the sub-matrix

$$M_{A,B} := [M_{ij}, i \in A, j \in B].$$

The following lemma gives $g(x)$ as a solution to a system of linear ordinary differential equations. We omit the proof since it follows along similar lines to the one of Lemma 4.2.

Lemma 4.3 *The column vector $g(x)$ satisfies*

$$(R_1)_{S_{20},S_{20}} g'(x) + Q_{S_{20},S_{20}} g(x) + Q_{S_{20},S_{2+}} e = 0, \tag{4.6}$$

with boundary conditions

$$g_j(0) = 0,\ j \in S'_-. \tag{4.7}$$

The solution is given by

$$g(x) = \sum_{i:Re(\lambda_i)\le 0} a_i e^{\lambda_i x} \phi_i + e,$$

where (λ_i, ϕ_i) are the eigenvalues and eigenvectors satisfying

$$(\lambda_i (R_1)_{S_{20},S_{20}} + Q_{S_{20},S_{20}})\phi_i = 0,$$

e is a column vector of ones of size N_{20}, and the coefficients a_i are computed from the boundary conditions as the solution to

$$\sum_{i:Re(\lambda_i)\le 0} a_i \phi_{ji} + 1 = 0,\ j \in S'_-.$$

Remark: For $j \in S,\ i \in S'_-,\ x \ge 0$ we note that $g_{ji}(x) := P(A_2(T) > 0, I(T) = i | I(0) = j, X_1(0) = x)$ can be computed as

$$P(I(T) = i | I(0) = j, X_1(0) = x) - P(A_2(T) = 0, I(T) = i | I(0) = j, X_1(0) = x)$$
$$= \alpha_{ji}(x) - \lim_{s\to\infty} \tilde{\psi}_{ji}(x, s).$$

Here

$$\alpha_{ji}(x) := P(I(T) = i | I(0) = j, X_1(0) = x)$$

and can be easily found by standard conditioning arguments as the solution to a system of differential equations similar to that of Lemma 4.2. Then $g_j(x) = \sum_{i\in S'_-} g_{ji}(x)$.

We can now obtain the expressions for γ_2 (the long-run fraction of time class 2 fluid is not being served) and ν_2 (the mean number of class 2 output off-periods per unit time) as given in the following Theorem.

Theorem 4.4 *Let* $\gamma = \sum_{i \in S'_-} \pi_i^1(0)$ *be the long-run fraction of time the system is in an on-period.*
Then

$$\gamma_2 = \gamma \sum_{i \in S'_-, p_2(i)=0} F_i(0) + 1 - \gamma,$$

and

$$\nu_2 = \gamma A,$$

where

$$\begin{aligned} A &:= \sum_{i \in S'_-,\ p_2(i)=0} F_i(0) \sum_{j \in S'_-, p_2(j)>0} q_{ij} + \sum_{i \in S'_-, p_2(i)>0} F_i(\infty) \sum_{j \in S'_+} q_{ij} + \\ &+ \sum_{i \in S'_-, p_2(i)=0} (F_i(\infty) - F_i(0)) \sum_{j \in S'_+} q_{ij} + \sum_{i \in S'_-, p_2(i)=0} F_i(0) \sum_{j \in S'_+, p_2(j)=0} q_{ij} g_j(0), \end{aligned}$$

is the rate at which class 2 off-periods are generated per time unit given that the system is in an on-period.

Proof: During periods of time with $X_1(t) > 0$ or $X_1(t) \geq 0$ and $I(t) \in S'_+$ (referred to as off-periods) class 2 is not served since there is no leftover service capacity available. The long-run fraction of time spent in off-periods is given by $1 - \gamma$. Class 2 is also not served during on-periods, i.e. during periods of time $X_1(t) = 0$ and $I(t) \in S'_-$, when there is no class 2 fluid in the buffer and there is no inflow of class 2. Therefore γ_2, the fraction of time class 2 fluid is not being served, is given by

$$\gamma_2 = \gamma \sum_{i \in S'_-, p_2(i)=0} F_i(0) + 1 - \gamma.$$

To derive the expression for ν_2 we note that output of class 2 is only possible while the system is in an on-period, i.e. there is leftover service capacity to serve class 2. Then the output bursts of class 2 alternate with class 2 output off-periods. The long-run fraction of time the system is in on-periods is given by γ. Given that the system is in an on-period the output bursts of class 2 can finish in four possible ways corresponding to the four terms in the expression for A as follows: The first term of A accounts for off-times of class 2 that end within an on-period due to jumps from a state in which $X_2 = 0$ and $p_2(i) = 0$ to a state $j \in S'_-$ with $p_2(j) > 0$. The second term of A counts the ends of class 2 off-times that result from a jump into a state $j \in S'_+$ from a state $i \in S'_-$ with $p_2(i) > 0$. The third term of A arises again from interruptions of the higher priority fluid while there is positive amount of class 2 fluid in the buffer during the on-period. Finally, the fourth term of A represents the cases when there is no class 2 in the buffer and it is not coming into the buffer (i.e. $p_2(i) = 0$) at the moment of interruption of class 1 fluid, i.e. when $I(t)$ jumps to a state in S'_+ but by the end of the off-period there is positive amount of class 2 in buffer.

Thus, we obtain the expression for the mean number of class 2 output off-periods per unit time, $\nu_2 = \gamma A$. $\diamondsuit$

5 Numerical Results

We illustrate the developed methodology in the case of four independent input sources with input rates

$$p = [p_1, p_2, p_3, p_4] = [8, 3, 10, 2],$$

and identical exponential on-times and off-times with respective parameters $\alpha_i = 4, \beta_i = 1, i = 1, \ldots, 4$. In Figures 5.2 and 5.3 we vary the service capacity μ and plot the corresponding mean output busy periods $B_i,\ i = 1, \ldots, 4$. The difference of the two figures is in the values of μ for which the points are plotted so that the sizes of the jumps of B_i for each i can be viewed more easily. Figure 5.2 also contains the numerical values of each $B_i, i = 1, \ldots, 4$. The system becomes unstable for

$$\mu \leq \sum_{i=1}^{4} \frac{p_i \beta_i}{\alpha_i + \beta_i} = \frac{23}{5} = 4.6.$$

Thus, we vary μ from 4.75 to $23.02 > \sum_{k=1}^{4} p_k$.

With the increase of μ, B_1 decreases as expected and it becomes equal to the mean on-time of source 1, $1/\alpha_1 = 0.25$ for values of $\mu > p_1$. Thus, the observed pattern of the first (highest priority) busy period is not surprising. The second priority shows a somewhat different behavior. When μ becomes larger than 8, there is a significant increase of B_2 after which it starts decreasing. The explanation lies in the fact that for service capacity that is less than $p_1 = 8$ even though $p_2 = 3$ is relatively small, the second class of fluid gets interrupted by the first, every time source 1 turns on. This does not happen for values of $\mu > 8$ that can handle the service of the first class and also provide leftover capacity for class 2. Class 3 being of a lower priority is getting interrupted more often. The leftover capacity for its service depends on class 1 and class 2 service requirements. This leads to two increases of B_3, when μ becomes greater than $p_1 = 8$ and when μ becomes greater than $p_1 + p_2 = 11$. Overall, larger values of B_3 are observed since it is transmitted at a much larger rate $p_3 = 10$ which leads to its significant accumulation in the buffer. The observed values of B_4 show the complexity of the problem. Being of the lowest priority, class 4 fluid is affected by the first three flows. It is accumulated in the buffer for values of $\mu \leq \sum_{k=1}^{3} p_k = 21$ after which as μ becomes closer to $\sum_{k=1}^{4} p_k = 23$ it approaches the mean input on-time of source 4, $1/\alpha_4 = 0.25$. Two differences in the behavior of B_4 in comparison to that of B_2 and B_3 are observed. First, the values of B_4 stay below 0.25 from one point on and there are no such big increases observed as in the cases of B_2 and B_3. A possible explanation is in the particular values of the parameters. The source mean on-times are 4 times shorter than the mean off-times and also the input rate of source 4 is relatively low ($p_2 = 2$).

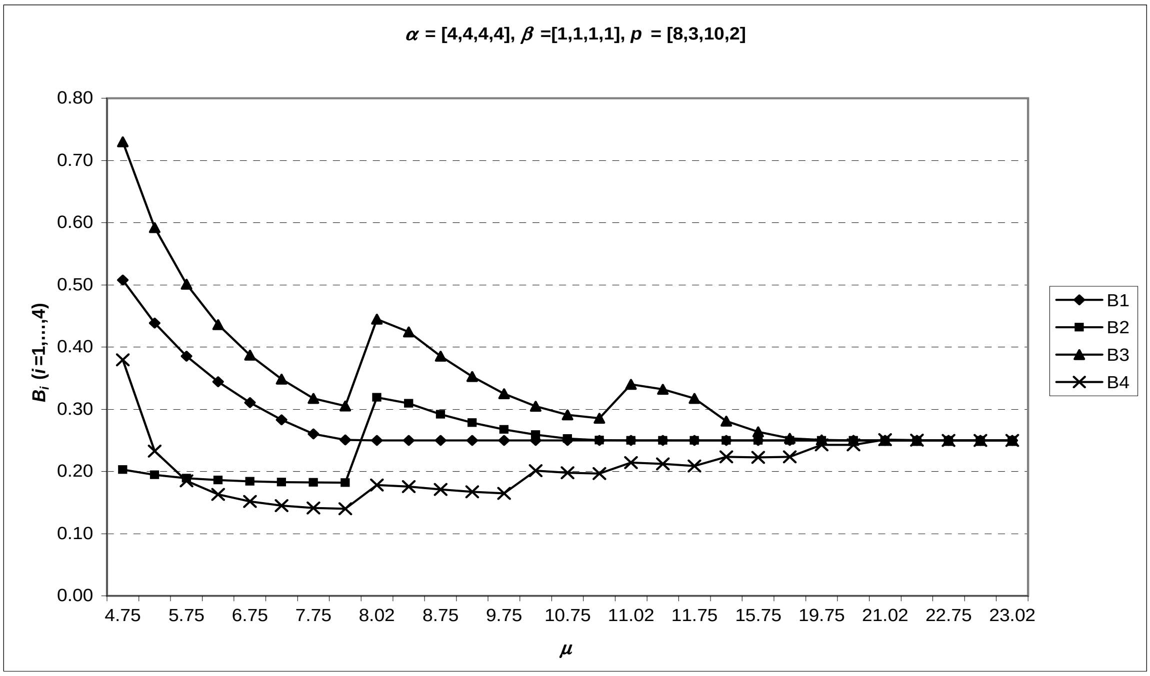

Figure 5.2: Output mean busy periods $B_i,\ i = 1, \dots, 4$ as μ varies from 4.75 to 23.02

References

[1] Chandy, K. M. and C. H. Sauer. (1978) Approximate models for analyzing queueing network models of computing systems. *ACM Computing surveys*, 10, pp 281-317

[2] Gelenbe, E. and I. Mitrani. (1980) Analysis and Synthesis of Computer Systems. *Academic Press, New York.*

[3] Hirasawa, Y. (2000). Approximating Traffic Parameters in Multi-Class Fluid Networks. *Ph.D. Thesis, Department of Operations Research, University of North Carolina, Chapel Hill, N.C. 27599-3180*

[4] Kaspi, H. and O. Kella. (1996) Stability of feed-forward fluid networks with Lévy input. *J. Appl. Prob.,* 33, pp 513-532.

[5] Kella, O. (1993) Parallel and tandem fluid networks with dependent Lévy inputs. *Annals. Appl. Prob.*, 3, pp 682-695.

[6] Kella, O. (1996) Stability and non-product form of stochastic fluid networks with Lévy inputs *Annals. Appl. Prob.*, 6, pp 186-199.

[7] Kella, O. and W. Whitt. (1990) A tandem fluid network with Lévy input. *Technical report 90-16, Yale University, CT 06511.*

[8] Kella, O. and W. Whitt. (1999) Linear stochastic fluid networks. *J. Appl. Prob.,* 36, pp 244-260.

[9] Kuehn, P. J. (1979) Approximate analysis of general queueing networks by decomposition. *IEEE Trans. Commun.,* COM-27, pp 113-126.

[10] Kulkarni V. G. and K. D. Glazebrook. (2002) Output analysis of a single-buffer multi-class queue: FCFS service. *J. Appl. Probab.*, 39, No. 2, pp 341-358

[11] Kulkarni V. G. (1997). Fluid models for single buffer systems. *Frontiers in Queueing; Models and Applications in Science and Engineering*, 321-338, Ed. J.H. Dshalalow, CRC Press.

[12] Reiser, M. and H. Kobayashi. (1974) Accuracy of the diffusion approximation for some queueing systems. *IBM J. Res. Dev.,* 18, pp 110-124.

[13] Tzenova E., Adan I.J.B.F., and V. Kulkarni (2005), Fluid models with jumps, *Sochastic Models,* 21, No.1, pp 37-56

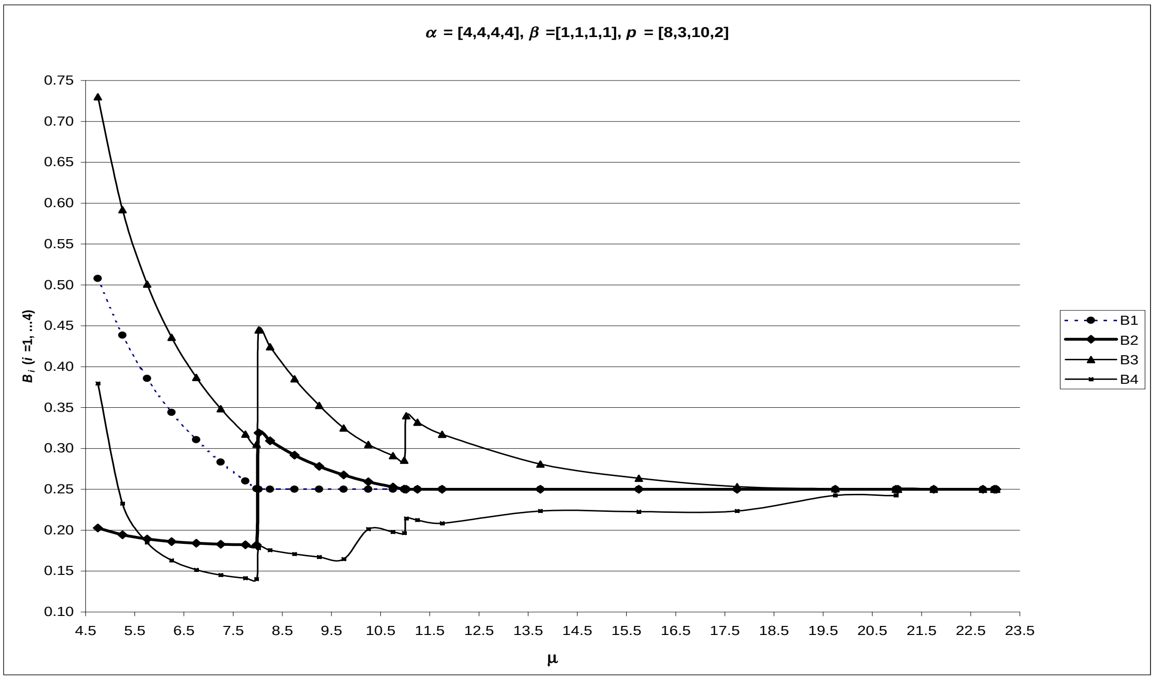

Figure 5.3: Output mean busy periods $B_i,\ i = 1, \ldots, 4$ as μ varies from 4.75 to 23.02

[14] Tzenova E., Adan I.J.B.F., and V. Kulkarni, A two-priority fluid flow model, *EURANDOM Report - 2005-16*

[15] Whitt, W. (1983) The queueing network analyzer. *The Bell Syst. Tech. J.,* 62, pp 2779-2815.