Analysis of fluid queues in saturation with additive decomposition

Miklós Telek and Miklós Vécsei

Budapest University of Technology and Economics Department of Telecommunications 1521 Budapest, Hungary {telek,vecsei}@webspn.hit.bme.hu

Abstract. Fluid queueing models with finite capacity buffers are applied to analyze a wide range of real life systems. There are well established numerical procedures for the analysis of these queueing models when the load is lower or higher than the system capacity, but these numerical methods become unstable as the load gets close to the system capacity. One of the available numerical procedures is the additive decomposition method proposed by Nail Akar and his colleagues.

The additive decomposition method is based on a separation of the eigenvalues of the characterizing matrix into the zero eigenvalue, the eigenvalues with positive real part and the eigenvalues with negative real part. The major problem of the method is that the number of zero eigenvalues increases by one at saturation. In this paper we present an extension of the additive decomposition method which remain numerically stable at saturation as well.

Key words: Markov fluid queue, additive decomposition method.

1 Introduction

Intuitively it is quite clear that infinite buffer queueing systems remain stable as long as the system load is below the system capacity. It is also widely accepted that finite buffer systems remain stable also when the system load is higher than the system capacity. This second statement suggests that finite buffer systems can be easily analyzed for any load level. In contrast, it turns out that standard solution methods suffer from severe numerical instabilities at the region where the load is close to the system capacity. It is interesting to note that analysis methods of finite buffer queueing systems used for the dimensioning of telecommunication network components are typically used for evaluating models close to saturation.

Apart of this practical issue, the common analysis approaches of finite buffer queueing systems exclude the case of saturation, because the discussion is restricted to the case when the load is below the system capacity and it is commonly left for the reader to invert the buffer content process if the load is higher than the system capacity. Unfortunately, this approach does not help when the load is equal to the system capacity. In this paper we consider Markov fluid queues (MFQs) with finite fluid buffers. There is a wide literature devoted to this subject (see e.g., [1-6]) for the case when the load is different from the capacity, but the case of saturation is considered only recently in [7] for the method proposed by Soares and Latouche in [3,8]. Here we investigate an other analysis method, the additive decomposition, which is proposed in [5], [6]. We propose an modification of the method which remains applicable in case of saturation.

The rest of the paper is organized as follows. Section 2 introduces Markov fluid queues (MFQs) with finite buffer and their analytical description. The next section discusses the additive decomposition method. The first subsection of Section 3 presents the solution method applicable for systems below and above saturation. The next subsection contains the proposed modification of the procedure for the case of saturation. Section 4 demonstrates the numerical properties of the standard and the proposed analysis methods. The paper is concluded in Section 5.

2 Markov fluid queue

The evolution of Markov fluid queue with single fluid buffer is determined by a discrete state of the environment and the continuous fluid level in the fluid buffer. The $Z(t) = \{M(t), X(t); t \ge 0\}$ process represents the state of the MFQ, where $M(t) \in S$ is the (discrete) state of the environment and $X(t) \in [0, b]$ is the fluid level in the fluid buffer at time t, where b denotes the buffer size. The fluid level cannot be negative or greater than b. We define $\hat{\pi}_j(t, x), \hat{p}_j(t, 0)$ and $\hat{p}_j(t, b)$ to describe the transient fluid densities at fluid level x and the transient probability masses of the fluid distribution at idle and full buffer as follows

$$\hat{\pi}_j(t,x) = \lim_{\Delta \to 0} \frac{\Pr(M(t) = j, x \le X(t) < x + \Delta)}{\Delta} ,$$
$$\hat{p}_j(t,x) = \Pr(M(t) = j, X(t) = x) \quad x = 0, b.$$

One of the main goal of the analysis of MFQ is to compute the stationary fluid density $\pi_j(x) = \lim_{t \to \infty} \hat{\pi}_j(t, x)$ and fluid mass at idle and full buffer $p_j(x) = \lim_{t \to \infty} \hat{p}_j(t, x), x = 0, b$. The row vector $\pi(x) = {\pi_j(x)}$, satisfies [9]

$$\frac{d}{dx}\pi(x)\mathbf{R} = \pi(x)\mathbf{Q} , \qquad (1)$$

where matrix $\mathbf{Q} = \{Q_{ij}\}\$ is the transition rate matrix of the environment process, and the diagonal matrix $\mathbf{R} = \text{diag}\langle R_j \rangle$ is composed by the fluid rates $R_j, j \in S$. R_j rate determines the rate at which the fluid level changes (increases when $R_j > 0$ or decreases when $R_j < 0$) when the environment is in state j. In this paper we assume that matrix \mathbf{Q} determines an irreducible Markov chain and exclude the case of $R_j = 0$. If there are states in the model where the fluid level remains constant then a censored process needs to be defined and investigated where sojourns in states with constant fluid level are excluded. Details of the censored analysis method can be find e.g. in [4]. A consequence of the exclusion of states with constant fluid level is that matrix \mathbf{R} is non-singular. We denote the set of states with positive fluid rates by S^+ and the set of states with negative fluid rates by S^- .

Kulkarni investigated the properties of the characterizing matrix of (1) in [1]. First of all, he defined the stability condition of infinite buffer MFQs. Let γ be the stationary distribution of the CTMC with generator matrix \mathbf{Q} . γ is the solution of the linear system $\gamma \mathbf{Q} = 0, \gamma \mathbb{1} = 1$, where $\mathbb{1}$ is the column vector of ones of appropriate size. An infinite buffer MFQ is stable if it "drift" is negative, where the drift is $d = \gamma \mathbf{R} \mathbb{1}$.

Further more differential equation in (1) suggests to find the solution of the fluid density function in a matrix exponential form. To find the matrix exponential solution [1] defines the relation of the number of states with positive and negative fluid rates and the number of eigenvalues of matrix \mathbf{QR}^{-1} with positive and negative real parts. These results are summarized in Table 1.

	d < 0	d = 0	d > 0	
positive eigenvalues	$ S^{-} - 1$	$ S^{-} - 1$	$ S^- $	
negative eigenvalue	$ S^+ $	$ S^+ - 1$	$ S^+ - 1$	
zero eigenvalue	1	2	1	

Table 1. Drift related properties of finite MFQs, where $|S^-|$ ($|S^+|$) is the number of states with negative (positive) fluid rate

The initial vector of the matrix exponential solution is determined by the boundary conditions.

$$p_i(0) = 0 \text{ for } i \in S^+, \ p_i(b) = 0 \text{ for } i \in S^-,$$
 (2)

and

$$-\pi_i(0)R_i + \sum_{j \in S^-} p_j(0)Q_{ji} = 0, \ \pi_i(b)R_i + \sum_{j \in S^+} p_j(b)Q_{ji} = 0.$$
(3)

(2) states that the fluid level cannot be 0 when the fluid rate is positive and it cannot be b when the fluid rate is negative. For $i \in S^-$ the first part of (3) means that the fluid level can be 0 due to a state transition of the environment from an other state with negative fluid rate or due to the fact that the fluid level reduced to 0 in a state with negative fluid rate. For $i \in S^+$ the first part of (3) represents that the fluid level can start increasing from 0 due to the fact that the process stayed in a state with negative fluid rate at level 0 and a state transition occurred to a state with positive fluid rate. The second part of (3) contains the counterpart statements for buffer level b.

3 The additive decomposition method

A numerically stable approach to the analysis of MFQs is the additive decomposition method [6]. It will be summarized in the following section. It's stability is based on the separation of the eigenvalues of the matrices in equation 1. The original additive decomposition algorithm from [6] can not be applied for fluid queues at saturation directly.

3.1 Fluid queues at non-zero drift

Due to the fact that states with constant fluid rates are excluded we can multiply both sides of (1) with \mathbf{R}^{-1} . If we denote \mathbf{QR}^{-1} with \mathbf{A} , this will result in the following differential equation:

$$\frac{d}{dx}\pi(x) = \pi(x)\mathbf{A} \tag{4}$$

The usual way of solving equations like (4) is based on it's spectral representation:

$$\pi(x) = e^{\lambda x} I$$

 λ is a scalar and Γ is a row vector. Substituting this form to (1), we find:

$$\lambda \Gamma = \lambda \mathbf{A} \tag{5}$$

After finding the eigenvalues λ_i and eigenvectors Γ_i one may search for $\pi(x)$ as the sum of the results of (4), with a_i parameters:

$$\pi(x) = \sum_{i} a_{i} e^{\lambda_{i} x} \Gamma_{i} \tag{6}$$

The limitation of this method may appear when we want to fit the formula to the boundary conditions at the buffer limit. The arising equations will define the a_i parameters in (6), hence they are crucial for solving the problem. If the buffer limit is large $(b \to \infty)$, than if $\lambda_i > 0 \to e^{\lambda_i b} \to \infty$ moreover if $\lambda_j < 0$, then $e^{\lambda_j b} \to 0$. This will result in badly conditioned linear equations for a_i .

The additive decomposition method solves this problem by separating the eigenvalues based on their sings, and by handling them separately. In [6] a procedure is described with which one may transform \mathbf{A} into a blockmatrix form. (It uses Schur-decomposition and solves a Lyapunov-equation in order to find it.)

$$\mathbf{Y}^{-1}\mathbf{A}\mathbf{Y} = \begin{pmatrix} 0 & 0 & 0\\ 0 & \mathbf{A}_{-} & 0\\ 0 & 0 & \mathbf{A}_{+} \end{pmatrix}$$
(7)

 \mathbf{A}_{-} (\mathbf{A}_{+}) is a square matrix, and all of it's eigenvalues are negative (positive). Let us denote different parts of \mathbf{Y}^{-1} with the following notations

$$\mathbf{Y} = \begin{pmatrix} \mathbf{L}_0 \\ \mathbf{L}_- \\ \mathbf{L}_+ \end{pmatrix} \tag{8}$$

 $\mathbf{L}_{\mathbf{0}}$ is the first row of \mathbf{Y}^{-1} while \mathbf{L}_{-} and \mathbf{L}_{+} have the same amount of rows as A_{-} and A_{+} respectively. In [6] it is proven, that the following form is also a complete solution of the differential equation (4):

$$\pi(x) = a_0 \mathbf{L}_0 + a_- e^{\mathbf{A}_- x} \mathbf{L}_- + a_+ e^{-\mathbf{A}_+ (b-x)} \mathbf{L}_+$$

 a_0 is a scalar and a_- and a_+ are row vectors with the same number of columns as \mathbf{A}_- and \mathbf{A}_+ respectively. They are the parameters we need to define from the boundary conditions. The linear equations in this case are numerically stable, because all of the eigenvalues of \mathbf{A}_- and $-\mathbf{A}_+$ are negative.

3.2 Fluid queues at saturation

The additive decomposition method was developed for fluid queues with nonzero mean drift, but the procedure as it is described in the previous subsection does not work in case of saturation. A slight enhancement is needed in order to apply the procedure for MFQs at saturation.

Theorem 1. In **A**'s normal Jordan form, there is one Jordan-block belonging to the zero eigenvalue, and it's size is 2×2 .

Proof. The numbers of eigenvalues of different signs are given in [1] and are summarized in Table 1. The multiplicity of the zero eigenvalue is 2 in saturation. Now we need to show that there is a single (linear independent) eigenvector associated with the zero eigenvalue, because in this case the Jordan decomposition contains a Jordan block of size 2.

The left eigenvector associated with the zero eigenvalue satisfies

$$\alpha \mathbf{Q} \mathbf{R}^{-1} = 0$$

Multiplying both sides with **R** shows that α should also be the left eigenvector of **Q** associated with the zero eigenvalue. Due to the fact that **Q** defines an irreducible Markov chain it has only a single (linear independent) eigenvector associated with the zero eigenvalue and it is γ .

Corollary 1. It is not possible to transform \mathbf{A} to the same form as in (7).

In case of a MFQ in saturation, instead of having a single matrix element associated with the zero eigenvalue, we have a Jordan-block of size 2×2 in the similar decomposition of **A** as the one in (7). Hence one needs to modify the original method for MFQs at saturation. The proposed modification is to transform A to the following form

$$\mathbf{Y}^{-1}\mathbf{A}\mathbf{Y} = \begin{pmatrix} \mathbf{A}_{\mathbf{0}} & 0 & 0\\ 0 & \mathbf{A}_{-} & 0\\ 0 & 0 & \mathbf{A}_{+} \end{pmatrix},\tag{9}$$

where $\mathbf{A}_{\mathbf{0}}$ corresponds to the 0 eigenvalues. Consequently $\mathbf{L}_{\mathbf{0}}$ will have two rows, a_0 will have two elements in (8), and the expression for $\pi(x)$ changes to

$$\pi(x) = a_0 e^{\mathbf{A}_0 x} \mathbf{L}_0 + a_- e^{\mathbf{A}_- x} \mathbf{L}_- + a_+ e^{-\mathbf{A}_+ (b-x)} \mathbf{L}_+$$

Unfortunately, this formula is not stable for large buffer limits. This happens, because one the off-diagonal elements of the Jordan-block A_0 are nonzero. For example, if it is an upper tridiagonal matrix then

$$\mathbf{A_0} = \begin{pmatrix} 0 & a_{12} \\ 0 & 0 \end{pmatrix} \to e^{\mathbf{A_0}x} = \begin{pmatrix} 1 & a_{12}x \\ 0 & 1 \end{pmatrix},$$

and $a_{12}x \to \infty$ as $x \to \infty$, therefore this matrix will be badly conditioned for large buffer limits. Thus one might experience numerical problems when fitting the parameters of the system to the boundary conditions.

4 Numerical examples

We analyzed the numerical properties of the algorithms for finite buffer MFQs using our MATLAB implementations, which are parts of the BuTools package (available at http://webspn.hit.bme.hu/~butools/). We compared the proposed procedure (Section 3.2), with the original additive decomposition method (Section 3.1) at two different drift values, one far from zero and one close to zero.

4.1 Comparison of methods when the drift is far from zero

First we evaluated the MFQ with buffer size b = 30, generator matrix and fluid rate matrix

	-4	0	2	1	1			1	0	0	0	0
	3	-6	0	2	1			0	1	0	0	0
$\mathbf{Q} =$	1	3	-5	1	0	,	$\mathbf{R} =$	0	0	-1	0	0
	3	1	1	-7	2			0	0	0	$^{-1}$	0
	1	1	0	1	-3			0	0	0	0	-1

respectively. The stationary distribution of the CTMC characterized by \mathbf{Q} is $\gamma = (0.314, 0.142, 0.154, 0.143, 0.247)$ and the drift is d = -0.00933. To quantify the difference between the results of the methods we used the following error measure:

$$\Delta = \sum_{i \in S} \int_0^o |\pi_i^O(x) - \pi_i^M(x)| dx + \sum_{i \in S} |p_i^O(0) - p_i^M(0)| + \sum_{i \in S} |p_i^O(b) - p_i^M(b)| ,$$

where $\pi_i^O(x)$ and $\pi_i^M(x)$ correspond to the fluid density for state *i* at level *x* for the original and the modified algorithms. $p_i^O(0)$ and $p_i^M(0)$ are the probabilities for the empty buffer and $p_i^O(b)$ and $p_i^M(b)$ are for the full buffer. The fluid density curves computed by the two methods are depicted in Figure 1.

We also calculated the difference between the methods for systems with state space cardinalities of 20 and 50. The results were similar. The average error was $\Delta \sim 10^{-5}$.



Fig. 1. The fluid density functions $(\pi_i(x))$ versus fluid level x) of the example with nonzero drift (b = 30, d = -0.00933). The left graph corresponds to the method proposed in [6], the right graph corresponds to the method proposed in Section 3.2.

4.2 Comparison of the methods when the drift is close to zero

In our second example the buffer size is b = 30 the generator matrix and the fluid rate matrix are

$$\mathbf{Q} = \begin{bmatrix} -5 & 3 & 1 & 0 & 1 \\ 5 & -8 & 0 & 2 & 1 \\ 1 & 0 & -4 & 2 & 1 \\ 4 & 1 & 0 & -6 & 1 \\ 1 & 0 & 0 & 2 & -3 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

The stationary distribution for this CTMC process is $\gamma = (0.349, 0.151, 0.087, 0.163, 0.250)$ and the drift is $d = -1.11 \cdot 10^{-16}$. The original method proposed in [6] and summarized in Section 3.1 failed in the phase of decomposition according to the signs of the eigenvalues using the standard numerical precision of MATLAB, while the modified method completes. The obtained fluid density curve is depicted in Figure 2. When the drift is close to zero the original procedure gets numerically instable as it is clearly visible on the figure.

4.3 Analysis of a communication system with RED

We analyze a communication system using the proposed method. The fluid level represents the amount of data in the buffer, and the data arrival and service processes are modulated by an environmental Markov chain with generator Q. There are N identical users in the system. They are either in the ON or in the OFF state. In the ON state they transmit data at rate r, otherwise they do not. The sojourn time in state ON (OFF) is exponentially distributed with parameter α (β). The service speed of the server is c, and reject incoming data with probability 1-s, consequently data arrive to the server at rate r*s. This last



Fig. 2. The fluid density function $(\pi_i(x))$ versus fluid level x) for a queue with zero drift $(d = -1.11 \cdot 10^{-16}, b = 30)$

functional property is referred to as "random early detection" (RED) mechanism [10]. The RED method filters the input data as a function of the fluid level, namely s(x) is a function of the fluid level x. Assuming that s(x) is a piecewise constant function the multi region version of the adaptive decomposition method [6] and its modification for the case of zero drift in Section 3.2 allows to analyze the described communication systems. The limits of the constant regions of s(x) are denoted by x_j (j = 0, 1, ..., k), such that $x_0 = 0$ and $x_k = B$.

Due to the identity of the N users a MFQ with N+1 states describe the system behavior with generator matrix

$$\mathbf{Q} = \begin{pmatrix} -N\beta & N\beta & 0 & 0 & 0 & 0 \\ \alpha & -\alpha - (N-1)\beta & (N-1)\beta & 0 & 0 & 0 \\ 0 & 2\alpha & -2\alpha - (N-2)\beta & (N-2)\beta & 0 & 0 \\ 0 & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & (N-1)\alpha - (N-1)\alpha - \beta & \beta \\ 0 & 0 & 0 & 0 & N\alpha & -N\alpha \end{pmatrix},$$

and fluid rate matrix

$$\mathbf{R}(x) = \begin{pmatrix} -c & 0 & 0 & 0 & 0 & 0 \\ 0 & rs(x) - c & 0 & 0 & 0 & 0 \\ 0 & 0 & 2rs(x) - c & 0 & 0 & 0 \\ 0 & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & (N-1)rs(x) - c & 0 \\ 0 & 0 & 0 & 0 & 0 & Nrs(x) - c \end{pmatrix}.$$

One of the most important performance measure of this system is the loss. The loss is the amount of lost data. Loss may be caused by two phenomenons. The first is the filtering of the RED mechanism. When n users are ON the loss rate is $L_1 = (1 - s)nr$. The second reason for the loss is the finite buffer. The server may also loose data when the buffer is full. As the buffer is served with speed c, the loss rate due to the finite buffer capacity is $L_2 = snr - c$. These two parts of the loss can be computed as

$$L_{1} = \int_{0}^{B} r_{i}(1 - s(x))f_{i}(x)dx + \sum_{j,k} p(x_{j},k)r_{k}(1 - s(x_{j})),$$
$$L_{2} = \sum_{k} p(B,k)(s(B)r_{k} - c)$$

where $f_i x$ is the stationary probability density for state *i*, and $p(x_j, k)$ is the probability at threshold level x_j for state *k*. Based on these loss rates the loss ratio is

$$L = \frac{L_1 + L_2}{\int\limits_0^B r_i f_i(x) dx + \sum\limits_{j,k} p(x_j,k) r_k}$$

We analyze the performance measures of interest through the MFQ model and the additive decomposition method. The model parameters are $\alpha = 2/3\frac{1}{s}$, $\beta = 1\frac{1}{s}$, r = 12.2kbps, N = 25, c = 190kbps and B = 30kb. Without RED filtering (s(x) = 1) the drift is d = 183kbps, and with decreasing RED acceptance probability the drift is decreasing to d = -c at s(x) = 0. We considered 2 kinds of piecewise constant functions for s(x). The (0, B) interval was divided into 3 and 6 identical subintervals. E.g., in the first case $x_1 = 10$, $x_2 = 20$, $x_3 = 30$ and vector (s_1, s_2, s_3) contains the acceptance probabilities for the intervals (0, 10), (10, 20), (20, 30), respectively. Figure 3 depicts the fluid density functions for different s(x) functions. In the first graph s(x) = 1, in the second graph $(s_1, s_2, s_3) =$ (0.905, 0.8041, 0.72), in the third graph $(s_1, s_2, s_3) = (1, 0.8127, 0.76)$, in the fourth graph $(s_1, s_2, s_3, s_4, s_5, s_6) = (0.908, 0.7936, 0.72, 0.69, 0.65, 0.54)$. The associated loss rations are 0.0121, 0.0995, 0.0492 and 0.100.

5 Conclusions

The problem of analyzing finite buffer MFQs in saturation has been considered recently in [7]. In that paper the numerical procedure by Soares and Latouche [3, 8] was generalized for the case of saturation. In this paper we considered the additive decomposition procedure by Nail et al. [5, 6] and generalized for the case of saturation.

The proposed modification seems to eliminate the numerical instabilities of the method for drift values close to zero and for moderate buffer sizes. The case of extremely large buffers still results in numerical problems, because in saturation a Jordan block of size 2×2 associated with the zero eigenvalue, which results in an exponentially increasing coefficient.

Acknowledgements

This work was supported by OTKA grant no. K-101150.

References

 V. G. Kulkarni, "Fluid models for single buffer systems," in *Models and Applica*tions in Science and Engineering, ser. Frontiers in Queueing, J. H. Dshalalow, Ed. CRC Press, 1997, pp. 321–338.



Fig. 3. Fluid density functions with s(x) = 1, $(s_1, s_2, s_3) = (0.905, 0.8041, 0.72)$, $(s_1, s_2, s_3) = (1, 0.8127, 0.76)$, $(s_1, s_2, s_3, s_4, s_5, s_6) = (0.908, 0.7936, 0.72, 0.69, 0.65, 0.54)$, respectively.

- S. Ahn, J. Jeon, and V. Ramaswami, "Steady state analysis of finite fluid flow models using finite qbds," *Queueing Systems*, vol. 49, pp. 223–259, 2005.
- A. da Silva Soares and G. Latouche, "Matrix-analytic methods for fluid queues with finite buffers," *Perform. Eval.*, vol. 63, pp. 295–314, May 2006.
- 4. M. Gribaudo and M. Telek, "Stationary analysis of fluid level dependent bounded fluid models," *Performance Evaluation*, vol. 65, pp. 241–261, 2008.
- K. Sohraby and N. Akar, "Infinite/finite buffer markov fluid queues: A unified analysis," *Journal of Applied Probability*, vol. 41, no. 2, pp. 557–569, 2004.
- H. Kankaya and N. Akar, "Solving multi-regime feedback fluid queues," *Stochastic Models*, vol. 24, no. 3, pp. 425–450, 2008.
- M. Telek and M. Vécsei, "Finite queues at the limit of saturation," in 9th International Conference on Quantitative Evaluation of SysTems (QEST). London, UK: Conference Publishing Services (CPS), sept. 2012, pp. 33–42.
- 8. A. da Silva Soares and G. Latouche, "Fluid queues with level dependent evolution," *European Journal of Operational Research*, vol. 196, no. 3, pp. 1041 – 1048, 2009.
- G. Horton, V. Kulkarni, D. Nicol, and K. Trivedi, "Fluid stochastic petri nets: Theory, applications, and solution techniques," *European Journal of Operational Research*, vol. 105, pp. 184–201, 1998.
- S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397–413, 1993.