Applied Mathematical Modelling 37 (2013) 3736–3747

Contents lists available at SciVerse ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

# A two-stage queueing network with MAP inputs and buffer sharing

Wenhui Zhou [a], Zhaotong Lian [b,*], Wei Xu [b], Weixiang Huang [a]

[a] School of Business Administration, South China University of Technology, Guangzhou, China
[b] Faculty of Business Administration, University of Macau, Macau SAR, China

ARTICLE INFO

*Article history:*
Received 13 June 2011
Received in revised form 7 June 2012
Accepted 3 July 2012
Available online 7 September 2012

*Keywords:*
Two-stage queueing network
Buffer sharing
MAP
Matrix filtrate technique

ABSTRACT

In this paper, we study a two-stage tandem queueing network with MAP inputs and buffer sharing. The two stages share the same buffer. By using Markov process, we give an exact analysis of the queueing network. Since the customer arrival is not a Poisson process, the PASTA (Poisson Arrivals See Time Averages) property does not hold. A matrix filtration technique is proposed to derive the probability distribution of queue length at arrivals. Our objective is to investigate how the buffer sharing policy is mitigate the tradeoff between the probability that an arriving customer is lost and the probability that the first-stage server is blocked. The numerical results show that buffer sharing policy is more flexible, especially when the inputs have large variant and are correlated.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

Considering a two-stage tandem queueing network, the settings of the two buffers can be finite or infinite. If the first-stage buffer is finite, the arriving customers could be lost due to the first-stage buffer becoming unavailable. If the second-stage buffer is finite, the first-stage server may be blocked because the second-stage buffer is full. Our goal is to investigate how the buffer sharing policy mitigates the tradeoff between the probability that an arriving customer is lost and the probability that the first-stage server is blocked, especially, when the customer arrivals are correlated. Motivated by this purpose, we consider a two-stage tandem queueing network with finite buffers (see Fig. 1). The customer arrivals follow a Markovian arrival process (MAP). Both buffers are finite and the whole or part of the buffers can be shared with each other.

There is extensive research about two-stage tandem queueing networks. When both buffers have infinite capacity in the two-stage queueing networks, if both nodes are quasi-reversibility, the network have product-form solution (Theorem 4.3 [1]) and is analytically tractable. When the quasi-reversibility is not follow, most of research define the number of customer in the first queue as a level and the number of customer in the second queue as a phase. The system is then involved in a two-dimensional infinite Quasi-Birth-and-Death (QBD) process, with which, it is difficult to derive the typical system performance measures, for example, the customer sojourn time and the busy period. Lian and Liu [2] define the total number of customers in the system as a level, and the number of customer in the first queue as a phase, so that the system can be handled by constructing a one-dimensional level-dependent QBD process.

When the first-stage buffer is infinite and the second-stage buffer is finite, the main focus of such tandem queueing network is “blocking”. Hunt [3] was perhaps the first work to study tandem queues with blocking. Without any claim to an exhaustive enumeration, two-stage tandem queueing network with a finite intermediate buffer or without intermediate waiting spaces were discussed by Avi-Itzhak and Halfin [4], Avi-Itzhak and Yadin [5], Grassmann and Drekic [6], Konheim

* Corresponding author.
*E-mail address:* lianzt@umac.mo (Z. Lian).

0307-904X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.apm.2012.07.003

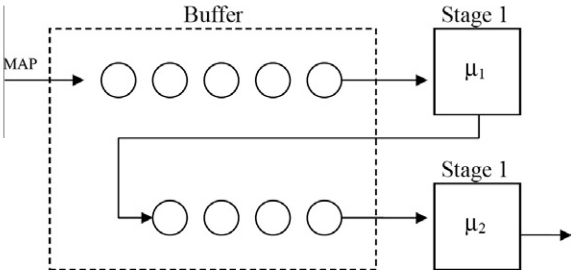


**Fig. 1.** A service system with a sharing buffer.

and Reiser [7,8], Latouche and Neuts [9], Neuts [10,11], Prabhu [12], Gómez-Corral and Martos. [13]. Hall and Sriskandarajah [14] and Perros [15] provide some survey on the tandem networks.

If two buffers are finite, not only the first-stage server may be blocked when the second-stage buffer is full, but also an arrival customer has been lost when the first-stage buffer is full. Most of research in this field studied buffer allocation. (For example, see [16–18].)

Zhou and Lian [19] first studied a two-stage tandem queueing network with buffer sharing policy. The total physical capacity of the system buffer is finite (denote by $N$ the buffer size). The buffer threshold for the first stage is $N_1$, and buffer threshold for the second stage is $N_2$ ($N_1 \leqslant N, N_2 \leqslant N$). As the buffer is finite, the arrival customer will be rejected if the buffer quota for the first stage is full, and the first server will be blocked if the buffer quota for the second stage is full.

The work presented in this paper is part of an ongoing study on [19]. In reality, customer arrival may not follow a Poisson process. In some cases, the customer inter-arrival times are even not be independently identical. Therefore in this paper, we consider an Markovian arrival process (MAP) input which the customer inter-arrival times are dependent. This paper is organized as follows. In Section 2, we present the model description. The stationary probability distribution and the stationary probability distribution at arrivals are derived in Sections 3 and 4, respectively. We obtain the tail probability distribution of the customer sojourn time in Section 5. Some numerical results are analyzed in Section 6. Section 7 concludes the paper.

## 2. Model description

We consider a two-stage tandem queueing network. Each stage has a single exponential server. The service rates for the two servers are $\mu_1$ and $\mu_2$, respectively. We denote $\mu = \mu_1 + \mu_2$. Customers arrive to queueing network following an MAP with an infinitesimal generator $D = D^0 + D^1$ in the state space $\{1, \ldots, m\}$, where $D^0 = (D^0_{i,j})_{m\times m}$ and $D^1 = (D^1_{i,j})_{m\times m}$. All the off-diagonal elements of $D^0$ and all the elements of $D^1$ are nonnegative, and all the diagonal entries of $D^0$ are non-positive. The transitions associated with $D^1$ are called type-1 transitions. A single customer arrives at and only at each type-1 transition instant.

Assume that the underlying Markov chain $D$ is irreducible and let $\mathbf{z}$ denote its stationary probability vector. In other words, $\mathbf{z}$ is uniquely determined by $\mathbf{z} \cdot (D^0 + D^1) = 0$ and $\mathbf{z} \cdot \mathbf{1} = 1$. The mean arrival rate of the MAP is $\lambda = \mathbf{z}D^1\mathbf{1}$, where $\mathbf{1}$ is a vector with all elements being equal to 1. Also, The variance $v$ of intervals between customer arrivals is

$$v = 2\lambda^{-1}\mathbf{z}(-D_0)^{-1}\mathbf{1} - \lambda^{-2}. \tag{1}$$

And the correlation coefficient $\rho$ of intervals between successive group arrivals is given by

$$\rho = (\lambda^{-1}\mathbf{z}(-D_0)^{-1}D_1(-D_0)^{-1}\mathbf{1} - \lambda^{-2})/v. \tag{2}$$

For further properties of the MAP, the readers are referred to [20,21]. To avoid trivial cases, we assume $D^1 \neq \mathbf{0}$ so that $\lambda > 0$, where $\mathbf{0}$ denotes a matrix of zeros.

The customers are served in each stage based on FCFS (First come first serve) discipline. The total physical capacity of the system buffer is finite (denoted by $N$). The two stages share the same buffer, in which the buffer quota for the first stage is $N_1$ ($0 \leqslant N_1 \leqslant N$) and the buffer quota for the second stage is $N_2$ ($0 \leqslant N_2 \leqslant N$). If there are $N_1$ customers waiting in the first-stage, or the number of customers waiting in the whole buffer equal to $N$, the new arriving customer will be rejected. Correspondingly, if there are $N_2$ customers waiting in the second-stage, or the number of customers waiting in the whole buffer equal to $N$, the first server must hold the completed customer and the server will be blocked.

(1) When $N_1 + N_2 = N$, there is no buffer sharing. This is a buffer allocation problem.
(2) When $N_1 + N_2 > N$ and $N_1 < N, N_2 < N$, we call the sharing policy "Partial Buffer Sharing" policy.
(3) When $N_1 = N, N_2 = N$, we call the sharing policy "Complete Buffer Sharing".

Throughout the paper, we adopt Block After Serve (BAS) policy, in other words, if a customer completes service at the first server and there are no empty space at the second stage, the customer will be blocked, i.e., the customer remains in the first

Download English Version:

https://daneshyari.com/en/article/1704647

Download Persian Version:

https://daneshyari.com/article/1704647

Daneshyari.com