Applied Mathematical Modelling 38 (2014) 6122–6132

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Approximation of throughput in tandem queues with multiple servers and blocking

Yang Woo Shin [a,*], Dug Hee Moon [b]

[a] Department of Statistics, Changwon National University, Changwon, Gyeongnam 641-773, Republic of Korea
[b] Department of Industrial and Systems Engineering, Changwon National University, Changwon, Gyeongnam 641-773, Republic of Korea

ARTICLE INFO

Article history:
Received 6 February 2013
Received in revised form 11 March 2014
Accepted 16 May 2014
Available online 2 June 2014

Keywords:
Tandem queues
Finite buffers
Exponential service times
Blocking
Throughput
Decomposition

ABSTRACT

In this paper, we develop an approximation method for throughput in tandem queues with multiple independent reliable servers at each stage and finite buffers between service stations. We consider the blocking after service (BAS) blocking protocol of each service stage. The service time distribution of each server is exponential. The approximation is based on the decomposition of the system into a set of coupled subsystems which are modeled by two-stage tandem queue with two buffers and are analyzed by using the level dependent quasi-birth-and-death (LDQBD) process.

1. Introduction

Tandem queues with finite buffers have been widely used for performance modeling of computer systems, telecommunication networks and manufacturing systems, and it is studied extensively in the literature e.g., Gershwin [1], Perros [2], Perros and Altiok [3] and the references therein. Models with finite buffers and exponential service times can be represented by finite state Markov chains [4]. However, the number of states of the Markov chain increases drastically as the number of stages increases, which makes numerical solutions intractable. Only small systems with one or two queues can be solved exactly; for exact methods we refer to [5–7,4] for the system with reliable servers and to [8,9,2] for the system with unreliable servers. The methods introduced for exact solution of two stage open queuing networks can include the ones which employs matrix geometric method [10,11] and spectral expansion [12], and approaches to ease state explosion problem such as dominant eigenvalues approach [13], three-dimensional approach [14] and hybrid solution approach [15] have been presented. The two stage tandem queue with single server and interarrival time of general distribution is studied by Klimenok et al. [7]. Kim et al. [5] deal with the single server two stage tandem queue with retrials and losses instead of blocking.

Many approximation methods are presented for the analysis of tandem queue with finite buffers. One of the most common method among the approximation techniques is decomposition method which decomposes the long line into subsystem with two service stations and one buffer, and derives a set of equations that determine the unknown parameters of each subsystem, and finally develops an iterative algorithm to solve these equations. For more about decomposition method for single server tandem queues, see e.g. Gershwin [16] for the system with unreliable server and exponential

* Corresponding author. Tel.: +82 552133462; fax: +82 552133460.
E-mail addresses: ywshin@changwon.ac.kr (Y.W. Shin), dhmoon@changwon.ac.kr (D.H. Moon).

http://dx.doi.org/10.1016/j.apm.2014.05.015
0307-904X/

distribution of service time, Helber [17] for the system with unreliable server and Cox-2 distribution of service time, Bierbooms et al. [18], van Vuuren and Adan [19] for the system with reliable server and general distribution of service time. The literature on the analysis of tandem queues with multiple servers and finite buffer is more scarce than the single server system. Maximum entropy method [20] and expansion method [21] are presented for approximation of the subsystem in multi-server queueing networks with finite buffers and exponential distribution of service time. van Vuuren et al. [22] decomposed the tandem queue with parallel reliable servers and general distribution of service time into subsystems with two service stations and one buffer space, and then approximate each multi-server subsystem by an aggregated single (super) server queue with state-dependent inter-arrival and service times. They used mixture of Erlang distribution and Cox-2 distribution for an approximation of individual service time and aggregated (super) server. An iterative algorithm to approximate the arrival and departures at the subsystems. The steady-state queue length distribution of each subsystem is determined by a spectral expansion method. In Diamantidis et al. [23], decomposition technique was applied directly to the tandem queues with multiple reliable servers whose service time distribution is exponential at each stage for approximation of throughput and mean number of customers in the system.

In this paper, we focus on multi-server tandem queues with exponential service times, reliable servers, finite buffers and blocking after service (BAS) protocol. We present a decomposition method using three-stage subsystem to reflect the dependence between consecutive stages and improve the quality of approximation. The system is decomposed into the subsystems with three-stage and two buffers and each subsystem is approximated by a tandem queue with two-stage and two buffers in which the arrival rate and departure rate are state dependent. The approximate subsystem is modeled with level dependent quasi-birth-and-death (LDQBD) process. Arrival and service rates of the subsystem are obtained iteratively.

The paper is organized as follows. In Section 2 we describe the model in detail and its decomposition. Some preliminaries and approximation of subsystem are presented in Sections 3 and 4, respectively. Section 5 provides an algorithm for calculating throughput. The results of the approximation method are compared numerically with simulation and existing methods in Section 6. Finally, Section 7 contains some concluding remarks.

2. The Model and decomposition

We consider a tandem queue L with $N+1$ stages labeled $S_i,\ i=0,1,\ldots,N$. Each stage S_i has a service station M_i with m_i parallel identical servers and there is a finite buffer B_i of size b_i in front of $M_i,\ i=1,2,\ldots,N$ as depicted in Fig. 1. When a customer completes its service at a stage and finds that the next stage is full at that time, the customer is held at the server where it just completed service until the destination can accommodate it. This type of blocking is called BAS (blocking after service) blocking. The servers in M_0 are never starved and the servers in the last stage M_N are never blocked. The service time distribution of each server in M_i is assumed to be exponential with rate μ_i.

The tandem queue L is decomposed into $N-1$ subsystems $L_1,\ L_2,\ldots,\ L_{N-1}$ as shown in Fig. 1. Subsystem L_i consists of two finite buffers B_i and B_{i+1} and three service stations $M_{i-1},\ M_i$ and M_{i+1}, where the service station M_{i-1} is a virtual station which supplies the customers into S_i. The departure process from L_i may be affected by blocking due to full buffer at S_{i+2}.

3. Preliminaries

Let $X_i(t)$ be the number of customers at S_i which includes the customers blocked at M_{i-1} but does not include the customers at M_i that are blocked to enter the next stage S_{i+1} at time t. It can be easily seen that the state space of $X_i(t)$ is $\{0,1,\cdots,K_i\}$, where $K_i=m_{i-1}+l_i$ and $l_i=m_i+b_i$.

Let $\boldsymbol{Z}_i=\{Z_i(t),\ t\geqslant 0\}$, where $Z_i(t)=(X_i(t),X_{i+1}(t))$. Let X_i and X_{i+1} be the stationary version of $X_i(t)$ and $X_{i+1}(t)$, respectively and $Z_i=(X_i,X_{i+1})$. Since $X_i+X_{i+1}\leqslant K_{i+1}+m_{i-1}+b_i$, the state space of $\boldsymbol{Z}_i$ is

$$\mathcal{S}_i=\{(n,j),\ 0\leqslant n\leqslant a_i(j),\ 0\leqslant j\leqslant K_{i+1}\}=\{(n,j),\ 0\leqslant j\leqslant d_i(n),\ 0\leqslant n\leqslant K_i\},$$

where $a_i(j)=K_i-\max(j-l_{i+1},0)$ and $d_i(n)=K_{i+1}-\max(n-m_{i-1}-b_i,0)$.

Let $\pi_i(n,j)=P(X_i=n,X_{i+1}=j),\ (n,j)\in\mathcal{S}_i$ and define the marginal distributions

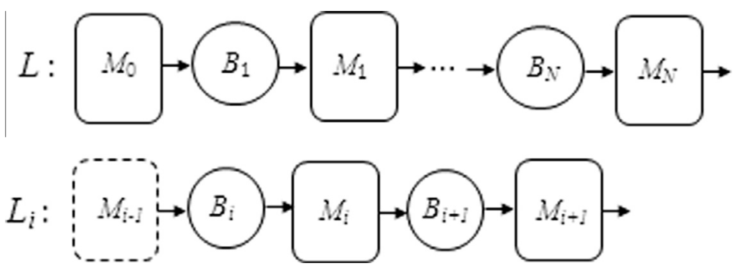

Fig. 1. Decomposition of tandem queue.