European Journal of Operational Research 235 (2014) 170–179

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Tandem queueing system with infinite and finite intermediate buffers and generalized phase-type service time distribution

Chesoong Kim [a,*], Alexander Dudin [b], Olga Dudina [b], Sergey Dudin [b]

[a] Sangji University, Wonju, Kangwon 220-702, Republic of Korea
[b] Belarusian State University, 4, Nezavisimosti Ave., Minsk 220030, Belarus

ARTICLE INFO

Article history:
Received 17 July 2013
Accepted 6 December 2013
Available online 15 December 2013

Keywords:
Queueing
Tandem queueing system
Marked Markovian arrival process
Phase-type distribution
Laplace–Stieltjes transform

ABSTRACT

A tandem queueing system with infinite and finite intermediate buffers, heterogeneous customers and generalized phase-type service time distribution at the second stage is investigated. The first stage of the tandem has a finite number of servers without buffer. The second stage consists of an infinite and a finite buffers and a finite number of servers. The arrival flow of customers is described by a Marked Markovian arrival process. Type 1 customers arrive to the first stage while type 2 customers arrive to the second stage directly. The service time at the first stage has an exponential distribution. The service times of type 1 and type 2 customers at the second stage have a phase-type distribution with different parameters. During a waiting period in the intermediate buffer, type 1 customers can be impatient and leave the system. The ergodicity condition and the steady-state distribution of the system states are analyzed. Some key performance measures are calculated. The Laplace–Stieltjes transform of the sojourn time distribution of type 2 customers is derived. Numerical examples are presented.

1. Introduction

Tandem queueing systems are very important part of the queueing theory that takes into account the possibility that a customer may need service from several sequentially arranged servers. The overwhelming majority of the results in the theory of tandem queues is obtained for systems with stationary Poisson arrival process and exponential service time distribution. Tandem queueing systems with correlated arrival flows were investigated not so extensively. For background information and overview of the present state of the art in the study of tandem queueing systems with correlated arrival flows and phase-type service time distribution see, e.g., Gomez-Corral (2002a, 2002b), Klimenok, Breuer, Tsarenkov, and Dudin (2005), Gomez-Corral and Martos (2006), Klimenok, Kim, Tsarenkov, Breuer, and Dudin (2007), Kim, Park, Dudin, Klimenok, and Tsarenkov (2010), and Kim, Dudin, Dudin, and Dudina (2013a). The papers Gomez-Corral (2002a, 2002b) and Gomez-Corral and Martos (2006) are devoted to the $MAP/PH/1 \rightarrow \bullet/G/1$ system with blocking. The tandem queue $BMAP/G/1/N \rightarrow \bullet/PH/1/M$ with losses is considered in Klimenok et al. (2005). The tandem-queue of the $BMAP/G/1 \rightarrow \bullet/PH/1/M$ type with losses and feedback is investigated in Klimenok et al. (2007). Analysis of the $BMAP/G/1 \rightarrow \bullet/PH/1/M$ tandem queue with retrials and losses is provided in Kim et al. (2010). In Kim et al. (2013a), a tandem queueing system with more than one server in each stage, Markovian arrival flow, and phase-type service time distribution at the second stage is considered.

In this paper, we analyze a tandem queueing system with two types of customers. Type 1 customers arrive to the first stage that represents a multi-server queue without buffer. After service completion at the first stage, type 1 customers leave the system or move to the second stage. Type 2 customers arrive to the second stage directly. The second stage is a multi-server queue with a finite buffer for type 1 customers and an infinite buffer for type 2 customers. Type 1 customers have non-preemptive priority over type 2 customers. They can be impatient and leave the intermediate buffer without service.

In order to study the effect of variation and correlation within the arrival process as well as the correlation of arrivals of different types of customers, we consider a Marked Markovian arrival process (*MMAP*) to model customer arrivals. Additionally, we assume that the service time of different types customers has a phase-type (*PH*) distribution with different parameters. This allows us to take into account the variance of the service time. Note that using a phase-type distribution significantly increases the dimension of the system state space and complicates the investigation of the system. To overcome these difficulties to some extent, we use the method proposed by Ramaswami (1985) and Ramaswami and Lucantoni (1985), for reducing the dimension of the state spaces of systems with a phase-type service time distribution,

* Corresponding author. Tel.: +82 33 730 0464.
E-mail addresses: dowoo@sangji.ac.kr (C. Kim), dudin@bsu.by (A. Dudin), dudina_olga@email.com (O. Dudina), dudin85@mail.ru (S. Dudin).

0377-2217/$ - see front matter
http://dx.doi.org/10.1016/j.ejor.2013.12.012

and consider a generalized phase-type distribution of the service times at the second stage instead of considering two separate phase-type distributions for each type of customers, that greatly facilitates the investigation of the system.

The queueing system under consideration is quite general and can be applied for modeling many real-world systems. Let us introduce several examples of such systems:

(i) Contact-center with an *IVR* (Interactive Voice Response). Type 1 customers can be interpreted as calls and type 2 customers are the text requests (e'mails). The first stage server represents an *IVR* and the second stage server is a contact center operator. E'mail requests arrive to the second stage directly and are always patient. Callers are firstly serviced by an *IVR*. If the caller cannot solve his (her) problem by using an *IVR*, he (she) can request to connect with an operator and move to the second stage. Type 1 customers can be impatient. The service times of e'mail requests and calls are different.
(ii) Manufacturing system. The system processes two types of details. Type 1 details need some preprocessing while type 2 details do not require such a preprocessing. After the preprocessing, type 1 details should be processed during a limited time (e.g., if the preprocessing includes heating of details, type 1 details should be processed before they become cool).
(iii) Database. The system processes queries from external users, who need preliminary identification and are non-patient, along with the requests from the internal users to retrieve or update information.
(iv) Information transmission system. More important information should be encrypted prior transmission and transmitted more urgent than some less important information which is sent without encryption.
(v) Medical system. Emergency surgery center has several rooms for anesthesiology and operation theatre. Depending on the type of injury, some arriving patients need more urgent treatment after a preliminary general anesthetic, while others may wait longer time and do not need general anesthetic at all.

The paper is organized as follows. In Section 2, the mathematical model is described. In Section 3, the process of the system states is considered. The ergodicity condition and the stationary state distribution are analyzed in Section 4. The expressions for the main performance measures of the system are given in Section 4. The Laplace–Stieltjes transform of the sojourn time distribution of an arbitrary type 2 customer is presented in Section 5. Section 6 contains numerical examples. Section 7 concludes the paper.

2. Mathematical model

We consider a tandem queueing system with two types of customers and two intermediate buffers. The structure of the system under study is presented in Fig. 1.

Customers arrive to the system according to the *MMAP*. The customers in the *MMAP* are heterogeneous and have different types. The arrival of customers is directed by the stochastic process $\nu_t,\ t \geqslant 0$, which is an irreducible continuous-time Markov chain with the state space $\{0, 1, \ldots, W\}$. The sojourn time of this chain in the state ν is an exponentially distributed with the positive finite parameter $\lambda^{(\nu)}$. When the sojourn time in the state ν expires, with probability $p^{(0)}_{\nu,\nu'}$ the process ν_t jumps to the state ν' without generation of a customer, $\nu, \nu' = \overline{0, W},\ \nu \neq \nu'$, and with probability $p^{(l)}_{\nu,\nu'}$ the process ν_t jumps to the state ν' with a generation of type l customer, $l = 1,\ 2,\ \nu, \nu' = \overline{0, W}$. The notation $\nu = \overline{0, W}$ means that the parameter ν takes values in the set $\{0, 1, \ldots, W\}$.

The behavior of the *MMAP* is completely characterized by the matrices D_0, $D_1^{(l)},\ l = 1, 2$, defined by the entries $(D_1^{(l)})_{\nu,\nu'} = \lambda^{(\nu)} p^{(l)}_{\nu,\nu'}$, $\nu, \nu' = \overline{0, W}$, $l = 1, 2$, and $(D_0)_{\nu,\nu} = -\lambda^{(\nu)}$, $\nu = \overline{0, W}$, $(D_0)_{\nu,\nu'} = \lambda^{(\nu)} p^{(0)}_{\nu,\nu'}$, $\nu, \nu' = \overline{0, W},\ \nu \neq \nu'$. The matrix $D(1) = D_0 + D_1^{(1)} + D_1^{(2)}$ represents the generator of the process $\nu_t, t \geqslant 0$.

The average total arrival intensity λ is defined by $\lambda = \boldsymbol{\theta}\left(D_1^{(1)} + D_1^{(2)}\right)\mathbf{e}$ where $\boldsymbol{\theta}$ is the invariant vector of the stationary distribution of the Markov chain $\nu_t, t \geqslant 0$. The vector $\boldsymbol{\theta}$ is the unique solution to the system $\boldsymbol{\theta} D(1) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. Hereinafter $\mathbf{e}$ denotes a column vector consisting of 1's, and $\mathbf{0}$ is a zero row vector. The average arrival intensity λ_l of type l customers is defined by $\lambda_l = \boldsymbol{\theta}\, D_1^{(l)}\mathbf{e}$, $l = 1, 2$.

The squared integral (without differentiating the types of customers) coefficient of variation c_{var} of the intervals between customer arrivals is defined by $c_{var} = 2\lambda\boldsymbol{\theta}(-D_0)^{-1}\mathbf{e} - 1$. The squared coefficient of variation $c_{var}^{(l)}$ of inter-arrival times of type l customers is defined by

$$c_{var}^{(l)} = 2\lambda_l\boldsymbol{\theta}\left(-D_0 - D_1^{(l')}\right)^{-1}\mathbf{e} - 1, \quad l' \neq l,\ l',\ l = 1, 2.$$

The integral coefficient of correlation c_{cor} of two successive intervals between arrivals is defined by

$$c_{cor} = (\lambda\boldsymbol{\theta}(-D_0)^{-1}(D(1) - D_0)(-D_0)^{-1}\mathbf{e} - 1)/c_{var}.$$

The coefficient of correlation $c_{cor}^{(l)}$ of two successive intervals between type l customers' arrivals is computed by

$$c_{cor}^{(l)} = \left(\lambda_l\boldsymbol{\theta}\left(D_0 + D_1^{(l')}\right)^{-1} D_1^{(l)}\left(D_0 + D_1^{(l')}\right)^{-1}\mathbf{e} - 1\right)/c_{var}^{(l)},$$
$$l' \neq l,\ l',\ l = 1, 2.$$

We assume that type 1 customers arrive to the first stage, while type 2 customers arrive to the second stage directly, not entering the first stage. The first stage is described by an R-server queueing system without buffer. The service time for each server at this stage has an exponential distribution with the parameter μ.

After receiving service at the first stage, type 1 customer proceeds to the second stage of the tandem with probability q, $0 \leqslant q \leqslant 1$, or leaves the system forever (is lost) with the complementary probability. The second stage represents an N-server queue with a finite buffer of capacity K for type 1 customers (buffer 1) and an infinite buffer for type 2 customers (buffer 2).

If there is a free server at the second stage during an arbitrary customer arrival epoch, the customer occupies this server immediately. If all servers are busy and buffer 1 is not full during an arbitrary type 1 customer arrival epoch, the customer is admitted to this buffer. Otherwise, type 1 customer leaves the system forever (is lost). If all servers at the second stage are busy during a type 2 customer arrival epoch, this customer moves to buffer 2.

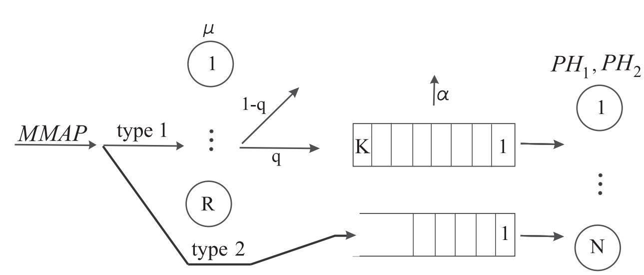

Fig. 1. Structure of the system.