Operations Research Letters 43 (2015) 203–208

Contents lists available at ScienceDirect

Operations Research Letters

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

Explicit expressions for moments of waiting times in Poisson driven deterministic two-node tandem queues with blocking

Hochang Lee, Dong-Won Seo *

School of Management, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, 130-701, South Korea

ARTICLE INFO

Article history:
Received 6 August 2014
Received in revised form
3 February 2015
Accepted 3 February 2015
Available online 11 February 2015

Keywords:
Max-plus algebra
Blocking
Higher moments
Tail probability
Stationary waiting time
Tandem queue

ABSTRACT

There has been no analytic expression for a multi-node queue even with constant processing times because of a correlation caused by blocking between nodes. This study introduces explicit expressions for moments and tail probability of stationary waiting times in a Poisson-driven deterministic 2-node tandem queue with blocking. Unlike the normal queueing theory, we derive these expressions from the previous results based on the max-plus algebraic approach. Two blocking policies are considered: blocking before service (BBS) and blocking after service (BAS).

1. Introduction

Analytical solutions are attractive to researchers interested in performance evaluations of prevalent systems in areas such as production (manufacturing) systems and telecommunications networks. However, obtaining general analytical solutions is not a simple task.

Although single-server single-node queues have been studied extensively and applied in the evaluation of various systems, there are few analytical solutions for multi-node non-Markovian queues using a computational perspective. Moreover, results regarding finite capacity queues are small in number because of blocking phenomena caused by finite capacities.

Obtaining explicit expressions for characteristics of stationary waiting time in finite-capacity queues is difficult, with some exceptions such as $M/M/1/K$, $M/G/c/c$, $M/G/1/1$, and $M/G/1/2$ queues (see, e.g., Takagi [15] and Gross and Harris [10]). Tijms [16] first provided the recursive formula for stationary distributions in $M/G/c/K$ queues. By using the generating function (z-transform), Brun and Garcia [8] showed analytical solutions of steady-state probability distributions in finite-capacity $M/D/1$ queues. Alouf [1] analytically derived the stationary distribution of the $M/D/1/K$ queue using Cohen's results of the $M/G/1/K$ queue (see Cohen [9]). For more general queues, on the other hand, most studies have introduced various approximation methods. For instance, Perros [12] numerically demonstrated several relationships between blocking policies and introduced a variety of approximation methods, all of which are forms of a weighted combination of exact (if available) expressions of two queues with deterministic and exponential distributions. Similarly, Smith [14] presented an approximation for an $M/G/c/K$ queue based on closed-form expressions derived from the finite-capacity exponential and deterministic queues.

The normal queueing theory has limitations in applying to general queues with multi-node, multi-server, or generally structured queues. As an effort to overcome these shortcomings, we adopt an alternative method, the so-called max-plus algebraic approach. Many types of networks belonging to a class of queueing networks, the so-called max-plus linear system, can be modeled properly by Timed Event Graphs, a special case of the Petri net. They can be analyzed using max-plus algebra, involving only two operators: 'max' and '+'. In brief, a max-plus linear system is a choice-free network of single-server queues with FIFO service discipline. As it is clear that a single-server two-node tandem queue with blocking belongs to a max-plus linear system, max-plus algebra is a useful tool in analyzing this finite-capacity queue. Additionally, as constant service times render the series expressions (see Section 2) simple and tractable, we focused solely on deterministic service times in this study.

This study aims to demonstrate closed-form (transform-free) expressions for higher moment and tail probability of stationary

* Corresponding author. Tel.: +82 2 9610396.
E-mail addresses: hochang@khu.ac.kr (H. Lee), dwseo@khu.ac.kr (D.-W. Seo).

http://dx.doi.org/10.1016/j.orl.2015.02.002
0167-6377/

waiting times in a deterministic two-node tandem queue with a Poisson arrival process and blocking. Blocking is typically represented by two policies: blocking before service (denoted as BBS or communication blocking) and blocking after service (denoted by BAS or production blocking) policies. The communication blocking policy allows a customer at node j to be served only when there is a vacant space in the buffer at node $j+1$. Therefore, under BBS, a customer at node j cannot begin his service unless there is an available buffer space at node $j+1$. On the other hand, under the production blocking policy, a customer can be served even if there is no vacant space in the buffer at node $j+1$, and the customer at node j is only allowed to move to node $j+1$ when there is a vacant space in the buffer at the moment node $j+1$ is available. The blocked customer stays in node j until a vacant space is available. During that time, node j is blocked from serving other customers.

The paper is organized as follows. Section 2 introduces the brief preliminaries on waiting time in terms of max-plus algebra. Our main results, the explicit expressions for moments and tail probabilities of stationary waiting times, are given in Section 3. The proof and numerical results are provided in Sections 4 and 5, respectively. The paper ends with concluding remarks.

2. Waiting times in max-plus linear systems

The basic reference algebra used throughout this study is the so-called max-plus algebra on the real line $\mathbb{R}$; namely, the semi-field with the two operations $(\oplus, \otimes)$ in which $\oplus$ refers to maximization and $\otimes$ refers to addition for scalars and max-plus algebra product for matrices (see Baccelli et al. [4]). The dynamics of max-plus linear systems with α nodes can be described by the α-dimensional vectorial recurrence equations

$$X_{n+1} = A_n \otimes X_n \oplus B_{n+1} \otimes T_{n+1} \tag{1}$$

with an initial condition of X_0, where $\{T_n\}$ is a non-decreasing sequence of real-valued random numbers (e.g. the epochs of the Poisson arrival process with rate λ), $\{A_n\}$ and $\{B_n\}$ are stationary and ergodic sequences of real-valued random matrices of size $\alpha \times \alpha$ and $\alpha \times 1$, respectively, and $\{X_n\}$ is a sequence of α-dimensional state vectors referring to the absolute time of the beginning of the nth service at each node. The components of the state vector represent absolute times that grow to ∞ when n increases unboundedly, and hence, one is more interested in the differences $W_n^i = X_n^i - T_n$ (like the waiting time of the nth customer until he joins server i). Let $\tau_n = T_{n+1} - T_n$ with $T_0 = 0$ and let $C(x)$ be the $\alpha \times \alpha$ matrix with all diagonal entries equal to $-x$ and all non-diagonal entries equal to $-\infty$. By subtracting T_{n+1} from both sides of (1), the new state vector W_{n+1} can be expressed as

$$W_{n+1} = A_n \otimes C(\tau_n) \otimes W_n \oplus B_{n+1},$$

for $n \geq 0$ and with the initial condition W_0. Baccelli et al. [4] previously demonstrated that under certain conditions, the dynamics of Poisson driven max-plus linear systems could be described by vectorial recurrence equations (also see Heidergott [11]). For all $\lambda < a^{-1}$, where λ is the arrival rate and a is the maximal Lyapunov exponent of the sequence $\{A_n\}$, a stationary waiting time W is determined by the matrix-series

$$W = D_0 \oplus \bigoplus_{k \geq 1} C(T_{-k}) \otimes D_k$$

with $D_0 = B_0$, $W_0 = B_0$, and for all $k \geq 1$

$$D_k = \left(\bigotimes_{n=1}^{k} A_{-n} \right) \otimes B_{-k}.$$

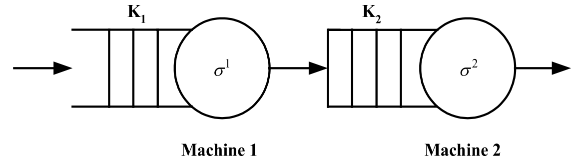

Fig. 1. 2-node tandem queue with a finite buffer K_2.

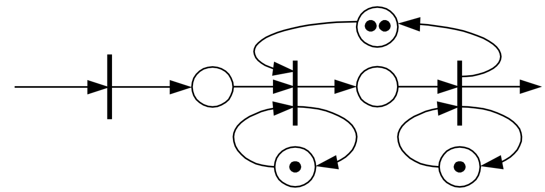

Fig. 2. Event graph for a 2-node tandem queue with a finite buffer $K_2 = 2$.

From this topology, Baccelli and Schmidt [7] obtained characteristics of waiting times in a class of stochastic networks as Taylor series expansions with respect to an arrival rate λ (also see Ayhan and Seo [2,3]). Note that the random vector D_n plays an important role in computing characteristics of waiting times, and its calculation is independent of the arrival rate. Additionally, the components of the random vector refer to the time elapsed from the arrival to the beginning of service at each node, and they can be interpreted as a critical path in a task graph.

Once the explicit expression of the random vector D_k for a max-plus linear system is derived, we can compute various characteristics of transient and stationary waiting times by inputting the expression into the series expansions given in Baccelli and Schmidt [7], Baccelli et al. [5,6], and Ayhan and Seo [2,3]. However, it is usually quite difficult to derive closed-form expressions for stationary waiting times. So, they assumed the ith element of the sequence $\{D_n\}$ to be 'ultimately periodic'. That is, in a class of max-plus linear system with constant service times, the ith elements of $\{D_n\}$ is given by

$$D_n^i = \begin{cases} \eta_n^i & \text{for } n = 0, \dots, \xi_i - 1 \\ \eta_{\xi_i}^i + (n - \xi_i) a_i & \text{for } n \geq \xi_i \end{cases} \tag{2}$$

for the constant real numbers $0 \leq \eta_0^i \leq \eta_1^i \leq \cdots \leq \eta_{\xi_i}^i$, a_i and some non-negative integers ξ_i. Not all deterministic max-plus linear systems fall into this category. However, this does cover many interesting queueing systems with deterministic service times such as tandem queues with various types of blocking, fork-and-join type queues, queueing networks embedded in Kanban systems, etc. By placing the explicit expressions for D_n^i satisfying the structure (2) into the closed-form formulas given by Ayhan and Seo [2, Theorem 3.5, Corollary 3.1] and [3, Theorem 2.3], we can compute the values of the Laplace transform and the higher moment and tail probability of stationary waiting times. Seo [13] applied their method to finite-capacity 2-node tandem queues with constant service times, assuming that the first node has infinite capacity but that the second node has finite capacity. Under BBS and BAS blocking policies, the values of expected waiting times are computed from the Taylor series expansion.

In Petri nets, a node's finite capacity can be treated by the number of markings (tokens) in a place (see Baccelli et al. [4]). Under BBS policy, hence, a 2-node tandem queue with a finite buffer capacity at node 2 shown in Fig. 1 can be depicted as the corresponding event graph in Fig. 2. These two systems behave in exactly the same way. Inserting a dummy transition then makes it possible to transform the corresponding (finite capacity) event graph into an infinite capacity event graph (see Fig. 3), which becomes tractable in the sense of drawing expressions.