# Explicit Formulae for Characteristics of Finite-Capacity M/D/1 Queues

Dong-Won Seo

**Even though many computational methods (recursive formulae) for blocking probabilities in finite-capacity M/D/1 queues have already been produced, these are forms of transforms or are limited to single-node queues. Using a distinctly different approach from the usual queueing theory, this study introduces explicit (transform-free) formulae for a blocking probability, a stationary probability, and mean sojourn time under either production or communication blocking policy. Additionally, the smallest buffer capacity subject to a given blocking probability can be determined numerically from these formulae. With proper selection of the overall offered load $\rho$, the approach described herein can be applicable to more general queues from a computational point of view if the explicit expressions of random vector $D_n$ are available.**

**Keywords: Blocking probability, finite-capacity queues, max-plus algebra, max-plus linear system, stationary probability, system sojourn time.**

## I. Introduction

An M/D/1 queue, the simplest queue with deterministic service time, has been studied extensively and has a variety of applications in the performance evaluation of production management, telecommunications networks, and other areas. However, because blocking phenomena caused by finite capacity raise many difficulties and introduce complexity into evaluations of various performance measures, results regarding finite-capacity queues are small in number.

With the exception of some special cases, such as Markovian queues (M/M/1/K), Erlang's loss queues (no waiting room at all, M/G/c/c), M/G/1/1 and M/G/1/2 queues, and so on, it is very difficult to obtain explicit expressions for a blocking probability or a stationary distribution in finite-capacity queues (for example, see Gross and Harris [1] and Takagi [2]).

On the one hand, the recursive formula for stationary distributions (or blocking probabilities) in M/G/c/K queues was first provided by Tijms [3]. Brun and Garcia [4] showed analytical (transform-free) solutions of steady-state probability distributions in finite-capacity M/D/1 queues via the use of the generating function (z-transform). Alouf and others [5] analytically derived the stationary distribution of the M/D/1/K queue using Cohen's results — which were based on the Laplace-Stieltjes transform (LST) — of the M/G/1/K queue of the service time distribution (see Cohen [6]). On the other hand, most studies have introduced various approximation methods for more general queues. For instance, Perros [7] numerically demonstrated several relations 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

Manuscript received Aug. 19, 2013; revised Nov. 1, 2013; accepted Nov. 8, 2013.

This research was supported by grant from Kyung Hee University in 2011 (KHU-20110900).

Dong-Won Seo (corresponding author, dwseo@khu.ac.kr) is with Kyung Hee School of Management and also with the Management Research Institute, Kyung Hee University, Seoul, Rep. of Korea.

distributions. Similarly, Smith [8] presented an approximation for an M/G/c/K queue based on closed-form expressions derived from finite-capacity exponential and deterministic queues. Sakasegawa and others [9] constructed an approximation formula for the overflow probability for GI/GI/c/N queues in terms of a queue-length distribution for the corresponding GI/GI/c/∞ queues.

The blocking probability plays an important role in the analysis of finite-capacity queues. Once a blocking probability is obtained, we are able to compute various performance measures of interest, including carried load, stationary probabilities, (higher) moments of stationary waiting time and sojourn time, and so forth. However, normal queueing theory has limitations in applications to general queues with multi-node, multi-server, or generally structured queues.

In an effort to overcome these shortcomings we adopt a different method from the usual queueing theory — namely, max-plus algebra. 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 analysed using max-plus algebra, which involves the use of only two operators: "max" and "+." In brief, a max-plus linear system is a choice-free network of single-server queues with first-in first-out service discipline.

Because it is clear that an M/G/1/K queue belongs to a max-plus linear system, a max-plus algebra is useful in analysing this finite-capacity queue. Additionally, constant service times render the series expressions (see section II) simple and tractable such that we focused solely on deterministic service times. The principal objective of this study is to demonstrate a closed-form (transform-free) expression of a blocking probability for an M/D/1/K queue under either communication or production blocking policy. In the case of communication blocking policy we can obtain the same expression as the one of Brun and Garcia [4] and Alouf and others [5], whereas in the case of production blocking policy we obtain a new formula. Moreover, other related expressions for stationary probability, mean system sojourn time, and an optimal buffer capacity are also provided. Similar to Sakasegawa and others [9], these expressions are written in terms of the blocking probability and the queue-length distribution of the corresponding M/D/1/∞ queue.

This paper is organized as follows. Brief preliminaries on max-plus algebra and on waiting times in a max-plus linear system are provided in section II. Section III includes our principal results on the explicit expression of blocking probability in finite-capacity M/D/1 queues. We show other related expressions in section IV and end with some concluding remarks in section V.

## II. Brief Preliminaries

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 the $\oplus$ refers to maximization and the $\otimes$ refers to addition for scalars and max-plus algebra product for matrices (see Baccelli and others [10]). The dynamics of a max-plus linear system with $\alpha$ nodes can be described by the $\alpha$-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 (for example, the epochs of the Poisson arrival process with rate $\lambda$), $\{A_n\}$ 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 $\alpha$-dimensional state vectors. The components of the state vector $X_n$ represent absolute times that grow to $\infty$ when $n$ increases unboundedly; hence, one is more interested in the differences $W_n^i = X_n^i - T_n$ (like the waiting time of the $n$th customer until they join 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 and others [10] 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 $\lambda$ is the arrival rate and $a$ is the maximal Lyapunov exponent of the sequence $\{A_n\}$, $W$ is determined by the matrix series

$$W = D_0 \oplus \bigoplus_{k \geq 1} C(T_{-k}) \otimes D_k, \tag{2}$$

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

$$D_k = (\bigotimes_{n=1}^{k} A_{-n}) \otimes B_{-k}. \tag{3}$$

From this topology, Baccelli and Schmidt [12] derived a Taylor series expansion for mean stationary waiting time with regard to the arrival rate in a Poisson-driven max-plus linear system. Their approach was generalized to other characteristics of stationary and transient waiting times, such as higher moments, Laplace transforms, and tail probabilities by Baccelli and others [13]–[14] and Ayhan and Seo [15]–[16]. Later, Heidergott [11] established Taylor series expansions with

regard to general parameters for max-plus linear systems with a renewal arrival process. The representation of the stationary waiting time in (2) and (3) holds for any max-plus linear system with a renewal arrival process. However, we assume the Poisson arrival process throughout this study to use the existing explicit results for the stationary waiting time. The reader can refer to the works of Baccelli and others [10] and Heidergott and others [17] for more details on basic max-plus algebra and to Baccelli and Schmidt [12], Baccelli and others [13]–[14], and Ayhan and Seo [15]–[16] for details on waiting times in a max-plus linear system.

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 it into the series expansions given in Baccelli and Schmidt [12], Baccelli and others [13]–[14], and Ayhan and Seo [15]–[16]. However, it is usually quite difficult to derive closed-form expressions for stationary waiting times. So, all authors in [12]–[16] assumed the *i*th element of the sequence $\{D_k\}$ to be 'ultimately periodic'; that is, in a class of max-plus linear systems with constant service times the *i*th element of $\{D_n\}$ is given by

$$D_n^i = \begin{cases} \eta_n^i & \text{for } n = 0, \ldots, \xi_i - 1, \\ \eta_{\xi_i}^i + (n - \xi_i)a_i & \text{for } n \geq \xi_i, \end{cases} \tag{4}$$

for 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 $\xi_i$. 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, queuing networks embedded in Kanban systems, and so on. By placing the explicit expressions for $D_n^i$ that satisfy the structure of (4) into the closed-form formulae given by Ayhan and Seo [15]–[16], we can compute the values of the Laplace transform of stationary waiting times, higher moments of stationary waiting times, and tail probability of stationary waiting times.

Recently, Seo [18] applied the method used by authors in [12]–[16] to finite-capacity 2-node tandem queues with constant service times, in which he assumed the first node to have infinite capacity but the second to have finite capacity. For this model, he considered two blocking policies: communication (blocking before service) and production (blocking after service). Under communication blocking, a customer at node *i* cannot begin service unless there is a vacant space in the buffer at node *j*+1. On the other hand, under production blocking, a customer served at node *j* moves to node *j*+1 only if the buffer of node *j*+1 is not full; otherwise, the blocked customer remains in node *j* until a vacancy becomes available. During that time, node *j* is blocked from serving other customers.

The aim of this study is to introduce explict expressions for blocking probabilities in M/D/1/K queues under two blocking policies. To the best of our knowledge, the explicit blocking formula under a production blocking policy has not been introduced, because it is not easy to handle with the usual queueing theory. However, there are a few results under a communication blocking policy in the literature, because it is more comfortable to treat.

The random vector $D_n$, $n \geq 0$, plays an essential part in deriving and computing explicit expressions for a blocking probability and other related characteristics of interest, as shown in the section below.

## III. Explicit Formulae for Blocking Probability

The following expressions of $D_n^i$, $i$ = 1, 2, under either a communication or a production blocking policy are given in Seo [18]. For node *i*, $i$ = 1, 2, let $\sigma^i$ be a deterministic service time and $K_i$ be a finite capacity with $K_1 = \infty$ and $K_2 < \infty$.

• Production blocking policy

If $K_2 \geq 1$,

$$D_n^1 = n\sigma^1 \quad \text{for } 0 \leq n \leq K_2, \tag{5}$$

$$D_n^1 = \max\left\{n\sigma^1, \sigma^1 + (n - K_2)\sigma^2\right\} \quad \text{for } n > K_2, \tag{6}$$

$$D_n^2 = \sigma^1 + \max\left\{n\sigma^1, n\sigma^2\right\} \quad \text{for } n \geq 0. \tag{7}$$

• Communication blocking policy

If $K_2 = 1$,

$$D_n^1 = n(\sigma^1 + \sigma^2) \quad \text{for } n \geq 0, \tag{8}$$

$$D_n^2 = \sigma^1 + n(\sigma^1 + \sigma^2) \quad \text{for } n \geq 0. \tag{9}$$

If $K_2 \geq 2$,

$$D_n^1 = n\sigma^1 \quad \text{for } 0 \leq n < K_2, \tag{10}$$

$$D_n^1 = \max\left\{n\sigma^1, \sigma^1 + (n - K_2 + 1)\sigma^2\right\}, \quad \text{for } n \geq K_2, \tag{11}$$

$$D_n^2 = \sigma^1 + \max\left\{n\sigma^1, n\sigma^2\right\} \quad \text{for } n \geq 0. \tag{12}$$

The series expansion using max-plus algebra assumes the stability condition ($\rho$ < 1) and unlimited capacity at the first node. Thus, an M/D/1/K queue must be transformed into the corresponding M/D/1/$\infty$ queue, which can be tractable through existing results such as the explicit expressions for the moment and for the tail probability of stationary waiting times.

Let $\sigma$ be a deterministic service time and $K$ ($\geq$ 2) be a finite

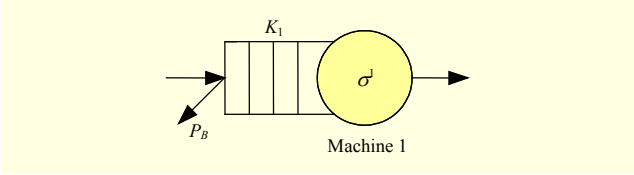


Fig. 1. M/D/1/4 queue.

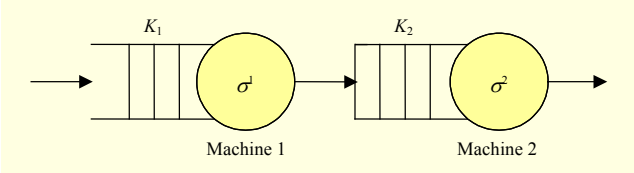


Fig. 2. M/D/1/∞ → ·/D/1/4 queue.

capacity for an M/D/1/K queue. To assess the finite-capacity M/D/1 queue, we extend it to a corresponding 2-node tandem queue by inserting a dummy node with zero service time and an infinite-capacity buffer in front of this M/D/1/K queue, both of which are depicted in Figs. 1 and 2 under communication blocking policy.

The corresponding extended 2-node tandem queue then has $\sigma^1 = 0$, $\sigma^2 = \sigma$, $K_1 = \infty$, and $K_2 = K$. By putting these $\sigma^i$ and $K_i$ into equations (5)–(7), and (10)–(12) we have, for $K \geq 2$.

• Production blocking policy

$$D_n^1 = 0 \quad \text{for } 0 \leq n \leq K, \tag{13}$$

$$D_n^1 = (n-K)\sigma \quad \text{for } n > K, \tag{14}$$

$$D_n^2 = n\sigma \quad \text{for } n \geq 0.$$

• Communication blocking policy

$$D_n^1 = 0 \quad \text{for } 0 \leq n < K, \tag{15}$$

$$D_n^1 = (n-K+1)\sigma \quad \text{for } n \geq K, \tag{16}$$

$$D_n^2 = n\sigma \quad \text{for } n \geq 0.$$

We are now ready to present our main idea. A customer at the first node is blocked, either before service or after service, depending on the blocking policy. The customer must remain at the first node for a certain period of time (prior to moving to the second node) whenever there are $K$ customers in the second node whose capacity is $K$. Recall that this blocking phenomenon is captured in the expression of $D_n$. On this point, the event that a customer's waiting time is greater than zero at the first node in the corresponding extended 2-node tandem queue is equivalent to the event that there are $K$ or more customers in the M/D/1/∞ queue. Recall that we assume the service time at the first node to be both deterministic and zero.

Therefore, we can conclude the following: the probability $E_K$, that $K$ or more customers are in an M/D/1/∞ queue at an arbitrary time, is equal to the probability that the stationary waiting time at the first node is greater than zero in the corresponding extended 2-node tandem queue. That is,

$$E_K = \sum_{j=K}^{\infty} \pi_j^{\infty} = \Pr(W^1 > 0), \tag{17}$$

where $\pi_j^{\infty}$ is a stationary probability that there are $j$ customers in an M/D/1/∞ queue at an arbitrary time and $W^1$ is a stationary waiting time at the first node in the corresponding extended 2-node tandem queue. More precisely, $W^1$ is the elapsed time from the arrival until the beginning of service at node 1.

The probability $E_K$ performs an important role in this study. Now, we generate the following theorem from the well-known simple blocking formula (see, for example, Takagi [2]) and the explicit expression for the tail probability of stationary waiting times shown in Ayhan and Seo [16].

**Theorem 1.** For an M/D/1/K queue with arrival rate λ and constant service time $\sigma$, the blocking probability $P_B$ is given by

$$P_B = \frac{(1-\rho)E_K}{1-\rho E_K}, \tag{18}$$

where the offered load $\rho = \lambda\sigma < 1$ and under a production blocking policy

$$E_K = 1-(1-\rho)\sum_{j=0}^{K} \frac{(-1)^j \rho^j (K-j)^j e^{\rho(K-j)}}{j!}. \tag{19}$$

***Proof.*** Because (17) already shows the relation between the tail probability of stationary waiting times and the probability of $E_K$, it suffices to show the explicit expression of $E_K$ for an M/D/1/∞ queue.

From (13) and (14), the expression of the random vector $D_n^1$ for the corresponding extended 2-node tandem queue can be written as

$$D_n^1 = \begin{cases} 0 & \text{for } n = 0, \ldots, K, \\ (n-K)\sigma & \text{for } n \geq K+1. \end{cases}$$

Then, it satisfies the structure provided in (4), and we can see that $\xi_1 = K+1$, $a_1 = \sigma$, $\eta_n^1 = 0$ for $0 \leq n \leq K$, $\eta_{\xi_1}^1 = \sigma$, and $\kappa = K+1$ when $t = 0$. Therefore, it satisfies the second case of Theorem 2.3 in [16] because $\xi_1 = K+1 > 0$ and $\eta_{\xi_1-1}^1 = \eta_K^1 = 0$. Note that $\kappa$ defined therein equals $K$ when $t = 0$ since

$$\begin{aligned} \kappa &= \min\left\{k \in (0,1,\ldots) : ka_1 > t - \eta_{\xi_1}^1 + \xi_1 a_1\right\} \\ &= \min\left\{k \in (0,1,\ldots) : k\sigma > 0 - \sigma + (K+1)\sigma\right\} \\ &= K+1. \end{aligned}$$

Thus, the probability $E_K$ is expressed as

$$E_K = \Pr(W^1 > 0) = 1-(1-\rho)\sum_{j=0}^{K}\frac{(-1)^j \rho^j (K-j)^j e^{\rho(K-j)}}{j!},$$

where $\rho = \lambda\sigma < 1$. We can then compute the blocking probability $P_B$ for an M/D/1/K queue using the simple blocking formula in (18). □

Our approach is also valid for a communication blocking policy. Similarly as above, we can obtain the following corollary without the proof. From (15) and (16), the expression of $D_n^1$ under a communication blocking policy can be written as

$$D_n^1 = \begin{cases} 0 & \text{for } n = 0, \dots, K-1, \\ \sigma + (n-K)\sigma & \text{for } n \geq K. \end{cases}$$

It also satisfies the structure given in (14). The fact that $\xi_1 = K$, $a_1 = \sigma$, $\eta_n^1 = 0$ for $0 \leq n \leq K-1$, and $\eta_{\xi_1}^1 = \sigma$ satisfies the second case of Theorem 2.3 in [16], because $\xi_1 = K > 0$ and $\eta_{\xi_1 - 1}^1 = \eta_{K-1}^1 = 0$. The $\kappa$ defined therein equals $K$ when $t = 0$ yields the probability $E_K$ as follows.

**Corollary 1.** For an M/D/1/K queue with arrival rate λ and constant service time $\sigma$, the blocking probability $P_B$ is given by

$$P_B = \frac{(1-\rho)E_K}{1-\rho E_K},$$

where the offered load $\rho = \lambda\sigma < 1$ and under a communication blocking policy

$$E_K = 1-(1-\rho)\sum_{j=0}^{K-1}\frac{(-1)^j \rho^j (K-j-1)^j e^{\rho(K-j-1)}}{j!}. \quad (20)$$

For a production blocking policy there is no similar result to Theorem 1 in the literature. However, we can see that our formula for a communication blocking policy is equivalent to the ones provided in [3]–[5].

**Remark 1.** Tijms [3] demonstrated that $P_B = \left(1-\rho+(\rho-1)\sum_{j=0}^{K-1}\pi_j^{\infty}\right)\left(1-\rho+\rho\sum_{j=0}^{K-1}\pi_j^{\infty}\right)^{-1}$. From the relation $\sum_{j=0}^{K-1}\pi_j^{\infty} = 1-E_K$, we can readily obtain the equivalent expression to ours. Besides, the fact that $E_1 = \rho$ when $K = 1$ derives the same formula mentioned in Takagi [2] as follows:

$$P_B = \frac{(1-\rho)E_1}{1-\rho E_1} = \frac{(1-\rho)\rho}{1-\rho^2} = \frac{\rho}{1+\rho}.$$

**Remark 2.** Brun annd Garcia [4] introduced analytical solutions for the steady-state distribution in a finite-capacity M/D/1 queue by using the generating function, which is completely different from our approach. Their key measure $b_n$ is defined as $b_0 = 1$ and for $n \geq 1$,

$$b_n = \sum_{j=0}^{n}\frac{(-1)^j \rho^j (n-j)^j e^{\rho(n-j)}}{j!}.$$

It is easy to find the following relation between the two measures $E_n$ and $b_n$:

$$E_0 = 1$$

and

$$E_n = 1-(1-\rho)b_{n-1} \text{ for all } n \geq 1.$$

**Remark 3.** For the stationary distribution of the M/D/1/K queue, Alouf and others [5] defined a measure $\alpha_j(\rho)$ as $\alpha_1(\rho) = 1$ and for $j \geq 2$,

$$\alpha_j(\rho) = \sum_{i+k=j-2}\frac{(-1)^k \rho^k (i+1)^k e^{\rho(i+1)}}{k!}.$$

The two measures $E_j$ and $\alpha_j(\rho)$ are related as follows:

$$E_1 = \rho$$

and

$$E_j = 1+(\rho-1)\alpha_j(\rho) \text{ for all } j \geq 2.$$

Remarks 2 and 3 show that our formula is equivalent to the ones provided in [4] and [5]. Thus, we could assert that our formula also works even when $\rho > 1$ without a mathematical proof because theirs hold regardless of the value of the offered load $\rho$. Such a proof will be carried out entirely differently from the approach described in [12] due to the stability condition — one of the fundamental assumptions of the Taylor series expansions.

**Remark 4.** Let $E_{K,C}$ and $E_{K,P}$ be the $E_K$ for a communication and a production blocking policy, respectively. For a given $K$ and $\rho$, we can see from the expressions given in (19) and (20) and the definition of $E_{K,C}$ that

$$E_{K,C} \geq E_{K+1,C} = E_{K,P}.$$

This relationship shows $E_{K,P} \leq E_{K,C}$. For a given $K$ and $\rho$, therefore, a production blocking policy provides less blocking probabilities than a communication blocking policy, since the blocking formula given in (18) is increaing in $E_K$.

Our explicit expressions allow us to compute exact blocking probabilities under two blocking policies. Figure 3 compares the blocking probabilities with varying $K$ when $\sigma = 5$ and $\lambda = 0.19$. It shows that the blocking after service (BAS) policy produces less blocking probabilities than the blocking before service (BBS) policy under the same environments; however, this difference becomes negligible as finite capacity increases.

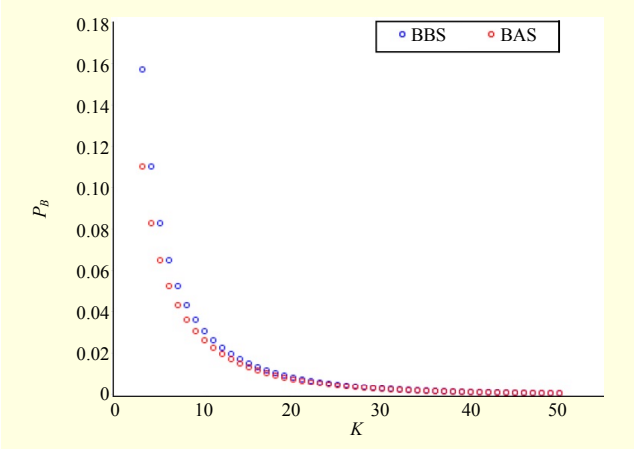


Fig. 3. Blocking probabilities with varying $K$.

Table 1. $P_B$ when $(K_1, K_2) = (5, 5)$.

| $\lambda$ | $\sigma^1 = 0.1, \sigma^1 = 0.9$ | | $\sigma^1 = 0.5, \sigma^2 = 0.9$ | |
|---|---|---|---|---|
| | $P_{B,S}$ | $P_{B,P}$ | $P_{B,S}$ | $P_{B,P}$ |
| 0.1 | 0.00000 ∓ 0.00000 | 8.2712E-13 | 0.00000 ∓ 0.00000 | 3.5334E-09 |
| 0.2 | 0.00000 ∓ 0.00000 | 3.2398E-11 | 1.8766E-07 ∓ 9.8926E-08 | 1.4446E-07 |
| 0.3 | 0.00000 ∓ 0.00000 | 1.3898E-09 | 1.6671E-06 ∓ 2.5313E-07 | 1.3419E-06 |
| 0.4 | 0.00000 ∓ 0.00000 | 5.7056E-08 | 8.4639E-06 ∓ 5.0956E-07 | 6.7800E-06 |
| 0.5 | 1.2009E-06 ∓ 4.6458E-07 | 1.2237E-06 | 3.4159E-05 ∓ 8.9659E-07 | 2.6045E-05 |
| 0.6 | 1.7674E-05 ∓ 1.9784E-06 | 1.5954E-05 | 1.2595E-04 ∓ 1.6398E-06 | 9.8234E-05 |
| 0.7 | 1.4309E-04 ∓ 5.2651E-06 | 1.4333E-04 | 4.8694E-04 ∓ 3.6751E-06 | 4.1745E-04 |
| 0.8 | 9.4617E-04 ∓ 1.6021E-05 | 9.4839E-04 | 0.00196 ∓ 8.2606E-06 | 0.00183 |
| 0.9 | 0.00470 ∓ 4.3470E-05 | 0.00473 | 0.00726 ∓ 1.8768E-05 | 0.00705 |
| 1.0 | 0.01747 ∓ 9.1046E-05 | 0.01754 | 0.02218 ∓ 3.4622E-05 | 0.02195 |
| 1.1 | 0.04738 ∓ 1.5243E-04 | 0.04748 | 0.05307 ∓ 5.0337E-05 | 0.05289 |

**Remark 5.** Whereas the approaches used in [4] and [5] are restrictive in the number of queues, our method remains valid for single-server multi-node queues if the explicit expressions of the random vector $D_n$ are available. For instance, Seo and others [19] numerically demonstrated that our approach is applicable to compute blocking probabilities in M/D/1/$K_1$ → ·/D/1/$K_2$ queues under a communication blocking policy. Table 1 shows the blocking probabilities, that is, $(P_{B,S})$ computed by simulation and $(P_{B,P})$ computed by our method when $\rho = \lambda \max\{\sigma^1, \sigma^2\} < 1$, where $\sigma^1$ and $\sigma^2$ are the constant service times at node 1 and 2, respectively. Their computational results show that our approach works quite well as $\rho$ increases. However, to obtain more accurate values of the blocking probabilities, further study is necessary to determine a method for the appropriate selection of an overall offered load $\rho$ for single-server multi-node systems.

## IV. Other Related Expressions

We can obtain the following explicit expressions of stationary distributions and mean system sojourn time in an M/D/1/∞ queue and an M/D/1/K queue, which are written in terms of the probability $E_K$. As we mentioned before, because the dynamic behaviors depending on a blocking policy are captured in the expression of $D_n$, we do not distinguish the two blocking policies in the below.

### 1. Explicit Formula for Stationary Probability

From the definition of $E_m$ for all $m \geq 0$, we have an explicit formula for the stationary probability $\pi_m^\infty$.

**Theorem 2.** In the M/D/1/∞ queue with $\rho = \lambda\sigma < 1$, the stationary probability $\pi_m^\infty$ is given by

$$\pi_m^\infty = E_m - E_{m+1} \text{ for all } m \geq 0,$$

where $E_m = \sum_{j=m}^{\infty} \pi_j^\infty$. More precisely,

$$\pi_m^\infty = (1-\rho)\left[\sum_{\ell=1}^{m} \frac{(-1)^\ell \rho^\ell (m-\ell)^{\ell-1} e^{\rho(m-\ell)}}{\ell!} \times \{(m-\ell) + \frac{\ell}{\rho}\} + e^{\rho m}\right].$$

***Proof.*** It is clear that $\pi_m^\infty = E_m - E_{m+1}$ from the definition of $E_j = \sum_{n=j}^{\infty} \pi_n^\infty$. From the explicit expression of $E_K$ in (13), along with some algebra, we can obtain the following:

$$\begin{aligned}
\pi_m^\infty &= E_m - E_{m+1} \\
&= \left[1 - (1-\rho)\sum_{j=0}^{m-1} \frac{(-1)^j \rho^j (m-j-1)^j e^{\rho(m-j-1)}}{j!}\right] \\
&\quad - \left[1 - (1-\rho)\sum_{j=0}^{m} \frac{(-1)^j \rho^j (m-j)^j e^{\rho(m-j)}}{j!}\right] \\
&= (1-\rho)\left[\sum_{j=0}^{m} \frac{(-1)^j \rho^j (m-j)^j e^{\rho(m-j)}}{j!} \right. \\
&\quad \left. - \sum_{\ell=1}^{m} \frac{(-1)^{\ell-1} \rho^{\ell-1} (m-\ell)^{\ell-1} e^{\rho(m-\ell)}}{(\ell-1)!}\right] \\
&= (1-\rho)\left[\sum_{j=1}^{m} \frac{(-1)^j \rho^j (m-j)^j e^{\rho(m-j)}}{j!} + e^{\rho m} \right. \\
&\quad \left. + \sum_{\ell=1}^{m} \frac{(-1)^\ell \rho^\ell (m-\ell)^{\ell-1} e^{\rho(m-\ell)}}{\ell!} \left(\frac{\ell}{\rho}\right)\right],
\end{aligned}$$

which completes the proof. □

Stationary probability at a departure time can be written as follows. We omitted the proof because it can be readily proven by Theorem 2. Then, stationary probabilities at an arbitrary time can be computed by the well-known relation $P_m = (1-P_B)\pi_m$ (see [6] for an example). In an M/D/1/K queue, for $0 \le m \le K-1$, stationary probabilities $P_m$ at an arbitrary time and $\pi_m$ at a departure time are computed by

$$\pi_m = \frac{E_m - E_{m+1}}{1-E_K},$$

where $E_m = \sum_{j=m}^{\infty} \pi_j^{\infty}$

## 2. Explicit Formula for Mean Sojourn Time

Before moving to the mean system sojourn time, we first demonstrate an expression written in terms of the probabilities $E_j$, $j = 1, \ldots, K$, for the expected number of customers in an M/D/1/∞ queue. This is the mean value of the number of customers truncated by $K$–1 and is useful in deriving expressions for the mean system sojourn time $W$ in an M/D/1/K queue.

**Lemma 1.**

$$\sum_{j=1}^{K-1} j\pi_j^{\infty} = \sum_{j=1}^{K} E_j - KE_K,$$

where $E_m = \sum_{j=m}^{\infty} \pi_j^{\infty}$ and $\pi_j^{\infty}$ is the probability at an arbitrary time that there are $j$ customers in the M/D/1/∞ queue.

***Proof.*** With the help of relations $\sum_{j=1}^{\infty} j\pi_j^{\infty} = \sum_{j=1}^{\infty} E_j$ and $\sum_{j=K+1}^{\infty} (j-K)\pi_j^{\infty} = \sum_{j=K+1}^{\infty} E_j$, we can see that

$$\begin{aligned}
\sum_{j=1}^{K-1} j\pi_j^{\infty} &= \sum_{j=1}^{\infty} j\pi_j^{\infty} - \sum_{j=K}^{\infty} j\pi_j^{\infty} \\
&= \sum_{j=1}^{\infty} E_j - \left( K\pi_K^{\infty} + \sum_{j=K+1}^{\infty} j\pi_j^{\infty} \right) \\
&= \sum_{j=1}^{\infty} E_j - \left[ K\pi_K^{\infty} + \sum_{j=K+1}^{\infty} (j-K)\pi_j^{\infty} + K \sum_{j=K+1}^{\infty} \pi_j^{\infty} \right] \\
&= \sum_{j=1}^{K} E_j - KE_K,
\end{aligned}$$

which completes the proof. □

By using Little's law and Lemma 1, we have the following explicit expressions for a mean system sojourn time $W$ in an M/D/1/K queue, which are written in terms of $E_K$. The mean system size (the number of customers in the system) is also computed via this expression.

**Theorem 3.** The expected system sojourn time $W$ in an M/D/1/K queue is expressed as follows:

$$W = \frac{1}{\lambda}\left(\frac{1}{1-E_K}\right)\left(\sum_{j=1}^{K} E_j\right) + \frac{K}{\lambda}\left(1 - \frac{1}{1-E_k}\right)\rho.$$

***Proof.*** Let $L$ be the mean number of customers of an M/D/1/K queue. From Little's law and the proportional relation between state probabilities in M/G/1/∞ and M/G/1/K, we have

$$\begin{aligned}
W &= \frac{L}{\lambda(1-P_B)} = \frac{1}{\lambda(1-P_B)}\left(\sum_{j=1}^{K} jP_j\right) \\
&= \frac{1}{\lambda(1-P_B)}\left(\sum_{j=1}^{K-1} jP_j + KP_K\right) \\
&= \frac{1}{\lambda(1-P_B)}\left\{\sum_{j=1}^{K-1} j\pi_j^{\infty}\left(\frac{P_0}{\pi_0^{\infty}}\right) + KP_K\right\} \\
&= \frac{1}{\lambda(1-P_B)}\left(\frac{P_0}{\pi_0^{\infty}}\right)\left(\sum_{j=1}^{K-1} j\pi_j^{\infty}\right) + \frac{K}{\lambda(1-P_B)}P_K.
\end{aligned}$$

Since $P_0 = 1-\rho(1-P_B)$, $\pi_0^{\infty} = 1-\rho$, and $1-E_K = \frac{(1-\rho)(1-P_B)}{1-\rho(1-P_B)}$ we have

$$\frac{P_0}{\pi_0^{\infty}}\frac{1}{(1-P_B)} = \frac{1-\rho(1-P_B)}{(1-\rho)(1-P_B)} = \frac{1}{1-E_K}.$$

Then, from Lemma 1 and $P_K = P_B$ we have

$$W = \frac{1}{\lambda}\left(\frac{1}{1-E_K}\right)\left(\sum_{j=1}^{K} E_j - KE_K\right) + \frac{K}{\lambda(1-P_B)}P_B.$$

The proof is completed by applying the following relation to the above equation:

$$\left(\frac{P_B}{1-P_B} - \frac{E_K}{1-E_K}\right) = -\frac{\rho E_K}{1-E_K} = \left(1 - \frac{1}{1-E_K}\right)\rho.$$ □

## 3. Minimal Buffer Capacity

Our closed-form blocking formulae given in Theorem 1 and Corollary 1 can be immediately applied to an optimal problem that determines the minimal buffer capacity $K^*$ that satisfies the given blocking probability $P_B^*$. Clearly, $E_k$ is monotonously decreasing in the finite buffer capacity $k$, since for any $k \ge 0$, the steady-state probability $\pi_k^{\infty}$ has a positive value and $E_k - E_{k+1} = \sum_{j=k}^{\infty} \pi_j^{\infty} - \sum_{j=k+1}^{\infty} \pi_j^{\infty} = \pi_k^{\infty} \ge 0$. Thus, we can numerically select the minimal buffer capacity. This optimization problem can be formulated as follows:

$$\begin{aligned}
K^* &= \min\left\{k \in (1,2,\ldots) : P_B^* \ge \frac{(1-\rho)E_k}{1-\rho E_k}\right\} \\
&= \min\left\{k \in (1,2,\ldots) : E_k \le \frac{P_B^*}{1-\rho(1-P_B^*)}\right\},
\end{aligned}$$

where $E_k$ is the same as in (13) and $\rho = \lambda\sigma < 1$.

## V. Concluding Remarks

Unlike the usual queueing theory, in this study we used max-plus algebra to provide explicit formulae for a blocking probability, stationary distributions, and mean system sojourn times in M/D/1/K queues under two blocking policies: communication and production. Whereas we had equivalent results in the literature for a communication blocking policy, blocking formulae for a production blocking policy are a new achievement. Due to the stability condition of the (Taylor) series expansion, we limited an offered load (traffic intensity) of $\rho < 1$. However, our expression is also valid when $\rho > 1$ because it is equivalent to the formulae given in [4] and [5]. Moreover, we believe that our approach is applicable to finite-capacity multi-node tandem queues with properly chosen overall offered load $\rho$ if the explicit expressions of the random vector $D_n$ are available.

It will be necessary to conduct further inquiries into more efficient computational algorithms when the capacity $K$ and the offered load $\rho$ are large. Additionally, our explicit expressions are also useful in obtaining better approximations for queues with general service times, since various existing approximation methods form the weighted combination of deterministic and exponential queues (see, for example, Smith [8] and Tijms [3]).

## References


[1] D. Gross and C.M. Harris, *Fundamentals of Queueing Theory*, New York, NY: John Wiley & Sons, 1998, pp. 74–81.

[2] H. Takagi, *Queueing Analysis: Vol. II: Finite Systems*, Amsterdam, Netherlands: North-Holland, 1993, pp. 197–222.

[3] H.C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*, Chichester, England: John Wiley & Sons, 1986, pp. 357–364.

[4] O. Brun and J.M. Garcia, "Analytical Solution of Finite Capacity M/D/1 Queues," *J. Appl. Probability*, vol. 37, no. 4, Dec. 2000, pp. 1092–1098.

[5] S. Alouf, P. Nain, and D. Towsley, "Inferring Network Characteristics via Moment-Based Estimators," *IEEE INFOCOM*, Anchorage, AK, USA, Apr. 22–26, 2001, pp. 1045–1054.

[6] J.W. Cohen, *The Single Server Queue*. Revised ed., Amsterdam, Netherlands: North-Holland, 1982, pp. 551–582.

[7] H.G. Perros, *Queueing Networks with Blocking*, New York, NY: Oxford University Press, 1994, pp. 91–126.

[8] J.M. Smith, "M/G/c/K Blocking Probability Models and System Performance," *Performance Evaluation*, vol. 52, no. 4, May 2003, pp. 237–267.

[9] H. Sakasegawa, M. Miyazawa, and G. Yamazaki, "Evaluation of the Overflow Probability Using the Infinite Queue," *Manag. Sci.*, vol. 39, no. 10, Oct. 1993, pp. 1238–1245.

[10] F. Baccelli et al., *Synchronization and Linearity: An Algebra for Discrete Event Systems*, Chichester, England, UK: John Wiley & Sons, 1992, pp. 63–88.

[11] B. Heidergott, *Max-Plus Linear Stochastic Systems and Perturbation Analysis*, New York, NY: Springer, 2007, pp. 10–20.

[12] F. Baccelli and V. Schmidt, "Taylor Series Expansions for Poisson Driven (Max,+)-Linear Systems," *Ann. Appl. Probability*, vol. 6, no. 1, Feb. 1996, pp. 138–185.

[13] F. Baccelli, S. Hasenfuss, and V. Schmidt, "Transient and Stationary Waiting Times in (Max, +)-Linear Systems with Poisson Input," *Queueing Syst.*, vol. 26, no. 3–4 Nov. 1997, pp. 301–342.

[14] F. Baccelli, S. Hasenfuss, and V. Schmidt, "Expansions for Steady-State-Characteristics of (Max, +) Linear Systems," *Stochastic, Models*, vol. 14, no. 1–2, 1998, pp. 1–24.

[15] H. Ayhan and D.-W. Seo, "Laplace Transform and Moments of Waiting Times in Poisson Driven (Max, +) Linear Systems," *Queueing Syst.*, vol. 37, no. 4, Mar.–Apr. 2001, pp. 405–438.

[16] H. Ayhan and D.-W. Seo, "Tail Probability of Transient and Stationary Waiting Times in (Max, +)-Linear Systems," *IEEE Trans. Autom. Contr.*, vol. 47, no. 1, Jan. 2002, pp. 151–157.

[17] B. Heidergott, G. J. Olsder, and J. van der Woude, *Max Plus at Work*, Princeton, NJ: Princeton University Press, 2006, pp. 13–24.

[18] D.-W. Seo, "Application of (Max, +)-Algebra to the Waiting Times in Deterministic 2-Node Tandem Queues with Blocking," *J. KORMS Society*, vol. 30, no. 1, Mar. 2005, pp. 149–159.

[19] D.-W. Seo, J. Lee, and B.-Y. Chang, "Approximation Method for Blocking Probabilities in $M/D/1/K_1 \rightarrow \cdot/D/1/K_2$ Queues," Preprint, submitted Mar. 2013.



**Dong-Won Seo** is an associate professor at the School of Management, Kyung Hee University, Seoul, Rep. of Korea. He received his PhD degree from the School of Industrial and Systems Engineering at Georgia Institute of Technology, Atlanta, USA. His current research interests are stochastic processes, series expansions, max-plus algebra, performance evaluation, and applications of management science and simulation.