International Journal of Management Science
Vol 14, No 1, May 2008

Stationary Waiting Times in m-node Tandem Queues with Communication Blocking[*]

Dong-Won Seo
College of Management and International Relations, Kyung Hee University
1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 449-701, Korea

Hochang Lee[**]
College of Management and International Relations, Kyung Hee University
1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 449-701, Korea

Sung-Seok Ko

Department of Industrial Engineering, Konkuk University
1 Hwayang-dong, Gwangjin-gu, Seoul, 143-701, Korea

(Received: April 17, 2007 / Revised: October 23, 2007 / Accepted October 27, 2007)

ABSTRACT

In this study, we consider stationary waiting times in a Poisson driven single-server m-node queues in series. We assume that service times at nodes are independent, and are either deterministic or non-overlapped. Each node excluding the first node has a finite waiting line and every node is operated under a FIFO service discipline and a communication blocking policy (blocking before service). By applying (max, +)-algebra to a corresponding stochastic event graph, a special case of timed Petri nets, we derive the explicit expressions for stationary waiting times at all areas, which are functions of finite buffer capacities. These expressions allow us to compute the performance measures of interest such as mean, higher moments, or tail probability of waiting time. Moreover, as applications of these results, we introduce optimization problems which determine either the biggest arrival rate or the smallest buffer capacities satisfying probabilistic constraints on waiting times. These results can be also applied to bounds of waiting times in more general systems. Numerical examples are also provided.

Keyword: (Max, +)-Algebra, Tandem Queues, Finite Buffers, Waiting Times, Running Head, Stationary Waiting Times in m-node Tandem Queues with Communication Blocking

---

[*] This research was supported by the Kyung Hee University Research Fund in 2006 (KHU-20060563).

[**] Corresponding author, E- mail: hochang@khu.ac.kr

1. Introduction

As a common model of telecommunication networks and manufacturing systems, tandem queues with infinite or finite buffers have been widely studied. Because of the computational complexity and difficulty in the analysis of performance evaluations for stochastic networks, most studies are focused on very restrictive and/or small size of stochastic networks over the past decades such as exponential service times, infinite buffers, finite buffers of size 1 or 2, and so on. Especially, they usually assumed infinite buffers at each node when they analyze the stationary waiting times at a node in a finite-buffer stochastic system. In our best knowledge, there is no result on stationary waiting times in sub-areas of finite-buffer tandem queues.

Recently, more generous system which so called a (max, +)-linear system has been studied. Various types of stochastic networks which are prevalent in telecommunication, manufacturing systems belong to the (max, +)-linear system. Many instances of (max, +)-linear systems can be represented by stochastic event graphs, a special type of stochastic Petri net, which allow one to analyze them by (max, +)-algebra, involving only two operators: 'max' and '+'. To be short, (max, +)-linear system is a choice-free net and consists of single-server queues under a FIFO (First-In First-Out) service discipline.

Baccelli and Schmidt [5] derived a Taylor series expansion for mean stationary waiting time with respect to the arrival rate in a Poisson driven (max, +)-linear system. Their approach was generalized to other characteristics of stationary and transient waiting times by Baccelli *et al.* [4], Ayhan and Seo [1, 2]. Recently, Seo [6, 7], by the same way, derived explicit expressions for stationary waiting times in all areas of deterministic 2-node and 3-node tandem queues with finite buffers under two blocking policies: communication and manufacturing. In Seo and Song [8], they considered an optimization problem which determines the size of the finite buffers for 2-node tandem queues with blocking.

Their methods are still valid for more complex (max, +)-linear systems. Thus, the goal of this study is to extend their study to m-node tandem queues with blocking. In either deterministic or non-overlapping m-node tandem queues with a communication blocking explicit expressions on stationary waiting times at all nodes are derived. They are functions of finite buffer capacities, and immediately applicable to the closed form expression for characteristics of stationary waiting times in Theorem 1 of

[6] (see also [1]), and to optimization problems for determining the largest arrival rate or the smallest buffer capacities which satisfies predetermined probabilistic constraints on stationary waiting times.

Reader can refer on basic (max, +)-algebra and some preliminaries on waiting times in (max, +)-linear systems to Baccelli *et al.* [3] (see also [1, 2, 6]). This paper is organized as follows. Section 2 includes our main results. Applications of our results are introduced in Section 3 and 4. Section 5 shows some numerical examples and concluding remarks are mentioned in Section 6.

2. Stationary Waiting Times in m-node Finite Queues in Series

In this section we introduce our main results on stationary waiting times in a Poisson driven single-server m-node tandem queues with finite buffers. We assume that service times at all nodes are distributed as either deterministic (constant) or non-overlapped, and only a FIFO service discipline is allowed. We also assume that the buffer at each node is finite except the one at the first node, and services at each node are given by a communication blocking policy.

We first analyze an m-node queue in series with constant service times and derive the explicit expressions for D_n^i, the i-th component of random vector D_n (see [1, 2, 6]), on stationary waiting times at all areas of the stochastic system. Note that the D_n^i can be interpreted as a critical path (the longest path of n-th arrival from the initial node to node i) in the corresponding task graph and written in terms of service times with only two operators 'max' and '+'. It also means that the polynomials $p_k(\cdot)$ and $q_k(\cdot)$ given in [1] (see also [1, 2, 6]) can be calculated independently of the arrival rate of a Poisson process. Then, from those explicit expressions we can intuitively inference the explicit expressions on stationary waiting times for a system with non-overlapping service times.

For node i, $i=1, 2, \cdots, m$, let σ^i be a deterministic service time, and let K_i be a buffer capacity at node , which includes a room for a customer in service. First of all, it is worth mentioning about the stationary waiting times in m-node queues in series with infinite buffers at all nodes. From the definition of random vector D_n with some algebra, the expressions for the i-th components of D_n can be obtained as the following expressions.

$$D_n^i = \sum_{j=1}^{i-1} \sigma^j + \max\{n\sigma^1, n\sigma^2, \ldots, n\sigma^i\} \text{ for } n \geq 0. \quad (1)$$

For a finite buffer tandem queue we only consider a communication blocking policy (blocking before service) in this study. Under a communication blocking rule, a customer at node j cannot begin his service unless there is a vacant space in the buffer at node $j+1$. One can obtain recursive equations for stationary waiting times in (max, +)-linear systems with finite buffers by using the same way as done in [1, 2], which draws a corresponding event graph and then converts it to an event graph with infinite buffers by inserting dummy nodes with zero service times. For example, the following Figure 1 shows the event graph of a 3-node tandem queue with finite buffers of size 3 at node 2 and 3 while Figure 2 shows the event graph of the one with infinite buffers at all nodes by inserting dummy nodes having zero service times.

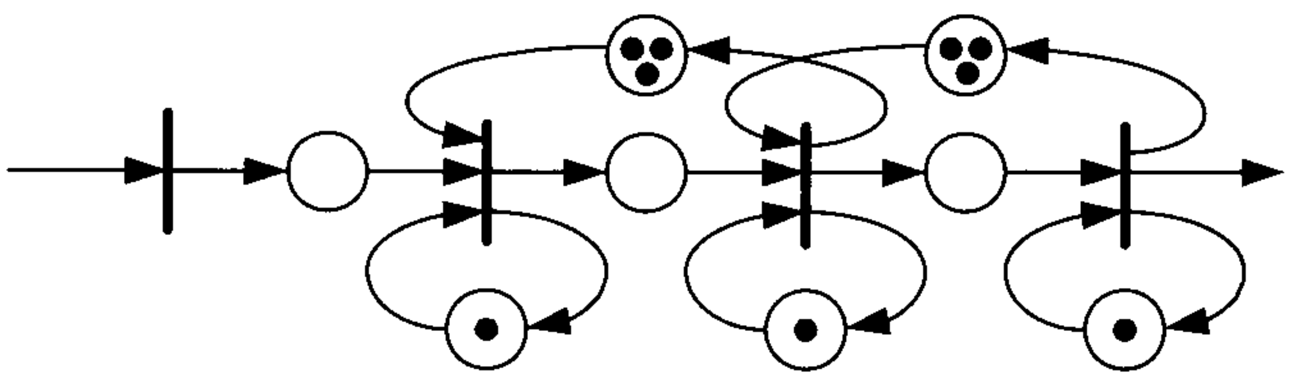

Figure 1. 3-node tandem queues with finite buffers of size 3 at node 2 and 3

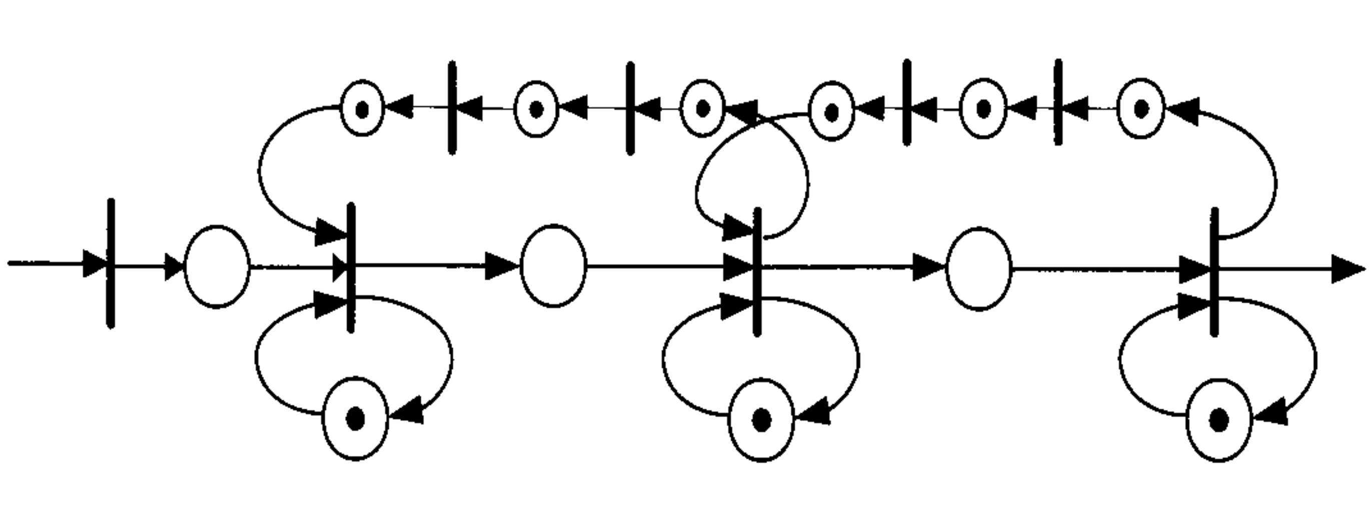

Figure 2. 3-node tandem queues with infinite buffers and dummy nodes

Remark 1: In the previous studies [6-8], they mentioned that the explicit expressions of D_n^i under communication blocking policy can be written in terms of a common formula when $K_i \geq 3$ for all i, but they recently figured out that these expressions are also valid when $K_i \geq 2$ for all i.

In this study, we only consider the cases for $K_i \geq 2,\ i = 2,\cdots,m$. Because this case covers more general systems and for other case omitted here one can easily obtain explicit expressions using the same way, we here derive the explicit expressions of D_n^i, the i-th components of the random vector D_n^i. By the similar way as done in infinite-buffer systems, one is able to obtain the following Proposition for the explicit expressions of D_n^i as functions of the sizes of finite buffers $K_j,\ j = i+1,\cdots,m$.

Proposition 1: For a deterministic m-node tandem queue under Communication Blocking when $K_j \geq 2,\ j = 2,\cdots,m$, and $K_1 = K_{m+1} = \infty$, for a node $i,\ i = 1,\cdots,m$,

$$D_n^i = \sum_{j=1}^{i-1} \sigma^j + \max\{n\sigma^1, n\sigma^2, \cdots, n\sigma^i\} \quad \text{for} \quad 0 \leq n < K_{i+1} \tag{2}$$

$$D_n^i = \sum_{j=1}^{i-1} \sigma^j + \max\{n\sigma^1, \cdots, n\sigma^i, l_i(i+1), \cdots, l_i(\kappa_i)\} \quad \text{for} \quad \sum_{j=i+1}^{\kappa_i} K_j \leq n < \sum_{j=i+1}^{\kappa_i+1} K_j \tag{3}$$

where $l_i(p)$ in (3) is defined as $l_i(p) = \sigma^i + 2\sum_{j=i+1}^{p-1} \sigma^j + \left[n - \left(\sum_{j=i+1}^{p} K_j \right) + 1 \right] \sigma^p$, κ_i is an integer value such that $\kappa_i = \min\left\{ q \in (i+1,\cdots,m) : \sum_{j=i+1}^{q} K_j \leq n < \sum_{j=i+1}^{q+1} K_j \right\}$, and with the convention that summation over an empty set is 0.

The following Remark confirms that Proposition 1 shows the same results as the previous study of Seo [7] for a Poisson driven finite 3-node tandem queue.

Remark 2: By letting $m = 3$ in Proposition 1, one can obtain the same expression for the random vector D_n^i as those in [7], in which he derive the explicit expression of D_n^i in a 3-node finite tandem queue with deterministic service times. That is, under communication blocking policy, when $K_2 \geq 2$ and $K_3 \geq 2$,

i) for $i = 1$,

$D_n^1 = n\sigma^1$ for $0 \leq n < K_2$,

$D_n^1 = \max\{n\sigma^1,\ \sigma^1 + (n - K_2 + 1)\sigma^2\}$ for $K_2 \leq n < K_2 + K_3$,

$$D_n^1 = \max\{n\sigma^1, \sigma^1 + (n - K_2 + 1)\sigma^2, \sigma^1 + 2\sigma^2 + [n - (K_2 + K_3) + 1]\sigma^3\}$$
$$\text{for } n \geq K_2 + K_3,$$

ii) for $i = 2$

$$D_n^2 = \sigma^1 + \max\{n\sigma^1, n\sigma^2\} \text{ for } 0 \leq n < K_3,$$
$$D_n^2 = \sigma^1 + \max\{n\sigma^1, n\sigma^2, \sigma^2 + (n - K_3 + 1)\sigma^3\} \text{ for } n \geq K_3,$$

iii) for $i = 3$

$$D_n^3 = \sigma^1 + \sigma^2 + \max\{n\sigma^1, n\sigma^2, n\sigma^3\} \text{ for } n \geq 0.$$

On the other hand, because every service time at a node can be different in a system with non-overlapping service times, let σ_{-n}^i be an i.i.d. random variable of the n-th service time at node i in the sense of Palm probability on the negative half line. From the basic probability theory, we can infer the expressions for a non-overlapping system by substituting a (simple) sum of degenerate random variables in Proposition 1 with a sum of random variables. That is, Proposition 1 can be written as follows.

Proposition 2: For a non-overlapping m-node tandem queue under Communication Blocking, when $K_j \geq 2$, $j = 2, \cdots, m$, and $K_1 = K_{m+1} = \infty$, for a node i, $i = 1, \cdots, m$,

$$D_n^i = \sum_{j=1}^{i-1} \sigma_0^j + \max\{\sum_{\ell=1}^{n} \sigma_{-\ell}^1, \sum_{\ell=1}^{n} \sigma_{-\ell}^2, \cdots, \sum_{\ell=1}^{n} \sigma_{-\ell}^i\} \text{ for } 0 \leq n < K_{i+1} \tag{4}$$

$$D_n^i = \sum_{j=1}^{i-1} \sigma_{-0}^j + \max\{\sum_{\ell=1}^{n} \sigma_{-\ell}^1, \cdots, \sum_{\ell=1}^{n} \sigma_{-\ell}^i, \hat{l}_i(i+1), \cdots, \hat{l}_i(\kappa_i)\} \text{ for } \sum_{j=i+1}^{\kappa_i} K_j \leq n < \sum_{j=i+1}^{\kappa_i+1} K_j \tag{5}$$

where $\hat{l}_i(p)$ in (3) is defined as $\hat{l}_i(p) = \sigma_{-n}^i + \sum_{j=i+1}^{p-1} \sum_{\ell=1}^{2} \sigma_{-(n+1-\ell)}^j + \sum_{\ell=1}^{\left[n - \left(\sum_{j=i+1}^{p} K_j\right) + 1\right]} \sigma_{-(n+1-\ell)}^p$, κ_i is the same as the one in Proposition 1, and with the convention that summation over an empty set is 0.

From the above Propositions, one is able to disclose the fact that deterministic m-node tandem queues with both infinite and finite buffers under communication blocking mechanism have same expressions of D_n^m for all $n \geq 0$. And, this is also true for those with non-overlapping service times (see equations (1), (2), and (4)).

These show the same results, by the totally different way, as those in Wan and Wolff [10] that when the first node's buffer capacity is infinite, a customer's sojourn time is not dependent of the finite buffer capacities and the order of nodes (see also Whitt [11]).

Corollary 3.1 in Ayhan and Seo [1], Theorem 1 in Baccelli *et al.* [4] and Theorem 2.3 in Ayhan and Seo [2] together with the explicit expressions allow one to compute characteristics of stationary waiting times in such stochastic systems with either deterministic or non-overlapping service times.

3. Application to Optimization Problems

Our results can be applicable to an optimization problem which determines the largest Poisson arrival rate or the smallest finite buffers capacities subject to probabilistic constraints on stationary waiting times in m-node tandem queues with finite buffers and either constant or non-overlapping service times.

For a node i, $i=1,\cdots,m$, let $\tau_i \geq 0$ be a pre-specified bound on stationary waiting time W^i, the elapsed time from the arrival until the beginning of its service at node i, and let $0<\beta_i<1$ be a pre-specified probability value, like a QoS (Quality of Service).

With the fact that the tail probability of stationary waiting time in an M/G/1 queue is convex with respect to an arrival rate λ (see [2]), and the explicit expressions of D_n^i derived in the previous section together with Theorem 2.3 in [2], the optimal arrival rate can be numerically computed as a solution of the following optimization problem. For a fixed i, $i=1,\cdots,m-1$, and given finite buffer sizes K_j s, $j=2,\cdots,m$,

$$\begin{aligned} \max \quad & \lambda \\ s.t. \quad & \Pr(W^i > \tau_i) \leq \beta_i \\ & \lambda \in [0, a_i^{-1}) \end{aligned}$$

where a_i is given in the i-th component of D_n vector defined in (2.5) of [1] (see also [2, 6]).

From the fact that the stationary waiting time at node m is independent of the sizes of finite buffers, an optimization problem determining the smallest finite buffer capacities can consider only sub-areas of a system. From the definition of W^i and D^i_n (see equations (2.2) and (2.3) in [1]), because for a fixed n, each argument of D^i_n is monotonously decreasing in K_j, $j = i+1,\cdots,m$, and 'max' function is a convex so that D^i_n is a decreasing convex for all $n \geq 0$. Consequently, W^i, the elapsed time from the arrival until the beginning of its service at node i, is a decreasing convex in K_j, $j \geq i+1$, because the composition of convex functions is also a convex function (see [9]). So, because W^i is stochastically non-increasing in K_j, $j = i+1,\cdots,m$, the optimal values of finite buffers can be numerically chosen. The optimal values of finite buffers can be computed as a solution of the following optimization problem.

$$\begin{aligned} \min \quad & \sum_{j=2}^{m} K_j \\ s.t. \quad & \Pr(W^i > \tau_i) \leq \beta_i \quad \textit{for } i = 1,\cdots,m-1 \\ & K_j \geq 2, j = 2,\cdots,m \end{aligned}$$

where a_i is given in the i-th component of D_n vector defined in (2.5) of [1].

As mentioned earlier, since the expressions of D^i_n are functions of K_j, $j = i+1,$ $\cdots,m$, this optimization problem can be solved in reverse order of node i, one by one from $i = m-1$ to $i = 1$. In other words, one can first choose the optimal value of K^*_m (when $i = m-1$), then choose the optimal value of $K_{m-1} + K^*_m$ (when $i = m-2$) by using this chosen value K^*_m, and so on.

4. Application to tandem queues with more General Service Times

In this section, we introduce a stochastic ordering result which is useful to obtain bounds for stationary waiting times in m-node tandem queues with more general service times. Shaked and Shanthikumar [9] showed the following well-known stochastic ordering result. For a random variable X with finite support, and letting l_X be the left endpoint and u_X be the right endpoint, then the following stochastic ordering holds.

$$E(X) \leq_{cx} X \leq_{cx} Z$$

where $E(X)$ is the expected value of X, $\leq_{cx}$ means stochastic convex ordering. A random variable Z is defined as

$$\Pr(Z = l_X) = \frac{u_X - E(X)}{u_X - l_X}, \ \Pr(Z = u_X) = \frac{E(X) - l_X}{u_X - l_X}.$$

From the definition of D_n^i one can easily see the convexities of waiting time W^i and tail probability $\Pr(W^i > \tau_i)$ since they are functions of service times involving only two operators '+' and 'max'. Together with our results, therefore, the stochastic convex ordering result allows one to compute lower and upper bounds for characteristic of stationary waiting times in finite-buffer m-node tandem queues with more general service times of finite supports. This finite support assumption is not unreasonable in real systems.

In order to illustrate our results, numerical examples are provided in next section.

5. Numerical Examples

Even though our method is valid for both deterministic and non-overlapping service times, to avoid computational complexity and difficulty we consider here only a deterministic tandem queues under communication blocking.

For instance, we consider a 5-node tandem queue with constant service times. No blocking occur at all downstream nodes of a node with the biggest service time. Without loss of generality, we assume an increasing order of service times. Let $\sigma^i = 0.1 \times i$ be a constant service time at node i, $i = 1, \cdots, 5$. In this example, the maximum of service times (Lyapunov maximum value, see [3]) a is 0.5, and when $K_j = 5$ for $j = 2, \cdots, 5$ the values of ξ_i, $i = 1, \cdots, 5$, defined in the structure D_n^i of (2.5) of [1] (see also [2]) are given as follows: $\xi_1 = 31$, $\xi_2 = 27$, $\xi_3 = 23$, $\xi_4 = 16$, and $\xi_5 = 0$. Note that ξ_i is the number of terms needed in Corollary 3.1 of [1] to compute characteristics of stationary waiting times. From the explicit expressions for

D_n^i together with Corollary 3.1 one is able to compute the exact values of mean stationary waiting times $E(W^i)$ at node i. Following Table 1–Table 3 show exact values and simulation values for the mean stationary waiting times at each node when the traffic intensity is $\rho = 0.5$.

Table 1. Mean Stationary Waiting Times with $K_1 = \infty$, $K_j = 5$, $j = 2, \cdots, 5$.

Waiting Time	Exact Solution	Simulation
$E(W^1)$	0.00555587	0.00556 ∓ 0.00003098
$E(W^2)$	0.125003	0.12504 ∓ 0.00011165
$E(W^3)$	0.364311	0.36439 ∓ 0.00028483
$E(W^4)$	0.733498	0.73376 ∓ 0.00067334
$E(W^5)$	1.25	1.2503 ∓ 0.0013900

As mentioned earlier, since the expressions for D_n^i are function of K_j, $j = i+1, \cdots, m$, ξ_i is also varying with finite buffer capacities K_j. Table 2 has $\xi_1 = 29$, $\xi_2 = 23$, $\xi_3 = 17$, $\xi_4 = 11$, and $\xi_5 = 0$, and Table 3 has $\xi_1 = 39$, $\xi_2 = 34$, $\xi_3 = 28$, $\xi_4 = 21$, and $\xi_5 = 0$.

Table 2. Mean Stationary Waiting Times with $K_1 = \infty$, $K_2 = 7, K_3 = 6, K_4 = 5$, and $K_5 = 4$.

Waiting Time	Exact Solution	Simulation
$E(W^1)$	0.00555556	0.00554 ∓ 0.000048929
$E(W^2)$	0.125	0.12497 ∓ 0.00017797
$E(W^3)$	0.364337	0.36429 ∓ 0.00041966
$E(W^4)$	0.734402	0.73435 ∓ 0.00085446
$E(W^5)$	1.25	1.2499 ∓ 0.00162

Table 3. Mean Stationary Waiting Times with $K_1 = \infty$, $K_j = 6$, $j = 2, \cdots, 5$.

Waiting Time	Exact Solution	Simulation
$E(W^1)$	0.00555557	0.00556 ∓ 0.000030980
$E(W^2)$	0.125	0.12504 ∓ 0.00011189
$E(W^3)$	0.364288	0.36439 ∓ 0.00028489
$E(W^4)$	0.73336	0.73362 ∓ 0.00066158
$E(W^5)$	1.25	1.2503 ∓ 0.00139

From the Tables, we can see that our expressions of D_n^i are accurate and verify the fact mentioned earlier that stationary waiting times at the last node are independent of finite buffer sizes K_j, $j = 2,\cdots,m$ (see equations (1), (2), and (4)). Moreover, comparing Table 1 with Table 3 shows the decreasing property of the mean waiting times with respect to finite buffer capacities as we expect.

By using the explicit expressions of D_n^i together with the previous results we can numerically determine the largest value of a Poisson arrival rate, the smallest values of K_j, $j = 2,\cdots,m$, as well as the bounds for characteristics of waiting times, but we omit them here.

6. Concluding Remark

By using (max, +)-algebra, we derived the explicit expressions on stationary waiting times in Poisson driven finite-buffer tandem queues with either deterministic or non-overlapping service times as functions of finite buffer. While we only consider a communication blocking policy, the same method can be applied to a production blocking policy. We show that these expressions are applicable to optimization problems subject to probabilistic constraints on waiting times such as QoS, and that they are useful to obtain bounds for waiting times in a finite-buffer system with more general service times.

In the future, these results can be extended to more general complex (max, +)-linear systems with finite buffers such as fork-and-join type queues (a special case of tandem queues), (maybe) mixture of several blocking policies, and so forth. It is also interesting to apply these methods to analysis of stochastic networks with multiple servers and of base-stock inventory models. It is necessary, however, to develop more efficient computational algorithms in order to overcome the computational complexity and difficulty.

References

[1] Ayhan, H. and D-W. Seo, "Laplace Transform and Moments of Waiting Times

in Poisson Driven (Max, +)-Linear Systems," *Queueing Systems* 37, 4 (2001), 405-438.

[2] Ayhan, H. and D-W. Seo, "Tail Probability of Transient and Stationary Waiting Times in (Max, +)-Linear Systems," *IEEE Transactions on Automatic Control* 47, 1 (2002), 151-157.

[3] Baccelli, F., G. Cohen, G. J. Olsder, and J-P. Quadrat, Synchronization and Linearity: An Algebra for Discrete Event Systems, John Wiley and Sons 1992.

[4] Baccelli, F., S. Hasenfuss, and V. Schmidt, "Expansions for Steady State Characteristics in (Max, +) Linear Systems," *Stochastic Models* 14 (1998), 1-24.

[5] Baccelli, F. and V. Schmidt, "Taylor Series Expansions for Poisson Driven (Max, +) Linear Systems," *Annals of Applied Probability* 6, 1 (1996), 138-185.

[6] Seo, D.-W., "Application of (Max, +)-algebra to the Waiting Times in Deterministic 2-node Tandem Queues with Blocking," *J. KORMS* 30, 1 (2005), 149-159.

[7] Seo, D.-W., "Application of (Max, +)-algebra to the Waiting Times in Deterministic 3-node Tandem Queues with Blocking," *J. KORMS* 30, 2 (2005), 73-80.

[8] Seo, D.-W. and B.-K. Song, "Application of (Max, +)-algebra to Optimal Buffer Size in Deterministic Queues in Series with Blocking," LNAI 3642 (2005), 671-677.

[9] Shaked M. and J. G. Shanthikumar, Stochastic Orders and Their Applications, Academic Press, 1994.

[10] Wan, Y.-W. and R. W. Wolff, "Bounds for Different Arrangements of Tandem Queues with Nonoverlapping Service Times," *Management Science* 39, 9 (1993), 1173-1178.

[11] Whitt, W., "The Best Order for Queues in Series," *Management Science* 31, 4 (1985), 475-487.