2005 한국경영과학회/대한산업공학회 춘계공동학술대회
2005 5 월 13 일~14 일, 충북대학교

3-盧드 유한 버퍼 일렬대기행렬에서의 최적 버퍼 크기에 대한 분석

Analysis of Optimal Buffer Capacities in 3-node Tandem Queues with Blocking

서동원

경희대학교 국제경영학부
dwseo@khu.ac.kr

Abstract

In this study, we consider characteristics of waiting times in single-server 3-node tandem queues with a Poisson arrival process, finite buffers and deterministic or non-overlapping service times at each queue. There are three buffers: one at the first node is infinite and the others are finite.

The explicit expressions of waiting times in all areas of the systems, which are driven as functions of finite buffer capacities, show that the sojourn time does not depend on the finite buffer capacities and also allow one to compute and compare characteristics of waiting times at all areas of the system under two blocking policies: communication and manufacturing blocking. As an application of these results, moreover, an optimization problem which determines the smallest buffer capacities satisfying predetermined probabilistic constraints on waiting times is considered. Some numerical examples are also provided.

Key Words: Tandem Queues, Finite Buffers, Waiting Times, (Max,+)-Linear Systems

1. Equation Section 1Introduction

As common models of telecommunication networks and manufacturing systems, finite buffered tandem queues have been widely studied. Many researchers have interests in characteristics in stochastic networks such as mean waiting times, system sojourn times, invariant probabilities, blocking probabilities, and so on. Since the computational complexity and difficulty in the analysis of performance evaluations for stochastic networks, most studies are focused on restrictive and/or small size of stochastic networks over the past decades.

In Poisson driven 2-node tandem queues with exponential service times, Grassmann and Drekic [7] studied the joint distribution of both queues by using generalized eigenvalues. For infinite buffered queues in series with non-overlapping service times and an arbitrary arrival process, Whitt [12] studied the optimal order of nodes which minimizes the mean value of sojourn times. Nakade [9] derived bounds for the mean value of cycle times in tandem queues with general service times under communication and manufacturing blocking policies.

Under the assumption of the capacities of buffers including the space for a customer in service and an arbitrary arrival process, Wan and Wolff [11] showed that the departure processes in tandem queues with finite buffers except for the first node and non-overlapping service times are independent of the size of finite buffers when it is greater than 2 under communication blocking or when it is greater than 1 under manufacturing blocking. Labetoulle and Pujolle [8] gave the same results for mean response time, but derived mean waiting times at each queue under the assumption of infinite buffer capacity for all nodes. However, in our best knowledge, there is no result on waiting times in all sub-areas of finite buffered queues in series.

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. (Max, +)-linear system is a choice-free net and consists of single-server queues under FIFO(First-In First-Out) service discipline. Discrete event systems (DESs) can be properly modeled by (max, +)-algebra, involving only two operators: 'max' and '+'.

Recently, Baccelli and Schmidt [6] 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, 5], Ayhan and Seo [1, 2].

By the similar way, Seo [10] derived explicit expressions for characteristics of stationary waiting times in all areas of deterministic 2-node tandem queues with a finite buffer under two blocking policies: communication and manufacturing. He also disclosed a relationship of stationary waiting times in all areas of the systems between the blocking policies.

The methods used in Seo [10] are still valid for more complex (max, +)-linear systems. Thus, the goal of this study is to extend his study to 3-node tandem queues which have three buffers: one at the first node is infinite and the others are finite. The explicit expressions for characteristics of stationary waiting times at all nodes in deterministic or non-overlapping 3-node tandem queues with finite buffers are derived. These expressions are functions of finite buffer capacities, immediately applicable forms to compute the characteristics of stationary waiting times, and allow one to disclose a relationship of waiting times between the two blocking policies. Moreover, we consider an optimization problem for determining the smallest buffer capacities which satisfy 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], Ayhan and Seo [1, 2] and Seo [10]. This paper is organized as follows. Section 2 contains our main results. An optimization problem and numerical examples are given in section 3 and 4. Conclusion and some future research topics are mentioned in Section 5.

2. Equation Section (Next)Waiting Times in 3-node Tandem Queues

This study investigates on waiting times in single-server 3-node tandem queues with finite buffers, a Poisson arrival process and deterministic or non-overlapping service times. The system has three buffers: one at the first node is infinite and the others are finite.

Let s^i and K_i be the deterministic or non-overlapping service time and the capacity of buffer at node i ($i = 1,2,3$). The buffer capacities include a room for a customer in service. We first mention about the waiting times in 3-node tandem queues with infinite buffers at all nodes. From the definition of random vector D_n, one can obtain the expressions of the components of D_n as

$$D_n^1 = n s^1 \quad for\ n \quad 0,$$

$$D_n^2 = s^1 + n \max\{s^1, s^2\} \quad for\ n \quad 0,$$

$$D_n^3 = s^1 + s^2 + n \max\{s^1, s^2, s^3\} \quad for\ n \quad 0 \qquad (1)$$

For 3-node tandem queues with finite buffers, we consider waiting times under two blocking policies: communication and manufacturing. Under communication blocking a customer at node j cannot begin his service unless there is a vacant space in the buffer at node $j+1$. For manufacturing 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 stays in node j until a vacancy is available. During that time, node j is blocked from serving other customers.

One can obtain recursive equations for

stationary waiting times in (max, +)-linear systems with finite buffers by using the same way as done in Seo [10], 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.

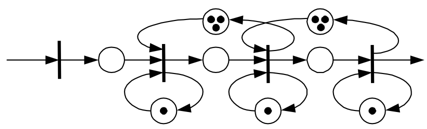

Figure 1: 3-node tandem queue with finite buffers of size 3 at node 2 and 3

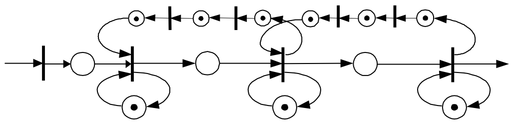

Figure 2: 3-node tandem queue with infinite buffers and dummy nodes

For the finite buffered system, we assume that we are interested in the cases: $3 ?\ K_2, K_3$ and $2 ?\ K_2, K_3$ for communication blocking and for manufacturing blocking, respectively. Since other cases the explicit expressions for the components of D_n can be easily obtained, we omit them here. Similarly as done in the infinite buffer case, one is able to obtain the following Propositions for the explicit expressions of random vector D_n, functions of finite buffer capacities K_2 and K_3. Theorem 1 and (2.4) in [10] together with the explicit expressions allow one to compute characteristics of waiting times in deterministic or non-overlapping systems.

Proposition 1: Under communication blocking,

when K_2 납$3, K_3 \quad 3$,

$$D_n^1 = n\,s^1 \quad for\ 0\ ?\ n \quad K_2,$$
$$D_n^1 = \max\{n\,s^1, s^1 + (n - K_2 + 1)s^2\} \quad for\ K_2\ ?\ n \quad K_2 + K_3,$$
$$D_n^1 = \max\{n\,s^1, s^1 + (n - K_2 + 1)s^2, s^1 + 2s^2 + [n - (K_2 + K_3) + 1]s^3\}, \quad for\ n\ ?\ K_2 \quad K_3$$

$$D_n^2 = s^1 + \max\{n\,s^1, n\,s^2\} \quad for\ 0\ ?\ n \quad K_3, \qquad (2)$$

$$D_n^2 = s^1 + \max\{n\,s^1, n\,s^2, s^2 + (n - K_3 + 1)s^3\} \quad for\ n\ ^3\ K_3, \qquad (3)$$

$$D_n^3 = s^1 + s^2 + n\max\{s^1, s^2, s^3\} \quad for\ n \quad 0. \qquad (4)$$

Proposition 2: Under manufacturing blocking,

when K_2 납$2, K_3 \quad 2$,

$$D_n^1 = n\,s^1 \quad for\ 0\ \#\ n \quad K_2,$$
$$D_n^1 = \max\{n\,s^1, s^1 + (n - K_2)s^2\} \quad for\ K_2 < n\ ?\ K_2 \quad K_3,$$
$$D_n^1 = \max\{n\,s^1, s^1 + (n - K_2)s^2, s^1 + s^2 + [n - (K_2 + K_3)]s^3\}, \quad for\ n > K_2 + K_3$$

$$D_n^2 = s^1 + \max\{n\,s^1, n\,s^2\} \quad for\ 0\ \#\ n \quad K_3, \qquad (5)$$

$$D_n^2 = s^1 + \max\{n\,s^1, n\,s^2, s^2 + (n - K_3)s^3\} \quad for\ n > K_3, \qquad (6)$$

$$D_n^3 = s^1 + s^2 + n\max\{s^1, s^2, s^3\} \quad for\ n \quad 0. \qquad (7)$$

From the above expressions, one is able to disclose three facts. One is that the expressions of D_n^3 for all $n\ ^3\ 0$ in tandem queues with finite buffer capacities and deterministic or non-overlapping service times under both blocking mechanisms are the same as those in the systems with infinite buffer capacities(see (1), (4) and (7)). It shows the same result in Wan and Wolff [11] 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 [12]).

The second one is that waiting times at node 2 in systems with manufacturing blocking have the same expressions in systems with communication blocking, except for one difference in the value of the finite buffer capacity at node 3. In other words, by substituting K_3 in (2) and (3) for $K_3 + 1$ can derive the same expressions for waiting times at node 2 as (5) and (6).

The third one is that waiting times in all subareas under manufacturing blocking are always smaller than or equal to those under communication blocking in systems with equal buffer capacities since all components of D_n under each blocking policy are nonincreasing in K_2 and K_3, and thus W^i is also stochastically nonincreasing(see (2.2) and (2.3) in [10]). Therefore, we can immediately conclude the following Theorem.

Theorem 1: In a Poisson driven deterministic or non-overlapping 3-node tandem queue with D_n^i satisfying the structure given in (2.5) of [10], when $K_1 =$ and $3 \# K_i$ $(i = 2, 3)$, then

$$E[G(W^i_{Communication\ Blocking\ with\ K_2, K_3+1})] = E[G(W^i_{Manufacturing\ Blocking\ with\ K_2, K_3})] \quad for\ i = 2, 3$$

and

$$E[G(W^i_{Communication\ Blocking\ with\ K_2, K_3})] \ ?\ E[G(W^i_{Manufacturing\ Blocking\ with\ K_2, K_3})] \quad for\ i \quad 1, 2, 3$$

where $G(\cdot)$ is an integrable, nonnegative, and differentiable function defined in (2.4) of [10].

3. Optimal Buffer Capacities

As an application of our results, we consider an optimization problem which minimizing the capacities of finite buffers subject to probabilistic constraints on stationary waiting times in (max, +)-linear systems where D_n^i sequence has the structure given in (2.5) of [10]. This probabilistic constraint, keeping waiting times within an acceptable range with a pre-specified probability, will ensure predictable completion times.

For node i, $i = 1, 2$, let t_i ³ 0 be a pre-specified bound on stationary waiting time W^i and let $0 < b_i < 1$ be a pre-specified probability value. Since W^i is stochastically nonincreasing in K_2 and K_3, one can numerically determine the smallest values of finite buffers K_2 and K_3 by using the explicit expressions of D_n^i together with Theorem 2.3 in [2]. The optimal buffer capacities can be computed as a solution of the following optimization problem, for a given $\lambda \in [0, a^{-1})$

$$\begin{aligned} \min \quad & K_2 + K_3 \\ s.t. \quad & P(W^i > t_i) \le b_i, \ for\ i \ \ 1,2 \\ & K_2, K_3 \in \infty \end{aligned}$$

where a_i is given in the i-th component of random vector D_m defined in (2.5) of [10].

Since the expressions for D_n^2 is a function of K_3 and D_n^1 is a function of K_2 with a fixed value of K_3, the optimization problem can be solved in two steps: one is for K_3 and then one for K_2 under the optimal value of K_3.

4. Examples

Even though our methods are still valid for non-overlapping service times, to avoid computation complexity we consider a finite buffered 3-node tandem queue with deterministic service times in this section.

Let $s^1 = 1, s^2 = 3, s^3 = 5$ be the constant service times at each node and $K_2 = 5, K_3 = 5$. In this particular example, the maximum of service times (Lyapunov maximum value) a is 5 and we assume that we are only interested in the mean value of waiting time W^i, the elapsed time from the arrival until the beginning of service at node i. From the explicit expressions of the random vector D_n^i together with Theorem 1 in [10] we are able to compute the exact value of mean waiting times. Therefore, the explicit expressions for D_n^i are given as followings:

- under communication blocking policy

$$D_n^1 = n \ for\ n = 0,\ldots,5,$$
$$D_n^1 = 7 + 3(n-6) \ for\ 6 \le n \ \ 14,$$
$$D_n^1 = 32 + 5(n-14) \ for\ n \ \ 14,$$

and

$$D_n^2 = 1 + 3n \ for\ n = 0,\ldots,8,$$
$$D_n^2 = 29 + 5(n-9) \ for\ n \ \ 9.$$

- under manufacturing blocking policy

$$D_n^1 = n \ for\ n = 0,\ldots,6,$$
$$D_n^1 = 7 + 3(n-7) \ for\ 7 \le n \ \ 16,$$
$$D_n^1 = 34 + 5(n-16) \ for\ n \ \ 16,$$

and

$$D_n^2 = 1 + 3n \ for\ n = 0,\ldots,10,$$
$$D_n^2 = 34 + 5(n-11) \ for\ n \ \ 11.$$

<Table 1> Waiting Times under Communication Blocking with $K_1 = \infty, K_2 \ \ 5, K_3 = 5$

Traffic Intensity ρ	$E(W^1)$		$E(W^2)$	
	Exact Solution	Simulation	Exact Solution	Simulation
0.1	0.01020	0.01041 ± 0.00122	1.09574	1.0921 ± 0.00749

0.2	0.02083	0.02013 m0.00137	1.20455	1.1980 m0.00808
0.5	0.05594	0.05445 m0.00175	1.65546	1.6417 m0.01754
0.8	0.47410	0.4861 m0.11565	4.20172	4.1143 m0.23868
0.9	4.75394	4.5113 m0.79959	12.79257	12.3161 m1.09133

<Table 2> Waiting Times under Manufacturing Blocking with K_1 = ? , K_2 5, K_3 = 5

Traffic Intensity r	$E(W^1)$		$E(W^2)$	
	Exact Solution	Simulation	Exact Solution	Simulation
0.1	0.01020	0.01034 m0.00132	1.09574	1.0919 m0.00789
0.2	0.02083	0.02005 m0.00140	1.20455	1.1980 m0.00806
0.5	0.05560	0.05465 m0.00073	1.64603	1.6372 m0.00852
0.8	0.28457	0.3039 m0.05227	3.55408	3.5639 m0.13519
0.9	3.44201	3.6750 m0.46990	10.89937	11.1558 m0.66525

<Table 3> Waiting Times under Communication Blocking with K_1 = ? , K_2 5, K_3 = 6

Traffic Intensity r	$E(W^1)$		$E(W^2)$	
	Exact Solution	Simulation	Exact Solution	Simulation
0.1	0.01020	0.00965 m0.00056	1.09574	1.0929 m0.00356
0.2	0.02083	0.02053 m0.00057	1.20455	1.2009 m0.00393
0.5	0.05583	0.05501 m0.00082	1.64603	1.6387 m0.00711
0.8	0.34415	0.3649 m0.05629	3.55408	3.5612 m0.12978
0.9	3.88841	4.1139 m0.47854	10.89937	11.1359 m0.65770

<Table 1<Table 1> and <Table 2> show exact and simulation values of the expected waiting times (just before the beginning of service) at node 1 and 2 for various values of traffic intensity. From the numerical results, we can see that our expressions of D_n^i are accurate and that the mean values of waiting times under manufacturing blocking are smaller than or equal to those under communication blocking. In addition, from <Table 2> and <Table 3 > we can obtain the exactly same values of mean waiting times at node 2 for the systems with one difference buffer capacity at node 3 under manufacturing and communication blocking policies which is

addressed in Theorem 1 of section 2.

By using the explicit expressions of D_n^i together with Theorem 2.3 in [2], one can numerically determine the smallest values of K_2 and K_3 that satisfy $R(W^i > t_i)$ b_i with various fixed values of t_i and b_i, for $i = 1,2$. For this particular example, <Table 4> shows the smallest values of finite buffer capacities K_2^* and K_3^* under communication blocking when the arrival rate $l = 0.16$ (traffic intensity $r = 0.8$).

<Table 4> The smallest buffer capacities when $r = 0.8$ under communication blocking

	$b_1 = b_2 = 0.3$		$b_1 = b_2 = 0.1$	
	K_2^*	K_3^*	K_2^*	K_3^*
$t_1 = 8.0$, $t_2 = 25.0$	4	3	4	8
$t_1 = 6.0$, $t_2 = 15.0$	3	5	3	10
$t_1 = 4.0$, $t_2 = 25.0$	5	3	5	8

5. Conclusion

In this paper, we studied waiting times in Poisson driven 3-node tandem queues with finite buffers and deterministic or non-overlapping service times, which is an extended version of the previous study (see [10]). Recursive expressions for waiting times in stochastic system with finite buffers under communication or manufacturing blocking rules can be obtained in (max, +)-algebra notation.

From these explicit expressions we show that the system sojourn times are independent of the capacities of finite buffers when the capacity of buffer at the first node is infinite. It is the same results as the previous studies but is obtained by totally different ways. A relationship on stationary waiting times in the system with finite buffers under two blocking policies is also disclosed. As an application of our results, we consider an optimization problem which determines the smallest buffer capacities under probabilistic constraints on the waiting times.

These results can be extended to more complex (max, +)-linear systems with finite buffers such as m-node tandem queues, fork-and-join type queues, mixture of them, and so forth.

References

1. Ayhan, H. and D. W. Seo, "Laplace Transform and Moments of Waiting Times in Poisson Driven (Max,+)-Linear Systems", *Queueing Systems*, Vol. 37, No. 4 (2001), pp. 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, Vol. 47, No. 1 (2002), pp. 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 & Sons, 1992.

4. Baccelli, F., S. Hasenfuss, and V. Schmidt, "Transient and Stationary Waiting Times in (Max, +) Linear Systems with Poisson Input", *Queueing Systems*, Vol. 26 (1997), pp. 301-342.

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

6. Baccelli, F., and V. Schmidt, "Taylor Series Expansions for Poisson Driven (Max,+) Linear Systems", *Annals of Applied Probability*, Vol. 6, No. 1 (1996), pp. 138-185.

7. Grassmann, W. K. and S. Drekic, "An Analytical Solution for a Tandem Queue with Blocking", *Queueing Systems*, Vol. 36 (2000), pp. 221-235.

8. Labetoulle, J. and G. Pujolle, "A Study of Queueing Networks with Deterministic Service and Application to Computer Networks", *Acta Informatica*, Vol. 7 (1976), pp. 183-195.

9. Nakade, K. "New Bounds for Expected Cycle Times in Tandem Queues with Blocking", *European Journal of Operations Research*, Vol. 125 (2000), pp. 84-92.

10. Seo, D.-W., "Application of (Max,+)-algebra to the Waiting Times in Deterministic 2-node Tandem Queues with Blocking", *Journal of the Korean Operations Research and Management Society*, Vol. 30, No. 1, (2005), pp 149-159.

11. Wan, Y.-W. and R. W. Wolff, "Bounds for Different Arrangements of Tandem Queues with Nonoverlapping Service Times", *Management Science*, Vol. 39, No. 9 (1993), pp. 1173-1178.

12. Whitt, W., "The Best Order for Queues in Series", *Management Science*, Vol. 31, No. 4 (1985), pp. 475-487.