

Approximate Analysis of Transfer Line with PH-service Time and Parts Assemble

Yang Woo Shin¹^a, Gyeong Min Baek²^b, Dong Ok Kim²^c and Dug Hee Moon³^d

¹Department of Statistics, Changwon National University, Changwon, Gyeongnam 51140, Korea

²Department of Eco-friendly Marine Plant FEED Engineering, Changwon National University, Changwon, Gyeongnam 51140, Korea

³School of Industrial Engineering and Naval Architecture, Changwon National University, Changwon, Gyeongnam 51140, Korea

Keywords: Tandem Queue, Phase Type Distribution, Decomposition, Blocking, Assembly.

Abstract: We consider a transfer line in which workstations are linked in series. Each station consists of single machine and two buffers of finite capacity, one for product and the other for part to be assembled. The product is supplied to the first workstation of the system and they are processed along the line, and a part is supplied to each station according to independent Poisson processes. The processing time distribution of each machine is of phase type (PH). Blocking-After-Service (BAS) rule is adopted. In this paper, we present an approximate analysis for the system based on the decomposition method. Some numerical examples are presented for accuracy of approximation.

1 INTRODUCTION

As a motivating example, we consider a railway vehicle manufacturing process. All railway vehicles are manufactured to the customer's specific order. In the railway vehicle assembly shop, various parts are assembled according to the customer's requirements. The layout of the assembly shop is divided into two types: fixed position production and flow production. Here, we consider the flow production layout. The assembly shop has limited space to store work in process due to space constraints. Due to the nature of the assembly shop that performs various processes, automation is difficult, and most of the processes are performed manually. Various parts are supplied to the assembly station as a kit to carry out the process according to the customer's order.

For modeling the flow production layout we consider a flow line in which workstations are linked in series and there is a buffer of finite capacity between two consecutive like this figure. Workstations are assumed to be reliable. The processing time is

assumed to be of phase type (PH) for modelling the manual operation. Parts are supplied to each workstation M_i according to a Poisson process with rate λ_i . The buffer for parts is finite and the parts are lost if the part buffer is full upon an arrival of a part. Each workstation does not work if it has no parts to be assembled. Transportation times of customers through buffers and workstations are assumed to be negligible comparing to processing time. By setting the arrival rate of the parts to be infinite, our model reduces to a transfer line without part assemble which has been widely used for performance modeling of computer systems and production systems. e.g. see the monographs Gershwin (1994), Buzzacott and Shanthikumar (1993), the survey papers Dallery and Gershwin (1992), Li et al. (2009), Papadopoulos and Heavey (1996) and the references therein. In this paper, we present an outline of approximate analysis for tandem queues with a buffer of finite capacity for products between two workstations. Each workstation has a single reliable server and a supplementary buffer of finite for part to be assembled. Our approach is based on well-known decomposition method, e.g. see Gershwin (1978, 1994), Shin and Moon (2018), Moon and Shin (2019).

The paper is organized as follows. The model is described in Section 2. In Section 3, an approxima-

^a  <https://orcid.org/0000-0002-3107-4569>

^b  <https://orcid.org/0000-0001-8421-3430>

^c  <https://orcid.org/0000-0002-7239-2002>

^d  <https://orcid.org/0000-0001-7660-4976>

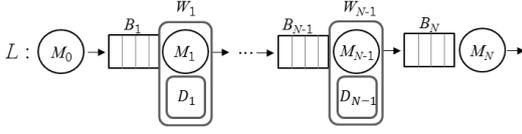


Figure 1: Tandem queue with parts assembly.

tion method is described briefly. Numerical examples are given in Section 4. Finally, concluding remarks are given in Section 5.

2 THE MODEL

We consider a transfer line L in which $\bar{N} = N + 1$ workstations W_i , $i = 0, 1, \dots, N$ are linked in series and there is a buffer B_i for customers (products) of capacity $0 \leq c_i < \infty$ between W_{i-1} and W_i as depicted in Figure 1. The station W_i , $1 \leq i \leq N - 1$ has a server (machine) M_i and a buffer D_i of capacity $0 \leq d_i - 1 < \infty$ for parts to be assembled and denote the station i by the pair $W_i = (M_i, D_i)$.

The customers are processed along the line and leave the system immediately after service completion at W_N . When a server completes its service at a stage, if the buffer for the customers of next station is full at that time, then the server is forced to stop its service and the customer is held at the station where it just completed its service until the destination can accommodate it. A server M_i is said to be starved if there are no customers to be served on the server M_i and the station is said to be lacked if there are no parts in W_i . The station W_i , $1 \leq i \leq N - 1$ does not work if it is in starved, lacked and blocked.

The initial station W_0 and the last station W_N are for preparation and investigation of final product, respectively, and assume that they consist of a server M_0 and M_N without buffers for parts, respectively, and they process their work without parts. We assume that the initial server M_0 is never starved and never lacked and it starts new service immediately after a service completion unless the server is blocked. The last server M_N is never blocked and never lacked, and the customer at M_N leaves the system immediately after completing its service.

Parts are supplied to each station W_i according to independent Poisson processes with rate λ_i and arriving parts enter the server M_i if there is an available space for the part in M_i . Note that the maximum number of parts that can be stored in the station W_i is d_i , $1 \leq i \leq N - 1$. If the buffer D_i is full, that is, if there are d_i parts in the station W_i , the arrivals of the parts to D_i is forced to stop and begin again at the epoch when there is an available space in W_i .

Service time distribution of M_i is of phase type with representation $PH(\boldsymbol{\alpha}_i, T_i)$, where $\boldsymbol{\alpha}_i = (\alpha_i(1), \dots, \alpha_i(h_i))$ is a probability distribution and T_i is a nonsingular matrix of size h_i with negative diagonal elements and nonnegative off-diagonal elements. Let $\mathbf{T}_i^0 = -\mathbf{T}_i \mathbf{e} = (t_i^0(1), \dots, t_i^0(h_i))^t$, where \mathbf{e} is the column vector of appropriate size whose components are all 1. See Neuts (1981) for phase type distribution. Transportation times of customers through buffers and servers are assumed to be negligible comparing to service time.

Stochastic processes. Let $D_i(t)$ be the number of parts in W_i at time t . The state space of $D_i(t)$ is $\{0, 1, \dots, d_i\}$. Define the state $M_i(t)$ of M_i at time t by

$$M_i(t) = \begin{cases} s, & M_i \text{ is starved} \\ j, & M_i \text{ is working with service phase } j \\ b, & M_i \text{ is blocked.} \end{cases}$$

The state space of $W_i(t) = (M_i(t), D_i(t))$ of the station $W_i = (M_i, D_i)$, $1 \leq i \leq N - 1$ is $\mathcal{W}_i = \{(0, 0), \mathbf{s}, \mathbf{w}, \mathbf{b}\}$, where $(0, 0)$ is the state that $X_i(t) \geq 1$ and $D_i(t) = 0$, $\mathbf{s} = \{(s, k), k = 0, 1, \dots, d_i\}$, $\mathbf{w} = \{(j, k), k = 1, 2, \dots, d_i, j = 1, 2, \dots, h_i\}$, $\mathbf{b} = \{(b, k), k = 1, 2, \dots, d_i\}$ and let $\mathbf{w}^* = \{(0, 0), \mathbf{w}\}$. Since W_0 and W_N consist of only one server M_0 and M_1 and have no buffers for part, the state space of $W_0(t) = M_0(t)$ and $W_N(t) = M_N(t)$ are given by $\mathcal{W}_0 = \{b\} \cup \{j, j = 1, 2, \dots, h_0\}$ and $\mathcal{W}_N = \{s\} \cup \{j, j = 1, 2, \dots, h_N\}$.

Let $X_i(t)$ be the total number of customers waiting in the buffer B_i , the customers that are being served, lacked or blocked at M_i , and the customers blocked at M_{i-1} at time t . Then $X_i(t)$ takes values on $\{0, 1, \dots, K_i\}$, where $K_i = c_i + 2$. Note that $X_i(t) = K_i$ is equivalent to $M_{i-1}(t) = b$.

Approximation method. Approximation is based on decomposition approach. The first step is to decompose the $N + 1$ station system into a set of subsystems L_i , $i = 1, 2, \dots, N$. Each subsystem L_i consists of upstream station W_{i-1} , downstream station W_i and a buffer B_i between them. Model the subsystem L_i with a Markov chain with generator, say Q_i and the unknowns in Q_i are calculated by an iteration method. Finally, performance measure such as throughput is calculated with the stationary distribution of L_N . Since W_i is a downstream station in L_i and upstream server in L_{i+1} , denote the downstream server in L_i by W_i^d and the upstream server in L_{i+1} by W_i^u , if necessary to distinguish them.

Note that an arrival of a customer to M_i^u is occurred by a service completion of M_{i-1} and its conditional rate given $W_i^u(t) = x$ is

$$a_i^u(x) = \sum_{j=1}^{h_{i-1}} \sum_{k=1}^{d_{i-1}} P(W_{i-1}(t) = w(j,k) | W_i^u(t) = x) t_{i-1}^0(j).$$

The components of the matrix $B_i^u(\mathbf{x}, \mathbf{y})$ corresponding to the transitions from the state in \mathbf{x} and the state in \mathbf{y} can be given in terms of $a_i^u(x)$ and \mathbf{T}_i^* . For example,

$$B_i^u(\mathbf{s}, \mathbf{w}_1^*) = \begin{pmatrix} a_i^u(s, 0) & O \\ O & \alpha_i \otimes \Delta[a_i^u(\mathbf{s})] \end{pmatrix},$$

where $a_i^u(\mathbf{s})$ be the vectors of size d_i whose k th components are $a_i^u(s, k)$ and $\Delta[\mathbf{x}]$ is the diagonal matrix whose diagonal vector is $\mathbf{x} = (x_1, \dots, x_n)$.

The matrices $A_i^d(n)$. Once a service completion occurs at M_{i-1} , if $X_i(t) = 0$, then the state of $M_i(t)$ changes from s to j with probability $\alpha_i(j)$ and if $X_i(t) \geq 1$, then there are no state transitions of $M_i(t)$. Thus $A_i^d(n) = \mathbf{I}$, $1 \leq n \leq K_i - 1$, $A_N^d(0) = \alpha_N$ and

$$A_i^d(0) = \begin{pmatrix} (s, 0) \\ (s, \cdot) \end{pmatrix} \begin{pmatrix} w(0, 0) & \mathbf{w} & \mathbf{b} \\ 1 & O & O \\ O & \alpha_i \otimes \mathbf{I} & O \end{pmatrix}.$$

The matrices $C_i^d(n)$. Note that when $M_i(t) = w(j)$, if $X_{i+1}(t) = K_{i+1} - 1$, then a service completion at M_i results in blocking of the server M_i and if $X_{i+1}(t) \leq K_{i+1} - 2$, then a departure from M_i occurs. It follows from the observations that the blocking rate $\delta_i^d(w(j, k), 0)$ and the departure rate $\delta_i^d(w(j, k), 1)$ of M_i given $W_i(t) = w(j, k)$ are

$$\begin{aligned} \delta_i^d(w(j, k), 0) &= P(X_{i+1}(t) = K_{i+1} - 1 | W_i(t) = w(j, k)) t_i^0(j), \\ \delta_i^d(w(j, k), 1) &= t_i^0(j) - \delta_i^d(w(j, k), 0). \end{aligned}$$

When $M_i(t) = b$, a departure from M_i is occurred by a departure from M_{i+1} and the rate is

$$\begin{aligned} \delta_i^d(b, k) &= \sum_{y \in \{\mathbf{w}, \mathbf{b}\}} P(W_{i+1}(t) = y | W_i^u(t) = (b, k)) \delta_{i+1}(y) \end{aligned}$$

with $\delta_N(j) = t_N^0(j)$, where

$$\{W_{i-1}^u(t) = (b, k)\} = \{W_{i-1}^u(t) \in \{(b_1, k), (b_2, k)\}\}.$$

The components of $C_i^d(n)$ are given in terms of $\delta_i^d(w(j, k), l)$, $l = 0, 1$ and $\delta_i^d(b, k)$.

The matrices $B_i^d(n)$. The components of $B_i^d(n)$ which are the transition rates of M_i^d without the changes of $X_i(t)$ are given in terms of known parameters of service time α_i , \mathbf{T}_i^* , arrival rate λ_i of parts and the unknowns $\delta_i^d(w(j, k), 0)$, and details are omitted.

3.3 Approximation of Transition Rates

Now we assume that the system is in stationary state and let $\boldsymbol{\pi}_i$ be the stationary distribution of Q_i . We express the parameters $a_i^u(x)$, $\delta_i^u(x, x')$, $p_i^u(x, x')$ for $A_i^u(n)$, B_i^u , $C_i^u(n)$ of the subsystem L_{i+1} and $\delta_{i-1}^d(y)$ in C_{i-1}^d , B_{i-1}^d of the subsystem L_{i-1} in terms of $\boldsymbol{\pi}_i$. For example, the transition rates $\delta_i^u(w_2(j, k), w_1)$ and the probability $p_i^u((b_2, k), w_1)$ for W_i^u are given by

$$\begin{aligned} \delta_i^u(w_2(j, k), w_1) &= \frac{P(X_i(t) = 2, W_i^d(t) = (j, k))}{P(X_i(t) \geq 2, W_i^d(t) = (j, k))} t_i^0(j), \\ p_i^u((b_2, k), w_1) &= \frac{P(X_i(t) = 2, W_i^d(t) = (b, k))}{P(X_i(t) \geq 2, W_i^d(t) = (b, k))} \end{aligned}$$

Similarly, the parameters for W_{i-1}^d in the subsystem L_{i-1} are given as follows, for example,

$$\begin{aligned} \delta_{i-1}^d(w(j, k), 0) &= \frac{P(X_i(t) = K_i - 1, W_{i-1}^u(t) = w(j, k))}{P(W_{i-1}^u(t) = w(j, k))} t_{i-1}^0(j). \end{aligned}$$

Throughput. Once the stationary distribution $\boldsymbol{\pi}_i$ of Z_i is obtained, the throughput can be obtained by

$$\Theta = \sum_{(j,k) \in \mathbf{w}} P(W_N(t) = (j, k)) t_N^0(j).$$

3.4 Algorithm

In this subsection, an iterative algorithm for solving the proposed decomposition equations for the components of Q_i is presented.

Boundary Conditions. Since M_0 is never starved while the iteration, the formulae for the matrices $A_0^u(n)$, $B_0^u(n)$ and $C_0^u(n)$ for M_0 are not changed during iteration. Similarly, the matrices $A_N^d(n)$, $B_N^d(n)$ and $C_N^d(n)$ corresponding to M_N^d are not changed during iteration.

Initial Step. Initially the upstream servers are assumed to be never starved. Then the matrices for upstream servers do not contain unknown parameters. Compute the stationary distribution $\boldsymbol{\pi}_N$ of Q_N .

Step 1. Backward iteration. For $i = N - 1, N - 2, \dots, 1$,

(1) Update $A_i^d(n)$, $B_i^d(n)$ and $C_i^d(n)$. In this step of the first iteration, use the matrices $A_i^u(n)$, $C_i^u(n)$ and $B_i^u(n)$ under the assumptions of no starvation and from the second iteration, use the matrices of general formulae whose components can be calculated by using the formulae for transition rates in subsection 3.3.

(2) Compute the stationary distribution $\boldsymbol{\pi}_i$ of Q_i .

If $i = 1$, GO TO next step.

Step 2. Forward iteration. For $i = 1, 2, 3, \dots, N - 1$,
 (1) Update $A_i^u(n)$, $B_i^u(n)$ and $C_i^u(n)$ using the formulae derived in subsection 3.3
 (2) Compute π_{i+1} of Q_{i+1} .
 If $i = N - 1$, then GO TO next step.

Step 3. Calculate throughput and check the stopping criterion

$$TOL = |\Theta^{(m)} - \Theta^{(m-1)}| < \epsilon, \quad (4)$$

where $\Theta^{(m)}$ is the throughput obtained in the m th iteration and $\epsilon > 0$ is the tolerance predetermined. If the stopping criterion is not satisfied, GO TO Step 1 and repeat the backward and forward iteration until the stopping criterion is satisfied.

4 NUMERICAL RESULTS

The accuracy of the method is investigated numerically by comparing approximations (App) with the simulations (Sim). We consider the system with 6 workstations where the service times are identical with common means 1.0 and the buffer size for customers between workstations are identical. The buffer size c_i between workstations are chosen as $c_i = 0, 3, 5$. We use the Erlang distribution of order k (E_k) for the squared coefficient of variation of service time $C_s^2 = \frac{1}{k} \leq 1$ and hyperexponential distribution of order 2 (H_2) with balanced mean and $C_s^2 = 2.0$. Tolerance $\epsilon = 10^{-5}$ is used for stopping criterion (4). Simulation models for the systems in the tables are developed with ARENA. Simulation run time is set to 100,000 unit times including 30,000 unit times of warm-up period. Ten replications are conducted for each case and 95% confidence intervals (c.i.) are calculated. The intervals are omitted in the following tables, but we have observed that all the approximation results are contained in the intervals. The deviation (\mathcal{D}) between approximation and simulation is calculated by $\mathcal{D}(\%) = \frac{App-Sim}{Sim} \times 100$. The throughput for the system is presented in Tables 1-2. Numerical results show that the approximation performs well.

5 CONCLUSIONS

In this paper, an approximate analysis for tandem queues with blocking and parts assembly has been presented. The approximation is based on the decomposition method. To reflect the dependence between consecutive stages, the states of the servers in subsystems are indicated by the state of the number of

Table 1: Throughput for the system with E_2 -service time.

λ_i	d_i	c_i 0	3	5
		Sim	Sim	Sim
		App($\mathcal{D}(\%)$)	App($\mathcal{D}(\%)$)	App($\mathcal{D}(\%)$)
0.5	1	0.316	0.371	0.386
		0.315(-0.4)	0.371(-0.1)	0.386(-0.0)
	3	0.388	0.424	0.435
	0.391(+0.8)	0.425(+0.3)	0.437(+0.4)	
	5	0.420	0.443	0.451
0.425(+1.3)	0.447(+0.8)	0.454(+0.7)		
1.0	1	0.479	0.605	0.628
		0.468(-2.4)	0.603(-0.3)	0.627(-0.1)
	3	0.546	0.725	0.752
	0.529(-3.0)	0.722(-0.4)	0.752(-0.0)	
	5	0.554	0.771	0.801
0.537(-3.2)	0.767(-0.6)	0.801(-0.0)		

Table 2: Throughput for the system with H_2 -service time.

λ_i	d_i	c_i 0	3	5
		Sim	Sim	Sim
		App($\mathcal{D}(\%)$)	App($\mathcal{D}(\%)$)	App($\mathcal{D}(\%)$)
0.5	1	0.279	0.345	0.361
		0.275(-1.3)	0.342(-0.9)	0.360(-0.5)
	3	0.336	0.401	0.414
	0.331(-1.7)	0.396(-1.1)	0.413(-0.4)	
	5	0.362	0.426	0.438
0.355(-1.8)	0.423(-1.0)	0.437(-0.2)		
1.0	1	0.363	0.500	0.536
		0.362(-0.4)	0.494(-1.2)	0.533(-0.6)
	3	0.391	0.571	0.619
	0.391(0.1)	0.559(-2.0)	0.611(-1.2)	
	5	0.393	0.594	0.649
0.395(0.5)	0.581(-2.2)	0.641(-1.2)		

customers in upstream subsystem as well as the states of the server (blocking, starvation, working, lack of parts), and the transitions among the states are considered. Numerical experiments indicated that the method works reasonably.

The approach can be applied to the more complex systems such as the system with unreliable servers and the system with more general service time by the versatility of PH-distribution, see e.g. Altioik (1985), Osogami and Harchol-Balter (2006) and references therein. Furthermore, our approach can be applied to the optimization problem for control of parts buffers and arrival rates of parts.

ACKNOWLEDGEMENTS

This paper was supported by Basic Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Grant Numbers NRF-2018R1D1A1A09083352 and NRF-2019R1F1A1057692.

REFERENCES

- Altiook, T. (1985). On the phase-type approximations of general distributions. *IIE Transactions* 17(2), 110-116.
- Buzzacott, J. A., Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*, Prentice-Hall.
- Dallery, Y., Gershwin, B. (1992). Manufacturing flow line systems: a review of models and analytical results. *Queueing Systems* 12, 3-94.
- Gershwin, S. B. (1987). An efficient decomposition algorithm for the approximate evaluation of tandem queues with finite storage space and blocking. *Operations Research* 35, 291-305.
- Gershwin, S. B. (1994). *Manufacturing systems engineering*. Prentice-Hall, Englewood Cliffs.
- Throughput analysis of production systems: recent advances and future topics. *International Journal of Production Research* 47(14), 3823-3851.
- Moon, D. H., Shin, Y. W. (2019). Approximation of tandem queues with blocking. In *Proceedings of the 8th International Conference on Operations Research and Enterprise Systems (ICORES 2019)*, Prague, February, 2019, pp. 422-428
- Neuts, M. F., (1981). *Matrix Geometric Solutions in Stochastic Models*. Dover Publishing Co., New York.
- Osogami, T., Harchol-Balter, M. (2006). Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluation* 62, 524-552.
- Papadopoulos, H. T., Heavey, C. (1996). Queueing theory in manufacturing systems analysis and design: a classification of models for production and transfer lines. *European Journal of Operational Research* 92, 1-27.
- Shin, Y. W., Moon, D. H. (2018). Approximation of discrete time tandem queueing networks with unreliable servers and blocking. *Performance Evaluation* 120, 49-74.