Analysis of a Fork/Join Synchronization Station with Inputs from Coxian Servers in a Closed Queuing Network

Ananth Krishnamurthy
Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180
Phone: 518-276 2958, Fax: 518-276 8227
Email: krisha@rpi.edu

Rajan Suri
Center for Quick Response Manufacturing
University of Wisconsin-Madison
1513 University Avenue, Madison, WI 53706-1572
Telephone: 608-262 0921, Fax: 608-265 4017
Email: suri@engr.wisc.edu

and

Mary Vernon
Department of Computer Sciences
University of Wisconsin-Madison
1210 W Dayton Street, Madison, WI-53706
Telephone: 608-262 7893, Fax: 608-262 9777
Email: vernon@cs.wisc.edu

Abstract

Fork/join stations are commonly used to model the synchronization constraints in queuing models of computer networks, fabrication/assembly systems and material control strategies for manufacturing systems. This paper presents an exact analysis of a fork/join station in a closed queuing network with inputs from servers with two-phase Coxian service distributions, which models a wide range of variability in the input processes. The underlying queue length and departure processes are analyzed to determine performance measures such as throughput, distributions of the queue length and inter-departure times from the fork/join station. The results show that, for certain parameter settings, variability in the arrival processes has a significant impact on system performance. The model is also used to study the sensitivity of performance measures such as throughput, mean queue lengths, and variability of inter-departure times for a wide range of input parameters and network populations.

To appear in Annals of Operations Research-special issue on Stochastic Models of Production-Inventory Systems

1 Introduction

Fork/join stations are used to model synchronization constraints between entities in a queuing network. The fork/join station of interest in this paper consists of a server with zero service times and two input buffers. As soon as there is one entity in each buffer, an entity from each of the buffers is removed and joined together. The joined entity exits the fork/join station instantaneously. Subsequent to its departure, the joined entity forks back into the component entities, which then get routed to other parts of the network. Fork/join stations find many applications in queuing models of manufacturing and computer systems. In queuing models of fabrication/assembly systems, the assembly station is typically modeled using a fork/join station (Harrison, 1973; Latouche, 1981; Hopp and Simon, 1989; Rao and Suri, 1994, 2000). Recently fork/join stations have also been used to model the synchronization constraints in models of material control strategies for multi-stage manufacturing systems (Di Mascolo *et al*., 1996; Krishnamurthy *et al.*, 2001). In computer systems analysis, queuing networks with fork/join stations have been studied in the context of parallel processing, database concurrency control, and communication protocols (Baccelli *et al*., 1989; Varki, 1999; Prabhakar *et al.,* 2000).

To develop efficient methods for analyzing these networks, several researchers have analyzed different aspects of the performance of fork/join stations in *isolation*. The typical inputs for such an analysis are the capacity of each input buffer and a description of the inter-arrival times of the entities to each input buffer. Performance measures of interest include synchronization delays, queue length distributions at the different input buffers, and station throughput. For the sake of analytical tractability, a majority of these efforts assume that the inter-arrival times of the inputs to the fork/join station have an exponential distribution (Harrison, 1973; Bhat, 1986; Lipper and Sengupta, 1986; Hopp and Simon, 1989; Som *et al.*, 1994; Takahashi *et al*., 1996; Varki, 1999). Although these results are useful, in many of the applications cited above the inter-arrival times have a distribution more different from the exponential. Recent studies by Takahashi *et al*. (2000), and Takahashi and Takahashi (2000) use matrix analytical methods to analyze fork/join stations where the inter-arrival times have phase type distributions. However, they assume that the arrivals are from infinite populations, i.e. the arrivals are an uninterrupted process and an arriving entity that finds the input buffer full is lost. Additionally, the computational complexity arising from the use of matrix analytical approaches seems to be a high price to pay for analyzing fork/join stations with inputs having inter-arrival times that have a distribution more general than the exponential.

When the fork/join station is part of a *closed* queuing network, (such as in models of multi-stage kanban systems or closed multi-level fabrication/assembly systems) then once the queue length of an input buffer equals the size of the population that can arrive to the buffer, the arrival process shuts down temporarily. In this paper, we propose an approach for analyzing such fork/join stations for a fairly general class of inputs, with reduced computational complexity. More specifically, we study the performance of a fork/join station in a closed queuing network with inputs from servers with two-phase Coxian service distributions. The choice of two-phase Coxian distribution allows us to model input processes with a wide range of mean $(0,\infty)$ and squared coefficient of variation, SCV $[0.5,\infty)$. We analyze the queue length process as a continuous time Markov chain to estimate the throughput and queue length distributions. Next we analyze the departure process as a semi-Markov process and derive expressions for the distribution of the inter-departure times after deriving the state transition probabilities for the embedded Markov chain.

The remainder of this paper is organized as follows. Section 2 defines our model of the fork/join station, lists the specific inputs and outputs, and describes our analysis approach. Section 3 provides a summary of the literature to date on the analysis of fork/join stations. Section 4 describes our analysis of the queue length process while Section 5 describes our analysis of the departure process from the fork/join station. Section 6 provides some numerical results and Section 7 presents our conclusions.

2 System Description and Overview of Analysis

2.1 System Description

We describe our model of the fork/join station and explain how it could be used to represent the synchronization behavior in particular manufacturing and computer systems. The model is illustrated in Figure 1. As shown in the figure the fork/join station has two input buffers B_1 and B_2. If an entity arriving in buffer B_1 (B_2) finds input buffer B_2 (B_1) empty, it waits for the corresponding entity to arrive in input buffer B_2 (B_1). As soon as there is at least one entity in each buffer, one entity is removed from each buffer. The removed entities join together, and immediately depart from the fork/join station. As a result the content of each input buffer is reduced by one. Subsequent to departure from the fork/join station, the joined entity forks back into two entities that are routed to separate stations with single servers. The service times at these servers have two-phase Coxian distributions. At these servers, each entity queues for service, and upon completion of service the entity revisits the fork/join station. There is a finite population of size K_i for the entity of type i. Consequently, the number of entities in input buffer B_i and at the corresponding Coxian server always sum up to K_i, i=1,2, and the arrival process to buffer B_i shuts down when there are K_i units in buffer B_i.

Fork/join stations with such characteristics are found in models of multi-stage kanban systems, closed multi-level fabrication/assembly systems, and communication systems. We provide examples for each case.

1. First, the fork/join station described above can represent a synchronization station before an assembly operation in a fabrication/assembly system (Rao and Suri, 1994, 2000). In this case K_i could correspond to the fixed number of automated guided vehicles (AGVs) transporting components of type i from the fabrication sub-network to the assembly station. Entities in buffers B_1 and B_2 correspond to the fabricated parts that are to be assembled. The join operation corresponds to the kitting operation, while the fork operation corresponds to the release of free AGVs to carry the parts required for assembly. These AGVs would go back to the fabrication sub-networks to be restocked with parts. The arrival of reloaded AGVs from the fabrication sub-networks could be modeled using a server with service times having a suitable two-phase Coxian distribution.

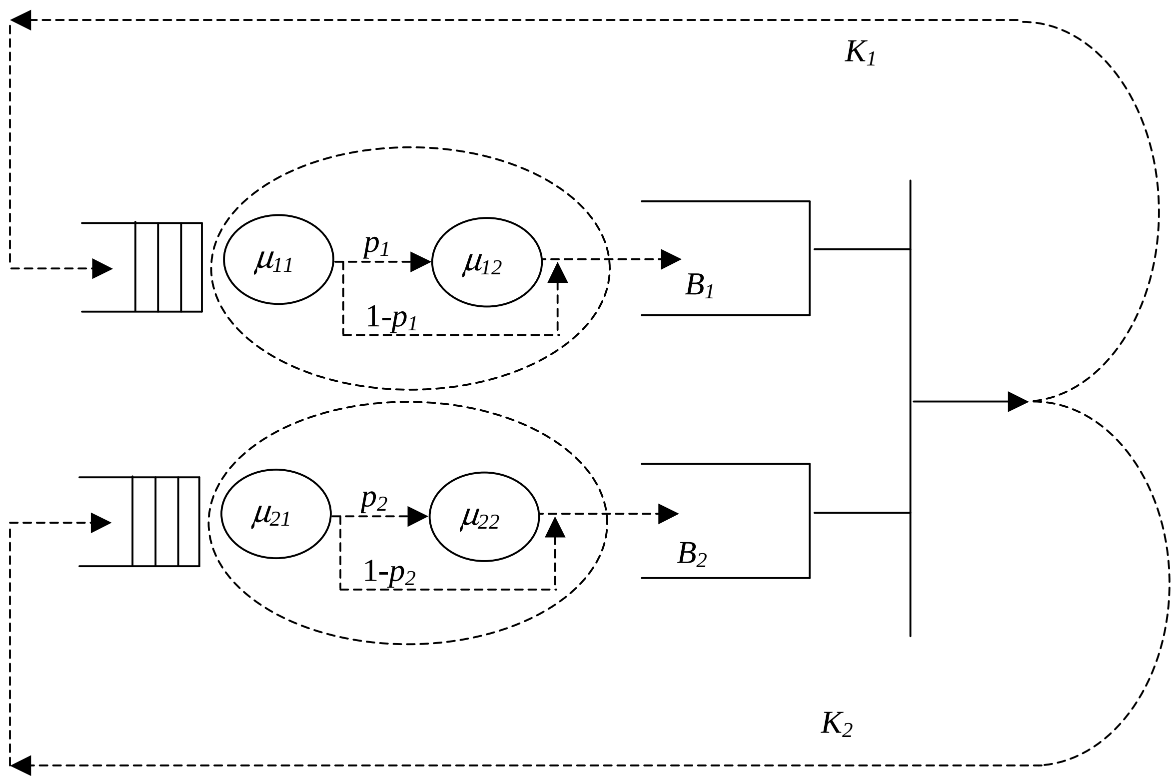

Figure 1. Fork/join station in a closed network having servers with two-phase Coxian service distributions

2. As a second example, the fork/join station model could represent the synchronization constraint in a kanban control system. In modern manufacturing, kanban systems are a popular form of material control (Di Mascolo *et al.*, 1996; Liberopoulos and Dallery, 2000). Here the fork/join station could model the synchronization constraint between an upstream and downstream stage in a multi-stage kanban system. Each entity in buffer B_1 would correspond to a part with an upstream kanban attached to it, while each entity in buffer B_2 would correspond to a free kanban returning from the downstream stage, and K_1 and K_2 would be the number of kanbans in the respective stage. During the join operation a part and upstream kanban are joined with a downstream kanban and during the fork operation, the upstream kanban is sent back, while the part and downstream kanban are sent to the next manufacturing stage. The manufacturing process in each stage could be modeled using a server with service times having a suitable two-phase Coxian distribution.

3. Finally, this fork/join station model can also be applied to represent the synchronization behavior of parallel programs contending for shared resources in a parallel or distributed computer system. For example, see Heidelberger and Trivedi (1983).

2.2 Inputs and Outputs

We assume that the two-phase Coxian distribution at server *i* is defined by three parameters, μ_{i1}, μ_{i2} and p_i, with cumulative distribution function

$$G_i(t) = 1 - C_{i1}e^{-\mu_{i1}t} - C_{i2}e^{-\mu_{i2}t} \text{ for } t \geq 0 \text{ and } i = 1,2 \tag{1}$$

where

$$C_{i1} = \frac{\mu_{i1}(1-p_i) - \mu_{i2}}{\mu_{i1} - \mu_{i2}} \text{ and } C_{i2} = 1 - C_{i1}, \text{ with } \mu_{i1} \neq \mu_{i2} \tag{2}$$

Note that this implies that whenever the number of entities at buffer B_i is less than K_i, entities arrive at the rate $\lambda_i = \dfrac{\mu_{i1}\mu_{i2}}{p_i\mu_{i1} + \mu_{i2}}$, and the SCV of the inter-arrival times is $c_i^2 = 1 - \dfrac{2p_i\mu_{i1}(\mu_{i2} - \mu_{i1}(1-p_i))}{(p_i\mu_{i1} + \mu_{i2})^2}$ for $i = 1,2$ (Altiok, 1996).

2.3 Overview of Our Analysis

For a fork/join station characterized as above, our goal is to compute the throughput χ_D, the mean queue lengths $\overline{L}_1$ and $\overline{L}_2$ at the input buffers B_1 and B_2 and the marginal distribution $G_D(t)$ of the inter-departure times from the fork/join station. Correspondingly, our analysis of the fork/join station consists of two parts: (1) the analysis of the *queue length process* and (2) the analysis of the *departure process*. We briefly summarize the analysis approach in the following paragraphs.

Analysis of the queue length process: To analyze the queue length process, we first define the state space for the queue length process and then analyze the queue length process as a continuous time Markov process. We solve the continuous time Markov chain to obtain the steady state probability distributions of the queue lengths at buffers B_1 and B_2, respectively. From these we obtain different performance measures such as throughput and the mean queue lengths. The details are described in Section 4.

Analysis of the departure process: We note that the output process is a Markov renewal process (Disney and Kiessler, 1987). In analyzing this Markov renewal process, we make use of the special structure of the Markov process embedded at departure instants to obtain the transition probability matrix and stationary probabilities of the Markov chain embedded at departure instants. Using the stationary probabilities, we obtain the marginal distribution of inter-departure times from the fork/join station. The details are given in Section 5.

3 Review of Previous Work

Fork/join stations with two or more inputs have been used to model synchronization constraints in queuing models of computer and manufacturing systems. Harrison (1973) analyzed a fork/join station with renewal input streams and showed that when there are no capacity limits for the input buffers, the fork/join station is unstable. However, if the buffers are bounded, then the fork/join station is stable. Bhat (1986) analyzed a fork/join station with bounded buffers and Poisson inputs and derived expressions for the queue length distributions at the input buffers. Kashyap (1965) studied a kitting process with input queues of components. The kitting process was modeled as a double-ended queue and expressions for the waiting time distributions were derived for the case of Poisson inputs. Som *et al.* (1994) and Takahashi *et al*. (1996) studied the departure process from a fork/join station with finite buffers and exponentially distributed inter-arrival times to these buffers. They derived expressions for the marginal distribution of the inter-departure times. Varki (1999) assumes a finite customer population in the fork/join queuing network but restricts all service times to be exponentially distributed.

In many applications the inter-arrival times have a distribution different from the exponential. Recent studies by Takahashi *et al*. (2000), and Takahashi and Takahashi (2000) use matrix analytical methods to analyze fork/join stations where the distribution of inter-arrival times is of a phase type. However, they assume that the arrivals are from infinite populations, i.e. the arrivals are an uninterrupted process and an arriving entity that finds the input buffer full is lost. When the fork/join station is a part of a closed queuing network, then once the contents in the input buffer reach a certain level, the arrival process shuts down temporarily.

In this paper, we analyze the performance of a fork/join station in a closed queuing network with inputs from servers with two-phase Coxian service distributions. The choice of two-phase Coxian distributions allows us to model a wide range of input processes, namely, input processes with mean inter-arrival times ranging over $(0,\infty)$ and with squared coefficient of variation (SCV) in the range of $[0.5,\infty)$. This range covers the values typically expected of traffic processes in many practical systems (Kamath *et al*., 1988; Buzacott and Shanthikumar, 1993).

4 Analysis of the Queue Length Process

In this section, we analyze the queue length process as a continuous time Markov process. Table 1 summarizes the notation used in the analysis and in the rest of the paper.

Let $N_1(t)$ and $N_2(t)$ denote the number of units in buffers B_1 and B_2 respectively at time *t.* From the operational characteristics of the fork/join station we note that it is not possible for both the buffers B_1 and B_2 to be non-empty for any finite time. Departures occur instantaneously from the fork/join station whenever both buffers are non-empty. Therefore, the number of units in both buffers at time *t* can be described uniquely using the one-dimensional random variable $N(t) = N_1(t) - N_2(t)$. $N(t)$ takes on values $-K_2,\ldots -1,0,1,\ldots,K_1$. For instance, $N(t) = -K_2$ implies that the buffer B_1 is empty and the buffer B_2 has K_2 units. If both input buffers are empty, $N(t)$ =0.

Table 1. Notation used for our analysis

Symbol	*Description*
μ_{j1}, μ_{j2} and p_j	Parameters of the two-phase Coxian distribution for the arrival process at buffer B_j, j=1,2
K_j	Size of the finite population from which arrivals occur to input buffer B_j, j=1,2
λ_j	Rate of arrivals to buffer B_j, when it is not shut down, j=1,2
c_j^2	SCV of inter-arrival times at buffer B_j, when it is not shut down, j=1,2
$N_j(t)$	Number of units in buffer B_j at time t, j=1,2
$J_j(t)$	Phase of pending arrival to buffer B_j time t, j=1,2
$(N(t), J_1(t), J_2(t))$	State of the fork/join station at time t, $N(t) = N_1(t) - N_2(t)$
$\overline{L}_j$	Average queue length at buffer B_j, j=1,2
χ_D	Throughput of the fork/join station

To describe the state of the system at any time t, we need to consider both-the number of units in each input buffer, as well as the phases of the pending arrivals. Then at time t, each buffer can be in one of three distinct states as defined below:

$$
J_1(t)\ [J_2(t)] = \begin{cases} 1 & \text{if the pending arrival to buffer } B_1\ [B_2] \text{ is in phase 1} \\ 2 & \text{if the pending arrival to buffer } B_1\ [B_2] \text{ is in phase 2} \\ 0 & \text{if the buffer } B_1\ [B_2] \text{ is full i.e. has } K_1\ [K_2] \text{ units and} \\ & \text{the arrival process to the buffer has shut down.} \end{cases} \tag{3}
$$

With these definitions, the state of the system is completely characterized by $(N, J_1, J_2) = \{(N(t), J_1(t), J_2(t)), t \geq 0\}$. Clearly, (N, J_1, J_2) is a continuous time Markov chain that describes the stochastic behavior of the system. The state space of state of (N, J_1, J_2) is given by

$$
\begin{aligned} S_Q = &\{(n, j_1, j_2) : n = -K_2 + 1, \ldots -1, 0, 1, \ldots, K_1 - 1;\ j_1 = 1,2;\ j_2 = 1,2\} \\ &\cup \{(-K_2, 1, 0), (-K_2, 2, 0), (K_1, 0, 1), (K_1, 0, 2)\} \end{aligned} \tag{4}
$$

Note that the Markov chain has 4(K_1+K_2) states. The state transition rates for the continuous time Markov chain representing the queue length process are illustrated in Figure 2. These transition rates are easily derived from the parameters of the two-phase Coxian distribution for the arrival processes, μ_{j1}, μ_{j2} and p_j, j=1,2. For example, if $n < K_1$, transitions from the state (n-1, 1, 2) to the state (n, 1, 2) occur when the arrival process for buffer B_1, completes stage 1 (at the rate μ_{11}) and stage 2 of the arrival process is not selected giving a total rate of $\mu_{11}(1-p_1)$.

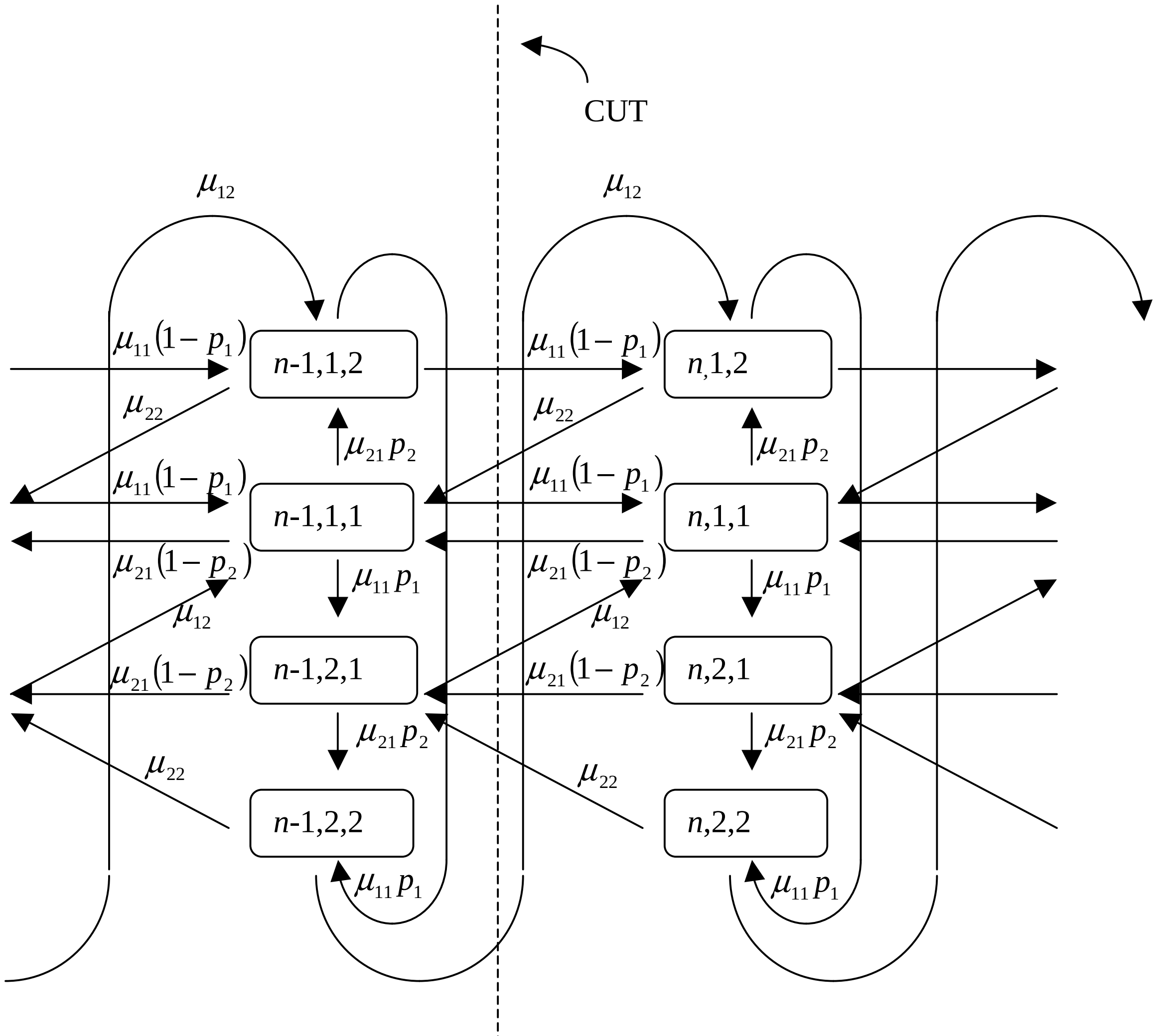

Figure 2. State transitions diagram when $-K_2+2 \le n \le K_1-2$

4.1 Transition Equations

For each $(n, j_1, j_2) \in S_Q$, let $P_Q(n, j_1, j_2)$ denote the steady state probability that the continuous time Markov chain (N, J_1, J_2) is in this state. These steady state probabilities can be computed by solving the corresponding state transition equations, as follows. Corresponding to whether one of the arrival processes to the buffers has *shut down* (*SD*) or both the arrival processes are *not shut down* (*NSD*) we partition the state space S_Q into S_Q^{SD} and S_Q^{NSD} where

$S_Q^{NSD} = \{(n, j_1, j_2) : n = -K_2 + 1, \ldots -1, 0, 1, \ldots, K_1 - 1;\ j_1 = 1,2;\ j_2 = 1,2\}$ and

$$S_Q^{SD} = \{(-K_2,1,0), (-K_2,2,0), (K_1,0,1), (K_1,0,2)\} = S_Q - S_Q^{NSD} \tag{5}$$

Then considering the partition of S_Q^{NSD}, we have that in steady state, for each $(n, j_1, j_2) \in S_Q^{NSD}$ and $\mathbf{P_Q}(n) = \left[P_Q(n,1,2) \quad P_Q(n,1,1) \quad P_Q(n,2,1) \quad P_Q(n,2,2)\right]^T$:

$$\begin{bmatrix} \mu_{11}(1-p_1) & \mu_{11}(1-p_1) & \mu_{12} & \mu_{12} \\ \mu_{11}(1-p_1) & 0 & 0 & \mu_{12} \\ 0 & -\mu_{11}p_1 & \mu_{12}+\mu_{21} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{P_Q}(n-1) = \begin{bmatrix} \mu_{22} & \mu_{21}(1-p_2) & \mu_{21}(1-p_2) & \mu_{22} \\ \mu_{11}+\mu_{22} & -\mu_{21}p_2 & 0 & 0 \\ 0 & 0 & \mu_{21}(1-p_2) & \mu_{22} \\ -\mu_{11}p_1 & 0 & -\mu_{21}p_2 & \mu_{12}+\mu_{22} \end{bmatrix} \mathbf{P_Q}(n) \tag{6}$$

For $(K_1,0,1)$ and $(K_1,0,2) \in S_Q^{SD}$ we have:

$$\begin{bmatrix} \mu_{11}(1-p_1) & 0 & 0 & \mu_{12} \\ 0 & \mu_{11}(1-p_1) & \mu_{12} & 0 \end{bmatrix} \mathbf{P_Q}(K_1-1) = \begin{bmatrix} -\mu_{21}p_2 & \mu_{22} \\ \mu_{21} & 0 \end{bmatrix} \begin{bmatrix} P_Q(K_1,0,1) \\ P_Q(K_1,0,2) \end{bmatrix} \tag{7}$$

For $(-K_2,1,0)$ and $(-K_2,2,0) \in S_Q^{SD}$ we have:

$$\begin{bmatrix} \mu_{22} & \mu_{21}(1-p_2) & 0 & 0 \\ 0 & 0 & \mu_{21}(1-p_2) & \mu_{22} \end{bmatrix} \mathbf{P_Q}(-K_2+1) = \begin{bmatrix} \mu_{11} & 0 \\ -\mu_{11}p_1 & \mu_{12} \end{bmatrix} \begin{bmatrix} P_Q(-K_2,1,0) \\ P_Q(-K_2,2,0) \end{bmatrix} \tag{8}$$

Additionally we have the following:

$$\sum_{(n,j_1,j_2)\in S_Q} P_Q(n, j_1, j_2) = 1 \tag{9}$$

Solving the system of equations 6-9 above, for each $(n, j_1, j_2) \in S_Q$, we obtain expressions for the steady state probabilities $P_Q(n, j_1, j_2)$.

4.2 Average Queue Length and Throughput

Given the steady state probabilities of the queue length process, the average queue lengths $\overline{L}_1$ and $\overline{L}_2$ at the buffers B_1 and B_2 are obtained using equations 10 and 11 below.

$$\overline{L}_1 = \sum_{n=1}^{K_1-1}\sum_{j_1=1}^{2}\sum_{j_2=1}^{2} nP_Q(n, j_1, j_2) + \sum_{j_2=1}^{2} K_1 P_Q(K_1, 0, j_2) \tag{10}$$

$$\overline{L}_2 = \sum_{n=-K_2+1}^{-1} \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} |n| P_Q(n, j_1, j_2) + \sum_{j_1=1}^{2} K_2 P_Q(-K_2, j_1, 0) \qquad (11)$$

Thus we have obtained the probability distribution of the queue length process as well as the mean queue lengths at both the input buffers. The throughput χ_D of the fork/join station is computed from the steady state probabilities of the queue length process as follows:

$$\chi_D = \left\{ \begin{array}{l} \sum_{n=-K_2+1}^{-1} \left(P_Q(n,1,2) + P_Q(n,1,1)\right)\mu_{11}(1-p_1) + \left(P_Q(n,2,2) + P_Q(n,2,1)\right)\mu_{12} + \\ \sum_{n=1}^{K_1-1} \left(P_Q(n,2,1) + P_Q(n,1,1)\right)\mu_{21}(1-p_2) + \left(P_Q(n,1,2) + P_Q(n,2,2)\right)\mu_{22} + \\ P_Q(-K_2,1,0)\mu_{11}(1-p_1) + P_Q(-K_2,2,0)\mu_{12} + P_Q(K_1,0,1)\mu_{21}(1-p_2) + P_Q(K_1,0,2)\mu_{22} \end{array} \right\} \qquad (12)$$

In section 6 we will provide some numerical results of our queue length analysis. First, we complete our analysis of the departure process of the fork/join station.

5 Analysis of the Departure Process

As noted in Section 2.3, the departure process of the fork/join station is a Markov renewal process (Disney and Kiessler, 1987). To analyze this process, we first identify the states in the Markov chain embedded at departure instants. Next, we recognize that this embedded Markov process has a special structure and use this information to identify the possible sample paths between successive departures. By deriving the conditional probability distributions for each sample path, we obtain the transition probability matrix P_D, of the Markov chain embedded at the departure instants. We solve the Markov chain to obtain the stationary probability vector $\prod$. Using $\prod$ and the distributions of the inter-arrival times at the two buffers, we obtain $G_D(t)$, the steady state marginal distribution of inter-departure times from the fork/join station. The analysis is summarized in Figure 3.

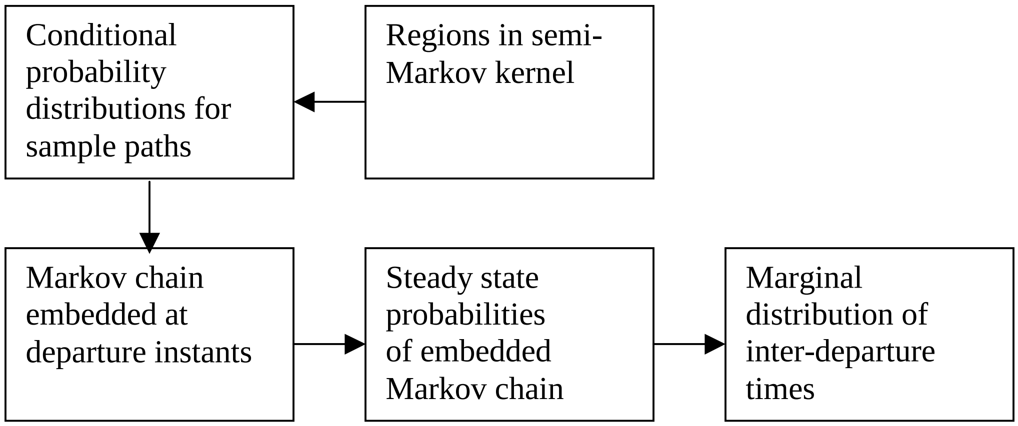

Figure 3. Analysis of the departure process at the fork/join station

5.1 The Semi-Markov Departure Process

A departure occurs (simultaneously from both the buffers) whenever at least one unit is available in each input buffer, B_1 and B_2. The sequence of states of the fork/join station at the departure instants and the corresponding sequence of departure times describe a pair of related random processes (X, D). The random process $D = \{\tau_m : m \in N\}$, where N is the set of all nonnegative integers, is a sequence of real valued random variables τ_m, where τ_m is the time of the m^{th} departure. Note that, each departure time for the fork/join station coincides with either an arrival time in buffer B_1 or an arrival time in buffer B_2. Associated with the random process D is the incremental process $T = \{T_m : m \in N\}$ where for $m > 0$, T_m is the time between consecutive departures m and m-1, i.e., $T_m = \tau_m - \tau_{m-1}$, and $T_0 = 0$. The states of the random process $X = \{X_m : m \in N\}$ is vector valued and consists of the triple (N_m, J_{1m}, J_{2m}) where $N_m = N(\tau_m^+)$, $J_{1m} = J_1(\tau_m^+)$ and $J_{2m} = J_2(\tau_m^+)$. τ_m^+ denotes the time instant just after a departure at τ_m, $N(\tau_m^+)$ denotes the number of units at the fork/join station, and $J_1(\tau_m^+)$ and $J_2(\tau_m^+)$ respectively denote the phases of the pending arrivals to input buffers B_1 and B_2 just after the m^{th} departure.

We derive the distribution of the inter-departure times by analyzing the pair of random processes (X, T). T_{m+1} depends on the present state X_m and the next state X_{m+1}. However, given these states, T_{m+1} is independent of the previous states $X_0, ..., X_{m-1}$ and $T_0, ..., T_m$. That is, the following relation holds

$$P(X_{m+1}, T_{m+1} \mid X_0, ..., X_m, T_0, ..., T_m) = P(X_{m+1}, T_{m+1} \mid X_m) \text{ for all } m \in N \tag{13}$$

and the output process (X, T) is a Markov renewal process. The state space of the output process (X, T) is $S_D \times R$, where $R = (0, \infty)$, and

$$S_D = \{(-K_2+1,1,1)\} \cup \{(n,1,1),(n,1,2), -K_2+1<n<0\} \cup \\ \{(0,1,1),(0,1,2),(0,2,1)\} \cup \{(n,1,1),(n,2,1), 0<n<K_1-1\} \cup \{(K_1-1,1,1)\} \quad (14)$$

In Figure 4 the set of feasible states in S_D are shown in gray. It can be easily verified that for each state in S_D all exit transitions lead to states in S_D.

To define these feasible states more precisely, we note that $S_D \subset S_Q$ and define $S_{Q-D} = S_Q - S_D$. States in S_{Q-D} are shown in dotted boxes in Figure 4. These states are not in S_D for the following reasons. First, it is obvious that neither of the buffer arrival processes is shut down immediately after a departure from the fork/join station, i.e., $S_Q^{SD} \subset S_{Q-D}$. For $X_m \in S_Q^{NSD}$ we reason in the following manner. If $N_m = -K_2+1$, buffer B_2 is full and its arrival process is shut down prior to the departure instant implying that $\{(-K_2+1,1,2),(-K_2+1,2,1),(-K_2+1,2,2)\} \subset S_{Q-D}$. Similarly, for $N_m = K_1 - 1$, the arrival process to buffer B_1 is shut down prior to the departure instant implying that $(K_1-1,1,2),\{(K_1-1,2,1),(K_1-1,2,2)\} \subset S_{Q-D}$. Furthermore, for $-K_2+1 < N_m < 0$, the departure time τ_m coincides with an arrival time in buffer B_1 and hence $\{(N_m,2,1),(N_m,2,2)\} \subset S_{Q-D}$. Similarly for $0 < N_m < K_1 - 1$, the departure time τ_m coincides with an arrival time in buffer B_2 and hence $\{(N_m,1,2),(N_m,2,2)\} \subset S_{Q-D}$. Finally, when $N_m = 0$, we have $\{(0,2,2)\} \subset S_{Q-D}$.

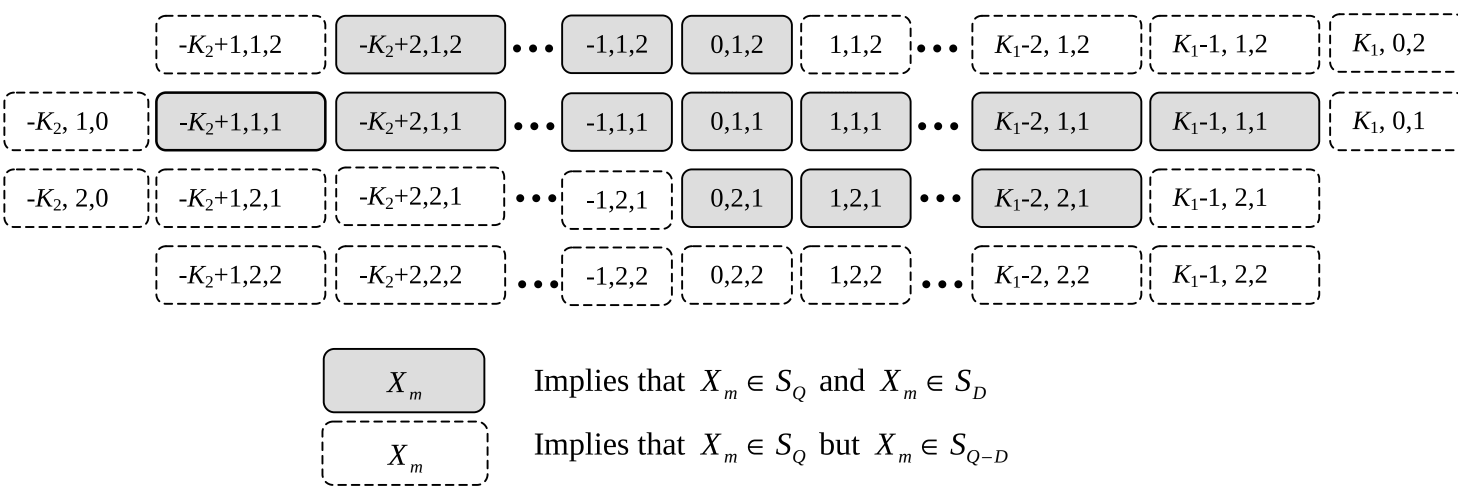

Figure 4. Identifying the states of the Markov chain embedded at departure instants

From Figure 4 and equation 14, it is easy to see that S_D has $2(K_1+K_2)-3$ states. Since the output process (X,T) is a Markov renewal process it is completely characterized by its semi-Markov kernel Q, where:

$$Q(X_m, X_{m+1}, t) = P\left(X_{m+1}, T_{m+1} \le t \mid X_m\right) \tag{15}$$

For convenience the elements of the semi-Markov kernel can be expressed in the Laplace transform domain as

$$Q^*\left(X_m, X_{m+1}, ds\right) = L\{Q(X_m, X_{m+1}, dt)\} \tag{16}$$

Section 5.2 below describes how to compute the values of $Q^*\left(X_m, X_{m+1}, ds\right)$. Using $Q^*\left(X_m, X_{m+1}, ds\right)$ several characteristics of the output process (X, T) can be obtained (Disney and Kiessler, 1987). For example, the state transition matrix P_D of the underlying Markov chain *X* embedded at times $\tau_m, m \in N$ is obained by setting $s = 0$ in equation 16 above, i.e.:

$$\begin{aligned} P_D\left(X_m, X_{m+1}\right) &= P\left(X_{m+1} = \left(N_{m+1}, J_{1m+1}, J_{2m+1}\right) \mid X_m = \left(N_m, J_{1m}, J_{2m}\right)\right) \\ &= Q^*\left(X_m, X_{m+1}, ds\right)|_{s=0} \end{aligned} \tag{17}$$

Let $\Pi = \{\Pi(X_k) : X_k \in S_D\}$, where $\Pi\left(X_k\right)$ is the steady state probability that the fork/join station is in state X_k at a departure instant. The stationary probability vector Π of the underlying Markov chain is obtained as follows:

$$\Pi = \Pi P_D \tag{18}$$

$$\sum_{k \in S_D} \Pi\left(X_k\right) = 1 \tag{19}$$

The system of equations 18 and 19 are solved to obtain Π. Section 5.3 describes how to compute $G_D(t)$, the cumulative distribution function of the inter-departure times using Π.

The computation of P_D and hence Π is the next step in estimating the cumulative distribution function $G_D(t)$ of the inter-departure times. In the next section, we construct P_D using the semi-Markov kernel *Q* of the output process (*X*, *T*) and use P_D to determine Π.

5.2 Construction of Semi-Markov Kernel Q and State Transition Probability Matrix P_D

To construct P_D, we note that between two successive departure states X_m and X_{m+1}, the fork/join station visits a finite sequence of intermediate states $Z_k, k = 1,2,...$ where $Z_k \in S_Q$. Each such sequence corresponds to sample paths leading the fork/join station from departure state X_m at time τ_m to state X_{m+1} at time τ_{m+1}. Let $SS\left(X_m, X_{m+1}\right)$ be the set containing all such sequences of intermediate states, and $\left|SS\left(X_m, X_{m+1}\right)\right|$ denote the number of sequences for the pair $\left(X_m, X_{m+1}\right)$. Let $SS_j\left(X_m, X_{m+1}\right) \in SS\left(X_m, X_{m+1}\right)$ be the j^{th} sequence in this set. Then

$SS_j(X_m, X_{m+1})$ can be written as the ordered sequence $(Z_0, Z_1, ..., Z_{l(j)-1}, Z_{l(j)})$ where $l(j)$ is the length of the sequence $SS_j(X_m, X_{m+1}), Z_0 \equiv X_m, Z_{l(j)} \equiv X_{m+1}$ and $Z_k, k = 1,2,..., l(j)-1$ correspond to the intermediate states. Additionally, for a given sample path that contains the sequence of states, let $t_0 \le t_1 \le ... \le t_{l(j)-1} \le t_{l(j)}$ be the times when the fork/join station transitions to states $Z_0, Z_1, ..., Z_{l(j)-1}, Z_{l(j)}$ respectively. Note $t_0 \equiv \tau_m$ and $t_{l(j)} \equiv \tau_{m+1}$. Figure 5 illustrates sample paths for each of three different sequences in $SS(X_m, X_{m+1})$, as well as the sequence of intermediate transitions in a sample path for $SS_1(X_m, X_{m+1})$.

Since the sequences of intermediate states $SS_j(X_m, X_{m+1})$ are mutually exclusive and equally likely, we have

$$Q^*(X_m, X_{m+1}, ds) \quad = \quad \sum_{j=1}^{|SS(X_{m+1}, X_m)|} \left(\prod_{k=1}^{l(j)} Q^*_{j,k}(Z_{k-1}, Z_k, ds) \right) \tag{20}$$

$$P_D(X_m, X_{m+1}) \quad = \quad \sum_{j=1}^{|SS(X_{m+1}, X_m)|} \left(\prod_{k=1}^{l(j)} Q^*_{j,k}(Z_{k-1}, Z_k, ds)|_{s=0} \right) \tag{21}$$

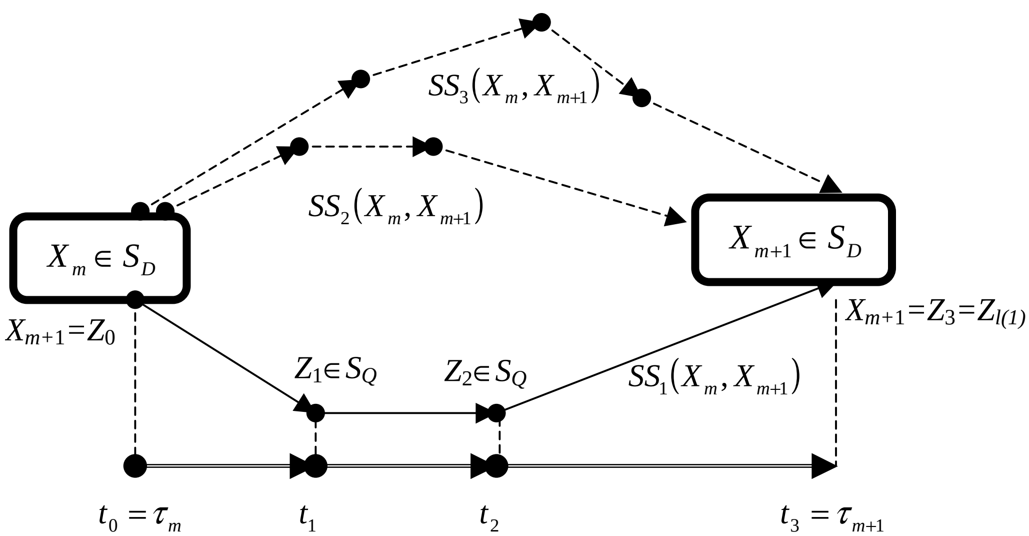

Figure 5. Possible sample paths between departures

In order to compute $P_D(X_m, X_{m+1})$ using equation 21 above, we need to identify all the sequences $SS_j(X_m, X_{m+1}) \in SS(X_m, X_{m+1})$ and then compute $Q^*_{j,k}(Z_{k-1}, Z_k, ds)|_{s=0}$ for each $k >$ 0, and each $Z_k \in SS_j(X_m, X_{m+1}) \in SS(X_m, X_{m+1})$. This may appear to be a cumbersome task. However, because arrivals to buffers B_1 and B_2 are composed of exponential phases, the semi-Markov kernel Q and hence, the transition matrix P_D has a special structure. In other words, depending upon the values of N, J_1 and J_2 of states X_m, and X_{m+1} the non-zero portion of P_D can be partitioned into 14 regions. We exploit this special structure to identify the set of sequences $SS(X_m, X_{m+1})$ for each pair (X_m, X_{m+1}) and also to compute $Q^*_{j,k}(Z_{k-1}, Z_k, ds)|_{s=0}$ for each $Z_k \in SS_j(X_m, X_{m+1})$. Figure 6 illustrates these regions (labeled 1 through 14) in P_D using the example of a fork/join station with two input streams and with K_1=K_2=5.

Regions 1-11 correspond to transitions between states X_m and X_{m+1} in which one of the buffers (B_1 or B_2) is non-empty at time instants τ_m and τ_{m+1}. Regions 12-14 correspond to the states when both the buffers B_1 and B_2 were empty at time instants τ_m and τ_{m+1}. Transitions between states in a region share similarity in the structure of the possible sample paths. We exploit this similarity while deriving expressions for $P_D(X_m, X_{m+1})$. The Appendix lists the formal definitions as well as the expressions for $P_D(X_m, X_{m+1})$ for each pair (X_m, X_{m+1}). Here, we illustrate the procedure for deriving these expressions using Region 11 as an example.

$P_D(X_m, X_{m+1})$		X_{m+1}																
		-4,1,1	-3,1,2	-3,1,1	-2,1,2	-2,1,1	-1,1,2	-1,1,1	0,1,2	0,1,1	0,2,1	1,1,1	1,2,1	2,1,1	2,2,1	3,1,1	3,2,1	4,1,1
	-4,1,1	5	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	-3,1,2	8	6	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	-3,1,1	5	7	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0
	-2,1,2	8	6	4	2	4	2	0	0	0	0	0	0	0	0	0	0	0
	-2,1,1	5	7	1	3	1	3	1	0	0	0	0	0	0	0	0	0	0
	-1,1,2	8	6	4	2	4	2	4	2	0	0	0	0	0	0	0	0	0
	-1,1,1	5	7	1	3	1	3	1	3	1	0	0	0	0	0	0	0	0
	0,1,2	8	6	4	2	4	2	4	13	13	13	9	11	9	11	9	11	10
X_m	0,1,1	5	7	1	3	1	3	1	14	14	14	1	3	1	3	1	7	5
	0,2,1	10	11	9	11	9	11	9	12	12	12	4	2	4	2	4	6	8
	1,1,1	0	0	0	0	0	0	0	0	1	3	1	3	1	3	1	7	5
	1,2,1	0	0	0	0	0	0	0	0	0	2	4	2	4	2	4	6	8
	2,1,1	0	0	0	0	0	0	0	0	0	0	1	3	1	3	1	7	5
	2,2,1	0	0	0	0	0	0	0	0	0	0	0	2	4	2	4	6	8
	3,1,1	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	7	5
	3,2,1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	6	8
	4,1,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7	5

Figure 6. Regions in the semi-Markov kernel Q and state transition matrix P_D

In Region 11 we have two sub-regions:

$$X_m = (0,1,2),\ X_{m+1} = (N_{m+1},2,1) \text{ for } 0 < N_{m+1} \le K_1 - 2 \text{ and}$$
$$X_m = (0,2,1),\ X_{m+1} = (N_{m+1},1,2) \text{ for } 0 > N_{m+1} \ge -K_2 + 2$$

Consider $P_D(X_m, X_{m+1})$ for the first sub-region where $X_m = (0,1,2)$ and $X_{m+1} = (N_{m+1},2,1)$ for $0 < N_{m+1} \le K_1 - 2$. In this case, $P_D(X_m, X_{m+1})$ is the probability of the event that in buffer B_1, exactly $(N_{m+1}+1)$ arrivals and only phase 1 of a two-phase $(N_{m+1}+2)^{nd}$ arrival are completed before the completion of phase 2 of the arrival in buffer B_2. Since arrivals to buffers B_1 and B_2 are independent and composed of exponential phases, for (X_m, X_{m+1}) we can write:

$P_D(X_m, X_{m+1}) =$
P(Exactly $(N_{m+1}+1)$ arrivals to B_1 completes before phase 2 of arrival to B_2 completes) $\times$
P(Phase 1 of a two-phase $(N_{m+1}+2)^{nd}$ arrival to buffer B_1 completes before phase 2 of the arrival to B_2 completes) $\times$
P(Phase 2 of the arrival to B_2 completes before phase 2 of $(N_{m+1}+2)^{nd}$ arrival to buffer B_1)
(22)

Noting that each of the $(N_{m+1}+1)$ arrivals in B_1 could be composed of one or two exponential phases, and thus there are $2^{(N_{m+1}+1)}$ possible sequences in $SS(X_m, X_{m+1})$ equation 22 leads to the following expression for $P_D(X_m, X_{m+1})$:

$$P_D(X_m, X_{m+1}) = \left[\left(\frac{(1-p_1)\mu_{11}}{\mu_{11}+\mu_{22}}\right)+\left(\frac{p_1\mu_{11}}{\mu_{11}+\mu_{22}}\right)\left(\frac{\mu_{12}}{\mu_{12}+\mu_{22}}\right)\right]^{N_{m+1}+1}\left[\frac{p_1\mu_{11}}{\mu_{11}+\mu_{22}}\right]\left[\frac{\mu_{22}}{\mu_{12}+\mu_{22}}\right] \qquad (23)$$

Using arguments similar to the above it can be shown that for this sub-region:

$$Q^*(X_m, X_{m+1}, ds) =$$
$$\left[\left(\frac{(1-p_1)\mu_{11}}{\mu_{11}+\mu_{22}+s}\right)+\left(\frac{p_1\mu_{11}}{\mu_{11}+\mu_{22}+s}\right)\left(\frac{\mu_{12}}{\mu_{12}+\mu_{22}+s}\right)\right]^{N_{m+1}+1}\left[\frac{p_1\mu_{11}}{\mu_{11}+\mu_{22}+s}\right]\left[\frac{\mu_{22}}{\mu_{12}+\mu_{22}+s}\right] \qquad (24)$$

From $Q^*(X_m, X_{m+1}, ds)$, the elements of the semi-Markov kernel Q corresponding to this sub-region can be obtained.

Using an approach similar to that described above, we can derive the expressions for $P_D(X_m, X_{m+1})$ and $Q^*(X_m, X_{m+1}, ds)$ when $X_m = (0,2,1)$ and $X_{m+1} = (N_{m+1},1,2)$ for $0 > N_{m+1} \ge -K_2 + 2$ in Region 11. The expressions for $P_D(X_m, X_{m+1})$ and $Q^*(X_m, X_{m+1}, ds)$ for each other pair (X_m, X_{m+1}) are derived in a similar manner. The full set of expressions for the elements of P_D is listed in the Appendix.

5.3 Marginal Distribution of Inter-departure Times

Using P_D derived in the previous section, we solve the system of equations 18 and 19 for the stationary probability vector Π. We use Π to obtain the marginal distribution of the inter-departure times. Note that since P_D is independent of *m*, the inter-departure times are identically distributed. Let $G_D(t)$ be the distribution function of the inter-departure times i.e., $G_D(t) = P\{T_m \le t\}$ for any *m*. $G_D(t)$ can be written in terms of Π and the distribution function for inter-arrival times at input buffers 1 and 2, namely, $G_j(t) = 1 - C_{j1}e^{-\mu_{j1}t} - C_{j2}e^{-\mu_{j2}t}$, $j = 1,2$, as follows:

$$\begin{aligned} G_D(t) &= \Pi_1\, G_2(t) + \Pi_2\, G_1(t) + \Pi(0,1,1)\, G_1(t)\, G_2(t) + \\ &\quad \Pi(0,1,2) G_1(t)(1 - e^{-\mu_{22}t}) + \Pi(0,2,1) G_2(t)(1 - e^{-\mu_{12}t}) \end{aligned}$$

where

$$\begin{aligned} \Pi_1 &= \sum_{S_{1D}} \Pi(X_m) \\ \Pi_2 &= \sum_{S_{2D}} \Pi(X_m) \\ S_D &= S_{1D} \cup S_{2D} \cup \{(0,1,2), (0,1,1), (0,2,1)\} \\ S_{1D} &= \{(n, j_1, 1) : 0 < n \le K_1 - 2;\ j_1 = 1,2\} \cup \{(K_1 - 1,1,1)\} \\ S_{2D} &= \{(n,1, j_2) : -K_2 + 2 \le n < 0;\ j_2 = 1,2\} \cup \{(-K_2 + 1,1,1)\} \end{aligned} \tag{25}$$

Π_1: The steady state probability that a departure results in a non-empty buffer B_1.
Π_2: The steady state probability that a departure results in a non-empty buffer B_2.

From $G_D(t)$, we obtain the mean inter-departure times, $\overline{D}\left(= 1/\chi_D\right)$.

$$\begin{aligned} \overline{D} &= \Pi_1\left(\frac{C_{21}}{\mu_{21}} + \frac{C_{22}}{\mu_{22}}\right) + \Pi_2\left(\frac{C_{11}}{\mu_{11}} + \frac{C_{12}}{\mu_{12}}\right) + \Pi(0,1,2)\left[\frac{1}{\mu_{22}} + \frac{C_{11}\mu_{22}}{\mu_{11}(\mu_{11}+\mu_{22})} + \frac{C_{12}\mu_{22}}{\mu_{12}(\mu_{12}+\mu_{22})}\right] + \\ &\Pi(0,1,1)\left[\left(\frac{C_{11}}{\mu_{11}} + \frac{C_{12}}{\mu_{12}} + \frac{C_{21}}{\mu_{21}} + \frac{C_{22}}{\mu_{22}}\right) - \left(\frac{C_{11}C_{21}}{\mu_{11}+\mu_{21}} + \frac{C_{11}C_{22}}{\mu_{11}+\mu_{22}} + \frac{C_{12}C_{21}}{\mu_{12}+\mu_{21}} + \frac{C_{12}C_{22}}{\mu_{12}+\mu_{22}}\right)\right] + \\ &\Pi(0,2,1)\left[\frac{1}{\mu_{12}} + \frac{C_{21}\mu_{12}}{\mu_{21}(\mu_{21}+\mu_{12})} + \frac{C_{22}\mu_{12}}{\mu_{22}(\mu_{22}+\mu_{12})}\right] \end{aligned} \tag{26}$$

Similarly, from $G_D(t)$, we can derive expressions for the second moment ED^2 of inter-departure times and hence c_D^2, the squared coefficient of variation (SCV) of inter-departure times.

Note that the expression for the marginal distribution $G_D(t)$ is also given by $G_D(t) = \Pi\, Q(t)U$, where *U* is a column vector with all elements equal to 1. Furthermore, the joint distribution of *k* successive inter-departure times from the fork/join station, $P\{T_m \le t_1, T_{m+1} \le t_2, ..., T_{m+k} \le t_k\}$ is equal to $\Pi\, Q(t_1)\, Q(t_2) \ldots Q(t_k) U$ for any m (Disney and Kiessler, 1987).

6 Numerical Results

In this section we present some numerical examples to illustrate the usefulness of our analysis. First, we compare the results of our analysis of a fork/join station in a network having servers with two-phase Coxian service distributions against the results that assume servers with exponential service distributions. Second, we study the sensitivity of performance measures such as throughput, mean queue lengths, and variability of average inter-departure times to the parameters of the input processes and to the customer populations. In the latter case we investigate the impact of (1) mean rates of the input processes (λ_1, λ_2), (2) different SCVs of the input processes (c_1^2, c_2^2), and (3) network populations K_1 and K_2 on performance measures such as throughput rate, average queue lengths at the input buffers and SCV of inter-departure times.

Figure 7 compares the values of K_1 and K_2 required to obtain a target throughput from a fork/join station with input processes having inter-arrival times with the same mean but with different SCVs. An application where such insights would be useful is closed loop fabrication assembly systems. In such systems, it would be necessary to decide the number of fixed pallets in the networks supplying the fork/join station so as to guarantee a required level of throughput. As seen in Figure 7, the throughput increases monotonically with K_1 and K_2. However, as the SCVs of the input processes increase, the values of K_1 and K_2 required to obtain a given throughput increase significantly. As discussed in Kamath *et al.* (1988), and Buzacott and Shanthikumar (1993), in practical manufacturing systems we could find SCVs ranging from 0 to 4.0.

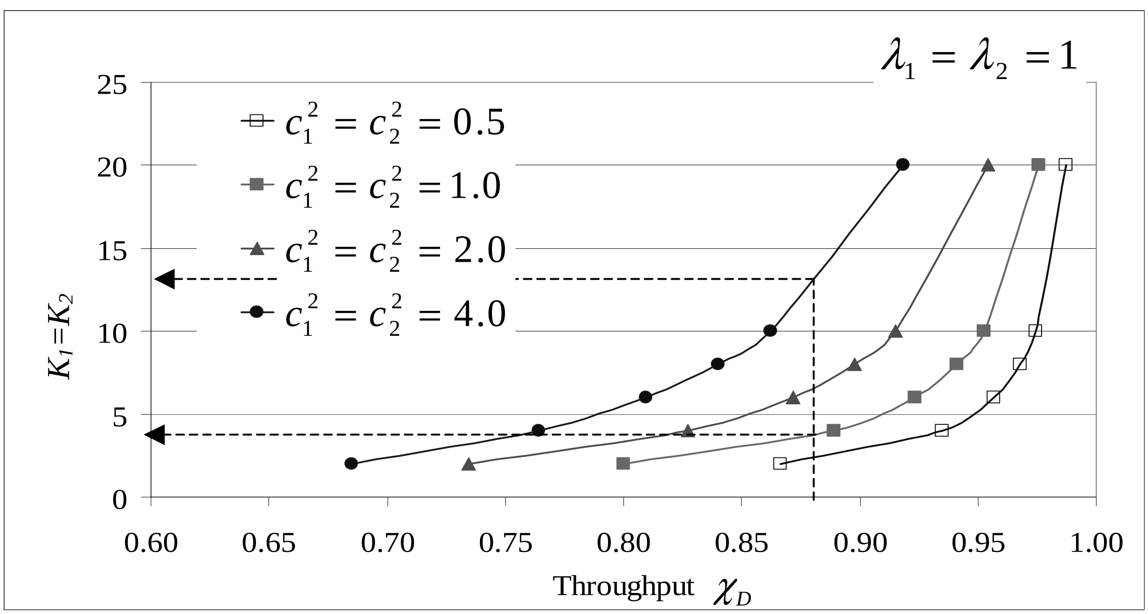

Figure 7. Impact of SCV on K_1 and K_2 required to obtain a given throughput

For example, for a throughput requirement of 0.88, as SCV increases from 0.5 to 4.0 the required values of K_1 and K_2 more than triple. For closed loop fabrication assembly systems, this would imply significant investment in pallets to buffer against variability of input processes. Conversely, ignoring the impact of variability in the input processes (for example, by using the number of pallets needed for the case of Poisson input processes) would result in significantly lower throughput than the target value if the SCV is actually equal to 4.0. This implies that analyzing fork/join stations for inputs more general than the Poisson process has important practical implications.

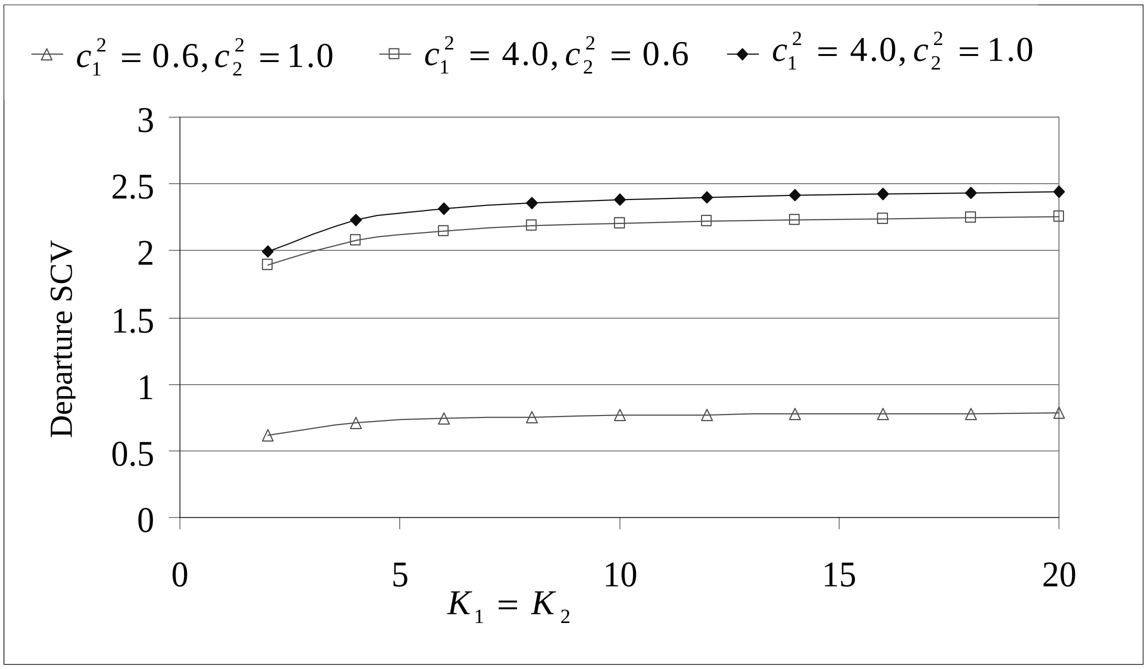

Figure 8. Impact of K_1, K_2, and input SCVs on departure SCVs ($\lambda_1 = \lambda_2 = 1$)

Figure 8 shows the SCV of the inter-departure times for a fork/join station with input processes having rates equal to 1 but different SCVs. As the values of K_1 and K_2 increase, the SCV of the inter-departure times tends to the average of the SCVs of the input processes. Therefore, while additional pallets help improve the throughput performance from the fork/join station, they would not help in reducing the variability of the output process.

In most applications one would expect that, through reasonable initial design of the system, the rates of the input processes at the fork/join synchronization station would be almost equal. However, in reality, imbalances in arrival rates could occur and it is useful to predict the performance impact of such imbalances. Figures 9, 10, and 11 indicate that, for certain parameter settings, throughput, mean queue length, and SCV of inter-departure times are very sensitive to imbalance in input rates.

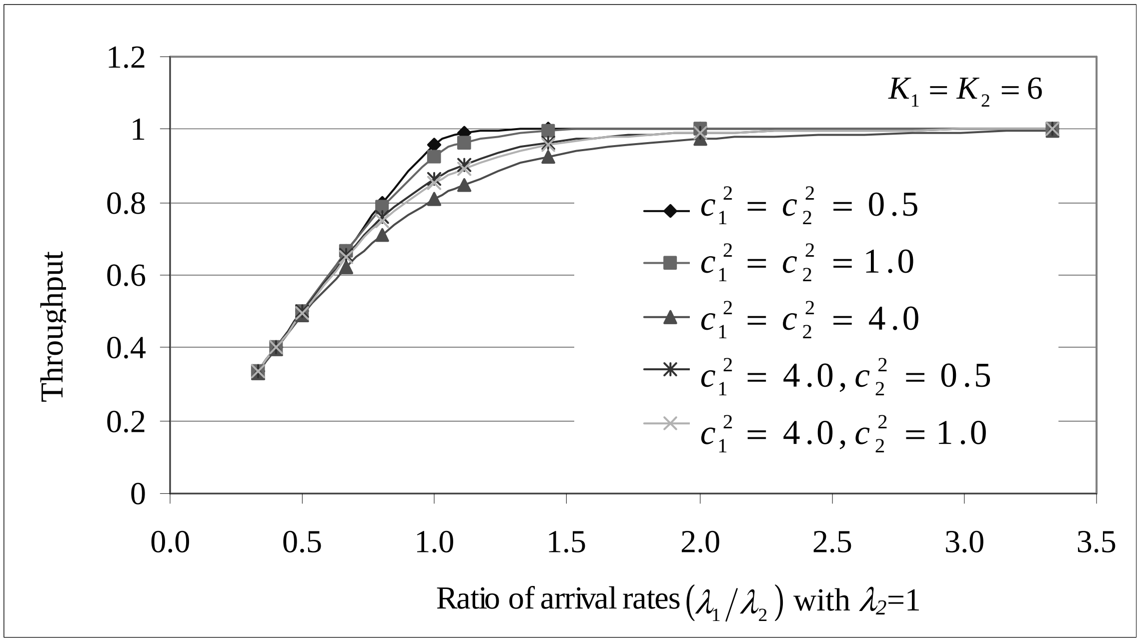

Figure 9. Impact of imbalanced input rates on throughput

In Figure 9 we see that, as would be expected, the throughput from the fork/join is primarily influenced by the arrival rate of the slower of the two input processes. In Figure 10 we see that substantial queues are observed at the buffer of the input process with a higher rate of arrivals.

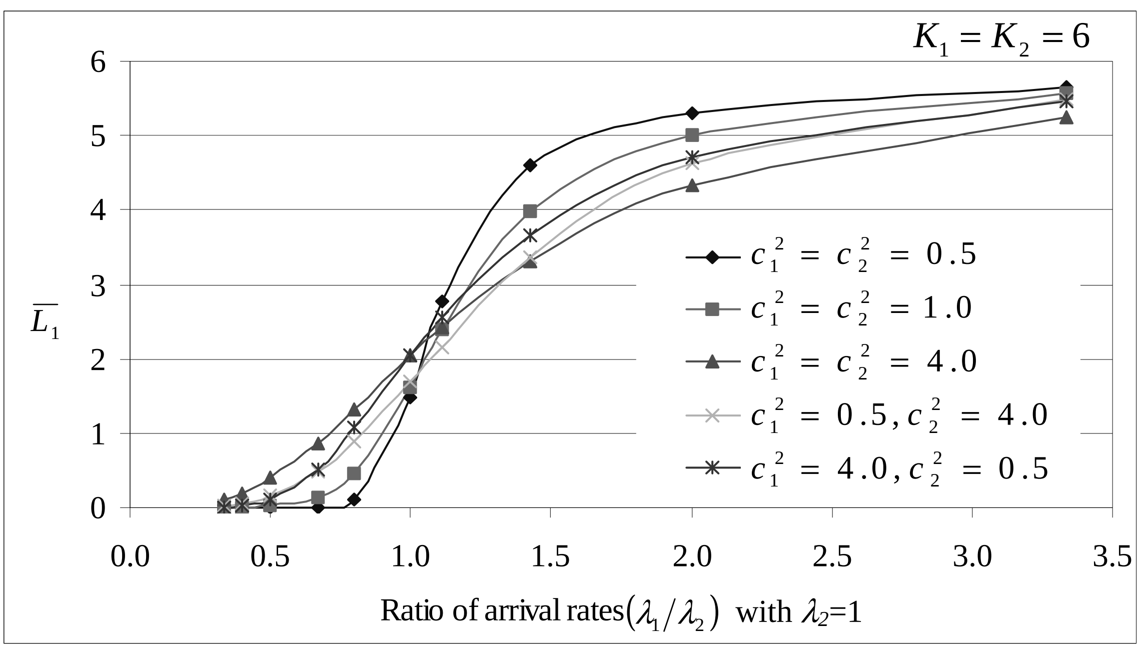

Figure 10. Impact of imbalanced input rates on mean queue length at buffer B_1

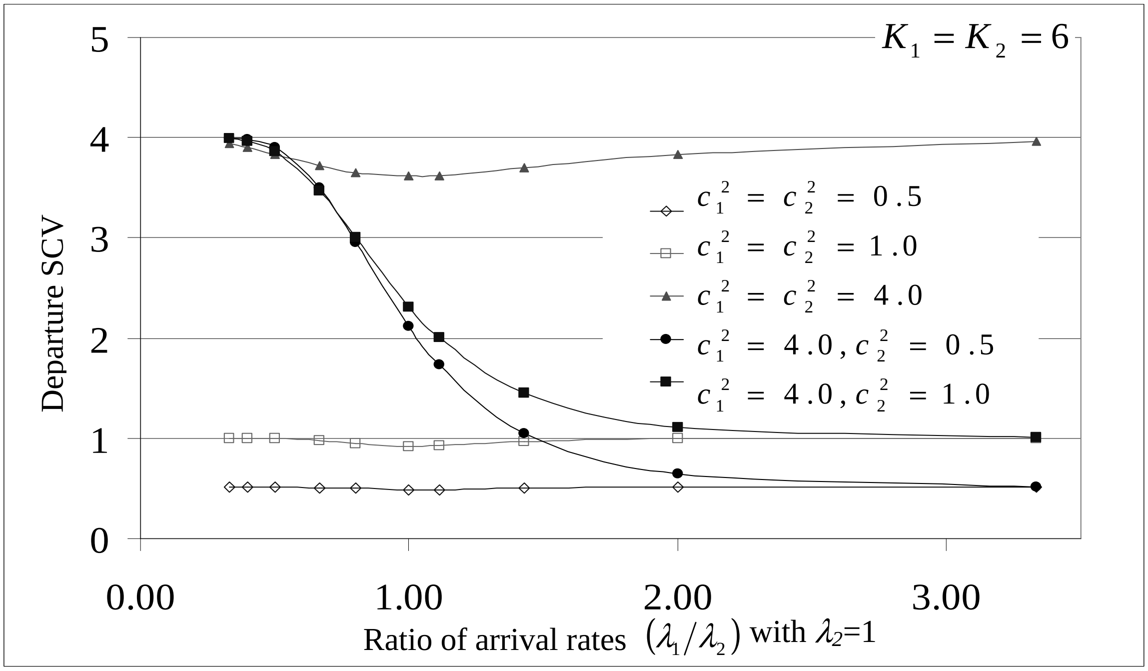

Figure 11. Impact of imbalanced input rates on SCV of inter-departure times

In Figure 11 we see that the SCV of the departure process from the fork/join station is influenced by the SCV of the slower of the two input processes. However, the magnitude of the influence depends on the ratio of the two arrival rates.

Finally, from all the figures it is clear that the SCVs of the input processes have a significant impact on the performance of the fork/join station primarily when the input processes have nearly equal rates. As discussed earlier, this would often be the case by design, hence predicting the impact of SCVs seems important.

Next, we discuss a typical scenario where the analysis presented in this paper could be used to make useful design decisions. We consider a fork/join station with input processes having different SCVs and marginally imbalanced inter-arrival times. The input process to buffer B_1 has inter-arrival times with mean 1.1 $(\lambda_1 = 0.91)$ and SCV equal to 4.0 while input process to buffer B_2 has inter-arrival times with mean 0.9 $(\lambda_2 = 1.11)$ and SCV equal to 0.5. For a given target throughput, we consider the impact of choosing different values of K_1 and K_2 on the output process. For a closed loop fabrication assembly systems, this would translate into deciding whether to invest in additional pallets of one type or the other. For example, Figure 12 indicates that the same target throughput of 0.88 can be obtained by setting $K_1 = K_2 = 10$ or by setting $K_1 = 18$ and $K_2 = 2$. However, as seen from Figure 13, setting $K_1 = K_2 = 10$ results in higher variability of the inter-departure times. This could be detrimental to the downstream work centers, or could require larger buffers downstream. Even setting $K_1 = 18$ and leaving $K_2 = 10$ would not reduce the variability of inter-departure times. To decrease the SCV of the inter-departure times, one would have to not only to set a *high* value of K_1 but also a *low* value of K_2. This large imbalance in K_1 and K_2 setting may not be intuitive to system designers, yet our model points out its benefits.

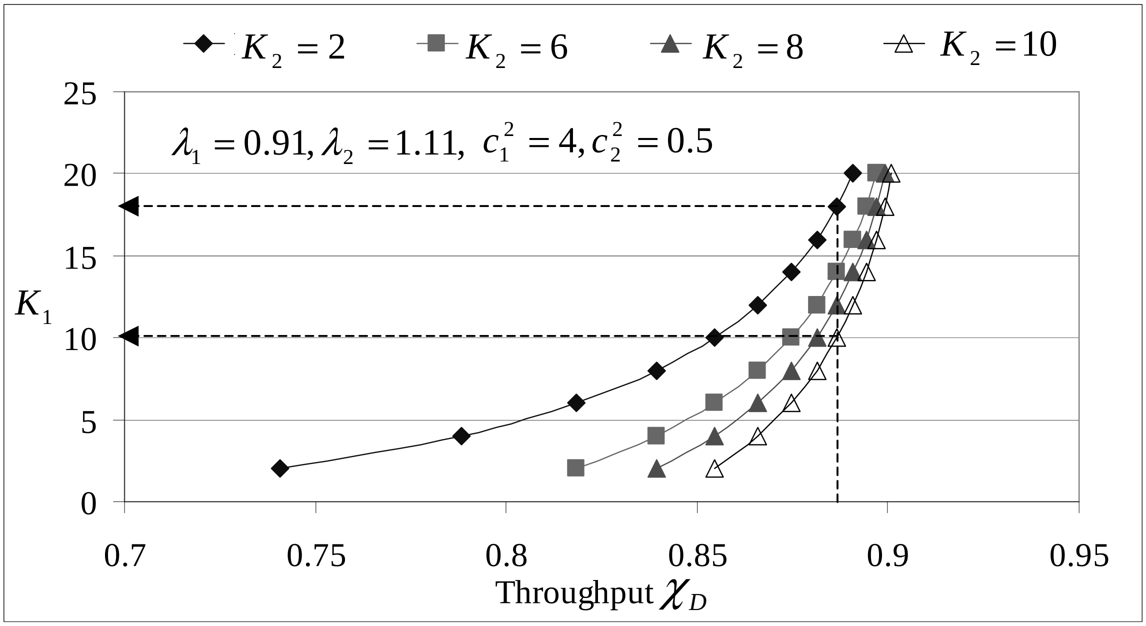

Figure 12. K_1 and K_2 required for a given throughput requirement

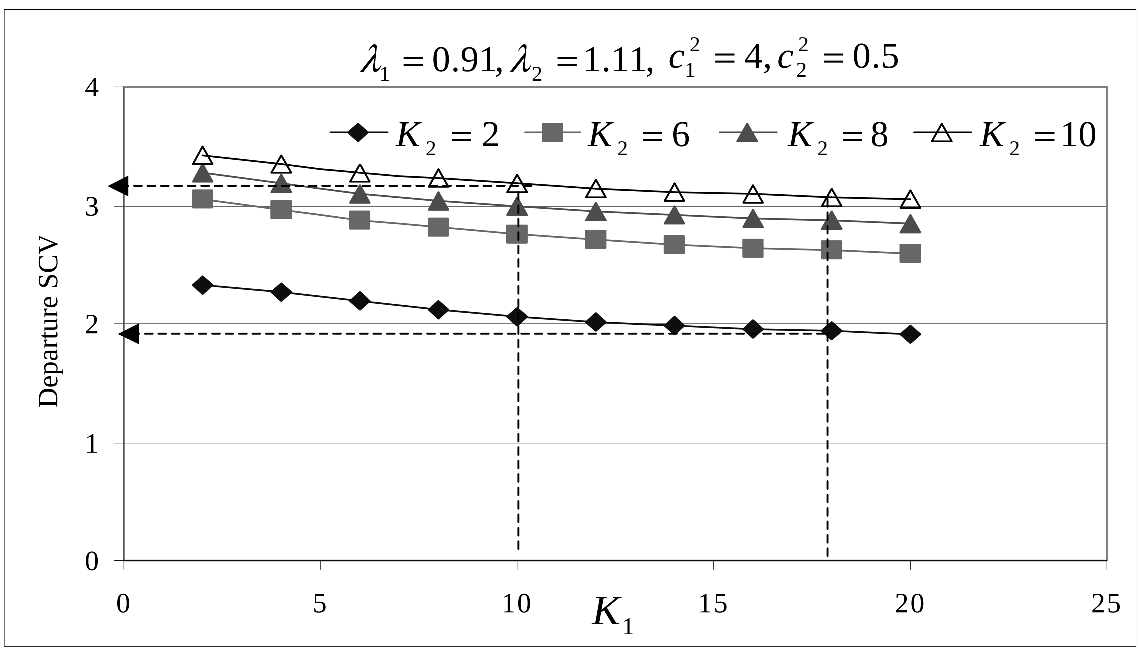

Figure 13. Impact of K_1 and K_2 on SCV of inter-departure times

Before we conclude, we comment on the computational complexity of our analysis. The transition probability matrix used for computing the queue length distribution has a size $4(K_1+K_2)$. For the departure process we consider the Markov chain embedded at departure instants. By analyzing the corresponding transition matrix of size $2(K_1+K_2)$-3, the throughput and marginal distribution of inter-departure times from the fork/join station are computed. The computation effort in constructing this transition matrix is further simplified by the fact that regardless of the parameters of the Coxian distributions and the values of K_1 and K_2, the non-zero portion of the matrix can be partitioned into 14 regions. The computational time required for estimating the throughput, distributions of queue lengths and inter-departure times was less than 3 seconds on a Pentium III, 300MHz personal computer while using MATLAB.

Next, we contrast the computational complexity of our analysis with similar works in the literature. Som *et al.* (1994) and Takahashi *et al.* (1996) analyze a fork/join station with Poisson inputs from infinite populations and finite buffers. In their analysis they assume that arrivals that find the buffers full are lost. The computations in these works involve transition matrices of size (K_1+K_2+1) and (K_1+K_2-1) for the queue length and departure processes respectively. Additionally, because of the memory-less property of the exponential distribution, the analysis presented in Som *et al.* (1994) and Takahashi *et al.* (1996) is also applicable for a fork/join stations where the inputs to each buffer is from a closed queuing network with an exponential server. Our analysis is a first step towards analyzing fork/join stations with arrivals from finite populations having inter-arrival times with distributions that are more general than the exponential. The choice of two-phase Coxian distribution allows us to model input processes with a wide range of mean $(0,\infty)$ and squared coefficient of variation, SCV $[0.5,\infty)$ and our computations involve transition matrices of size $4(K_1+K_2)$ and $2(K_1+K_2)$-3 for the queue length and departure processes respectively. Takahashi *et al*. (2000), and Takahashi and Takahashi (2000) analyze fork/join stations with inter-arrival times having phase type distributions (with p phases) and arrivals from *infinite* populations. Their computations involve transition matrices of size $p^2(K_1+K_2+1)$ and $p^2(K_1+K_2-1)$ for the queue length and departure processes respectively. While a similar analysis for the case of finite populations would be very useful, a very common practical motivation for considering more general input processes is to study the impact of variability in the input processes in addition to those of the means and analyze their impact on decisions regarding systems design. Our analysis for the case of Coxian inputs permits such a study without significant increase in computational effort.

7 Conclusion

Models for analyzing fork/join stations can be useful in the analysis of practical manufacturing and computer systems, as well as for developing efficient decomposition methods for queuing network models of such systems. This paper analyzes a fork/join station in a closed queuing network with inputs from servers having two-phase Coxian service distributions. Using an exact analysis, we show that when the Coxian servers governing the input processes have almost equal rates, variabilities in the arrival processes can have significant impact on throughput, queue lengths and variability in the inter-departure times. Such insights are not available via models that assume exponential interarrival times, since variability cannot be modeled separately from the mean arrival rates. Our choice of the two-phase Coxian distribution allows us to analyze more general input processes without much added computational complexity.

We also study the sensitivity of throughput, queue lengths and variability in the inter-departure times for a wide range of input parameters and network populations. In some cases our results point out behavior that would not be intuitively obvious, and thus our model could provide valuable insights to system designers. The insights here could be used directly in the design of simple systems containing a fork/join station. In addition, when the fork/join station is part of a larger network, information about the inter-departure times could help to characterize the output process which might, in turn, be the input process for other stations in the network.

Finally, we note that the analysis here is also useful in extending decomposition methods for performance analysis of queuing networks. Such methods require efficient "two-moment" approximations to characterize the performance of stations in the network. However, such approximations were previously not available for fork/join stations. The analysis and insights in this paper have been used in the development and validation of two-moment approximations for fork/join stations (Krishnamurthy *et al.*, 2002). These approximations are enabling the extension of decomposition methods to complex queuing networks with fork/join stations (Krishnamurthy, 2002).

Appendix

We list the expressions for $P_D\left(X_m, X_{m+1}\right)$ for each pair $\left(X_m, X_{m+1}\right)$ in Table A below. To simplify the notation in the table, we define the following quantities:

$$a_1 = \frac{(1-p_1)\mu_{11}}{\mu_{11}+\mu_{21}} \qquad \tilde{a}_1 = \frac{(1-p_1)\mu_{11}}{\mu_{11}+\mu_{22}}$$

$$b_1 = \frac{p_1\mu_{11}}{\mu_{11}+\mu_{21}} \qquad \tilde{b}_1 = \frac{p_1\mu_{11}}{\mu_{11}+\mu_{22}}$$

$$c_1 = \frac{\mu_{12}}{\mu_{12}+\mu_{21}} \qquad \tilde{c}_1 = \frac{\mu_{12}}{\mu_{12}+\mu_{22}}$$

$$a_2 = \frac{(1-p_2)\mu_{21}}{\mu_{11}+\mu_{21}} \qquad \tilde{a}_2 = \frac{(1-p_2)\mu_{21}}{\mu_{12}+\mu_{21}}$$

$$b_2 = \frac{p_2\mu_{21}}{\mu_{11}+\mu_{21}} \qquad \tilde{b}_2 = \frac{p_2\mu_{21}}{\mu_{12}+\mu_{21}}$$

$$c_2 = \frac{\mu_{22}}{\mu_{12}+\mu_{22}} \qquad \tilde{c}_2 = \frac{\mu_{22}}{\mu_{12}+\mu_{22}}$$

Additionally we define:

$$s_j = a_j + b_j c_j,\ j = 1,2$$

$$\tilde{s}_j = \tilde{a}_j + \tilde{b}_j \tilde{c}_j,\ j = 1,2$$

$$v = \left|N_{m+1} - N_m\right| + 1$$

and

$$H(r_j, v) = \sum_{i=0}^{i=v} r_j^i \text{ for } r_j = \frac{\tilde{s}_j}{s_j},\ j = 1,2$$

Table A. Expressions for $P_D\left(X_m, X_{m+1}\right)$ *for each pair* $\left(X_m, X_{m+1}\right)$

Region	X_m	X_{m+1}	$P\left(X_m, X_{m+1}\right)$
1	$(N_m,1,1)$ for $0 < N_m \le K_1 - 1$	$(N_{m+1},1,1)$ for $N_m - 1 \le N_{m+1} \le K_1 - 2$	$a_2 s_1^{\,v} + b_2 c_2 H(\tilde{s}_1/s_1, v) + b_1 \tilde{c}_1 \tilde{b}_2 c_2 H(\tilde{s}_1/s_1, v\text{-}1)$
	$(N_m,1,1)$ for $N_m = 0$	$(N_{m+1},1,1)$ for $0 < N_{m+1} \le K_1 - 2$	Same as above
	$(N_m,1,1)$ for $0 > N_m \ge -K_2 + 1$	$(N_{m+1},1,1)$ for $N_m + 1 \ge N_{m+1} \ge -K_2 + 2$	$a_1 s_2^{\,v} + b_1 c_1 H(\tilde{s}_2/s_2, v) + \tilde{b}_1 c_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}1)$
	$(N_m,1,1)$ for $N_m = 0$	$(N_{m+1},1,1)$ for $0 > N_{m+1} \ge -K_2 + 2$	Same as above
2	$(N_m,2,1)$ for $0 < N_m \le K_1 - 2$	$(N_{m+1},2,1)$ for $N_m - 1 < N_{m+1} \le K_1 - 3$	$b_1 c_1 \tilde{a}_2 s_1^{\,v-1} + \tilde{b}_1 \tilde{c}_1 \tilde{b}_2 \tilde{c}_2 \tilde{s}_1^{\,v-1}$ $+ \tilde{b}_1 c_1 \tilde{c}_2 \left(b_2 H(\tilde{s}_1/s_1, v\text{-}1) + b_1 \tilde{c}_1 \tilde{b}_2 H(\tilde{s}_1/s_1, v\text{-}2)\right)$
	$(N_m,2,1)$ for $N_m = 0$	$(N_{m+1},2,1)$ for $0 < N_{m+1} \le K_1 - 3$	Same as above
	$(N_m,2,1)$ for $0 < N_m \le K_1 - 2$	$(N_{m+1},2,1)$ for $N_{m+1} = N_m - 1$	$\tilde{s}_2$
	$(N_m,1,2)$ for $0 > N_m \ge -K_2 + 2$	$(N_{m+1},1,2)$ for $N_m + 1 > N_{m+1} \ge -K_2 + 3$	$c_2 b_2 \tilde{a}_1 s_2^{\,v-1} + \tilde{b}_1 \tilde{c}_1 \tilde{b}_2 \tilde{c}_2 \tilde{s}_2^{\,v-1}$ $+ \tilde{c}_1 \tilde{b}_2 c_2 \left(b_1 H(\tilde{s}_2/s_2, v\text{-}1) + \tilde{b}_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}2)\right)$
	$(N_m,1,2)$ for $N_m = 0$	$(N_{m+1},1,2)$ for $0 > N_{m+1} \ge -K_2 + 3$	Same as above
	$(N_m,1,2)$ for $0 > N_m \ge -K_2 + 2$	$(N_{m+1},1,2)$ for $N_{m+1} = N_m + 1$	$\tilde{s}_1$
3	$(N_m,1,1)$ for $0 < N_m \le K_1 - 1$	$(N_{m+1},2,1)$ for $N_m - 1 \le N_{m+1} \le K_1 - 3$	$b_1 \tilde{s}_2 s_1^{\,v} + \tilde{b}_1 \tilde{c}_2 \left(b_2 H(\tilde{s}_1/s_1, v) + b_1 \tilde{c}_1 \tilde{b}_2 H(\tilde{s}_1/s_1, v\text{-}1)\right)$
	$(N_m,1,1)$ for $N_m = 0$	$(N_{m+1},2,1)$ for $0 < N_{m+1} \le K_1 - 3$	Same as above
	$(N_m,1,1)$ for $0 > N_m \ge -K_2 + 1$	$(N_{m+1},1,2)$ for $N_m + 1 \ge N_{m+1} \ge -K_2 + 3$	$b_2 \tilde{s}_1 s_2^{\,v} + \tilde{c}_1 \tilde{b}_2 \left(b_1 H(\tilde{s}_2/s_2, v) + \tilde{b}_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}1)\right)$
	$(N_m,1,1)$ for $N_m = 0$	$(N_{m+1},1,2)$ for $0 > N_{m+1} \ge -K_2 + 3$	Same as above

Table A. Expressions for $P_D(X_m, X_{m+1})$ *for each pair* (X_m, X_{m+1}) *contd.*

Region	X_m	X_{m+1}	$P(X_m, X_{m+1})$
4	$(N_m,2,1)$ for $0 < N_m \le K_1 - 2$	$(N_{m+1},1,1)$ for $N_m \le N_{m+1} \le K_1 - 2$	$c_1 a_2 s_1^{\,v-1} + \tilde{c}_1 \tilde{b}_2 c_2 \tilde{s}_1^{\,v-1} + c_1 c_2 b_2 H(\tilde{s}_1/s_1, v\text{-}1)$ $+ c_1 c_2 \tilde{c}_1 b_1 \tilde{b}_2 H(\tilde{s}_1/s_1, v\text{-}2)$
	$(N_m,2,1)$ for $N_m = 0$	$(N_{m+1},1,1)$ for $0 < N_{m+1} \le K_1 - 2$	Same as above
	$(N_m,1,2)$ for $0 > N_m \ge -K_2 + 2$	$(N_{m+1},1,1)$ for $N_m \ge N_{m+1} \ge -K_2 + 2$	$a_1 c_2 s_2^{\,v-1} + c_1 \tilde{c}_2 \tilde{b}_1 \tilde{s}_2^{\,v-1} +$ $c_1 c_2 b_1 H(\tilde{s}_2/s_2, v\text{-}1) + c_1 c_2 \tilde{c}_2 \tilde{b}_1 b_2 H(\tilde{s}_2/s_2, v\text{-}2)$
	$(N_m,1,2)$ for $N_m = 0$	$(N_{m+1},1,1)$ for $0 > N_{m+1} \ge -K_2 + 2$	Same as above
5	$(N_m,1,1)$ for $0 \le N_m \le K_1 - 1$	$(K_1 - 1,1,1)$	$(1\text{-}b_2) s_1^{\,v} + b_2 H(\tilde{s}_1/s_1, v) + b_1 \tilde{c}_1 \tilde{b}_2 H(\tilde{s}_1/s_1, v\text{-}1)$
	$(N_m,1,1)$ for $0 \ge N_m \ge -K_2 + 1$	$(-K_2 + 1,1,1)$	$(1\text{-}b_1) s_2^{\,v} + b_1 H(\tilde{s}_2/s_2, v) + \tilde{b}_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}1)$
6	$(N_m,2,1)$ for $0 \le N_m \le K_1 - 2$	$(K_1 - 2,2,1)$	$c_1 b_1 \tilde{s}_2 s_1^{\,v-1} + \tilde{b}_1 \tilde{c}_1 \tilde{b}_2 \tilde{c}_2 \tilde{s}_1^{\,v-1}$ $+ \tilde{b}_1 c_1 \tilde{c}_2 \left(b_2 H(\tilde{s}_1/s_1, v\text{-}1) + b_1 \tilde{c}_1 \tilde{b}_2 H(\tilde{s}_1/s_1, v\text{-}2)\right)$
	$(N_m,1,2)$ for $0 \ge N_m \ge -K_2 + 2$	$(-K_2 + 2,1,2)$	$b_2 c_2 \tilde{s}_1 s_2^{\,v-1} + \tilde{b}_1 \tilde{c}_1 \tilde{b}_2 \tilde{c}_2 \tilde{s}_2^{\,v-1} +$ $\tilde{c}_1 \tilde{b}_2 c_2 \left(b_1 H(\tilde{s}_2/s_2, v\text{-}1) + \tilde{b}_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}2)\right)$
7	$(N_m,1,1)$ for $0 \le N_m \le K_1 - 1$	$(K_1 - 2,2,1)$	$b_1 \tilde{s}_2 s_1^{\,v} + \tilde{b}_1 \tilde{c}_2 \left(b_2 H(\tilde{s}_1/s_1, v) + \tilde{c}_1 \tilde{b}_2 b_1 H(\tilde{s}_1/s_1, v\text{-}1)\right)$
	$(N_m,1,1)$ for $0 \ge N_m \ge -K_2 + 1$	$(-K_2 + 2,1,2)$	$b_2 \tilde{s}_1 s_2^{\,v} + \tilde{c}_1 \tilde{b}_2 \left(b_1 H(\tilde{s}_2/s_2, v) + \tilde{b}_1 b_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}1)\right)$
8	$(N_m,2,1)$ for $0 \le N_m \le K_1 - 2$	$(K_1 - 1,1,1)$	$(1\text{-}b_2) c_1 s_1^{\,v-1} + \tilde{c}_1 \tilde{b}_2 \tilde{s}_1^{\,v-1}$ $+ c_1 b_2 H(\tilde{s}_1/s_1, v\text{-}1) + \tilde{c}_1 \tilde{b}_2 b_1 c_1 H(\tilde{s}_1/s_1, v\text{-}2)$
	$(N_m,1,2)$ for $0 \ge N_m \ge -K_2 + 2$	$(-K_2 + 1,1,1)$	$(1\text{-}b_1) c_2 s_2^{\,v-1} + \tilde{b}_1 \tilde{c}_2 \tilde{s}_2^{\,v-1} +$ $b_1 c_2 H(\tilde{s}_2/s_2, v\text{-}1) + \tilde{b}_1 b_2 c_2 \tilde{c}_2 H(\tilde{s}_2/s_2, v\text{-}2)$

Table A. Expressions for $P_D(X_m, X_{m+1})$ *for each pair* (X_m, X_{m+1}) *contd.*

Region	X_m	X_{m+1}	$P(X_m, X_{m+1})$
9	(0,1,2)	$(N_{m+1},1,1)$ for $0 < N_{m+1} \le K_1 - 2$	$c_2 \tilde{s}_1{}^v$
	(0,2,1)	$(N_{m+1},1,1)$ for $0 > N_{m+1} \ge -K_2 + 2$	$c_1 \tilde{s}_2{}^v$
10	(0,1,2)	$(K_1 - 1,1,1)$	$\tilde{s}_1{}^v$
	(0,2,1)	$(-K_2 + 1,1,1)$	$\tilde{s}_2{}^v$
11	(0,1,2)	$(N_{m+1},2,1)$ for $0 < N_{m+1} \le K_1 - 2$	$\tilde{b}_1 \tilde{c}_2 \tilde{s}_1{}^v$
	(0,2,1)	$(N_{m+1},1,2)$ for $0 > N_{m+1} \ge -K_2 + 2$	$\tilde{b}_2 \tilde{c}_1 \tilde{s}_2{}^v$
12	(0,2,1)	(0,2,1)	$b_1 c_1 \tilde{s}_2 + \tilde{b}_1 \tilde{c}_2 \left(\tilde{b}_2 \tilde{c}_1 + b_2 c_1\right)$
	(0,2,1)	(0,1,2)	$\tilde{b}_2 \tilde{c}_1 \tilde{s}_2$
	(0,2,1)	(0,1,1)	$c_1 (s_2 + \tilde{s}_2) + \tilde{b}_2 \tilde{c}_1 c_2$
13	(0,1,2)	(0,1,2)	$\tilde{b}_2 \tilde{c}_1 \left(b_1 c_2 + \tilde{b}_1 \tilde{c}_2\right) + b_2 c_2 \tilde{s}_1$
	(0,1,2)	(0,2,1)	$\tilde{b}_1 \tilde{c}_2 \tilde{s}_1$
	(0,1,2)	(0,1,1)	$c_2 (s_1 + \tilde{s}_1) + \tilde{b}_1 \tilde{c}_2 c_1$
14	(0,1,1)	(0,1,1)	$2 s_1 s_2 + b_1 c_1 \tilde{s}_2 + c_2 b_2 \tilde{s}_1 + \tilde{b}_2 \tilde{c}_1 b_1 c_2 + \tilde{b}_1 c_1 b_2 \tilde{c}_2$
	(0,1,1)	(0,2,1)	$b_1 s_1 \tilde{s}_2 + b_2 \tilde{b}_1 \tilde{c}_2 (s_1 + \tilde{s}_1) + b_1 \tilde{b}_1 \tilde{c}_1 \tilde{b}_2 \tilde{c}_2$
	(0,1,1)	(0,1,2)	$b_2 \tilde{s}_1 s_2 + b_1 \tilde{b}_2 \tilde{c}_1 (s_2 + \tilde{s}_2) + b_2 \tilde{b}_1 \tilde{b}_2 \tilde{c}_1 \tilde{c}_2$

References

Altiok, T., 1996, *Performance Analysis of Manufacturing Systems*, Springer Series in Operations Research, Springer, New York.

Baccelli, F., Massey, W.A., and Towsley, D., 1989, "Acyclic Fork/Join Queuing Networks", *Journal of the ACM*, 36, 615-642.

Bhat, U.N., 1986, "Finite Capacity Assembly like Queues", *Queuing Systems*, 1, 185-101.

Buzacott, J.A., and Shanthikumar, J. G., 1993, *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.

Di Mascolo, M., Frein, Y., and Dallery, Y., 1996, "An Analytical Method for Performance Evaluation of Kanban Controlled Production Systems", *Operations Research*, 44(1), 50-64.

Disney, R.L., and Kiessler, R.L., 1987, *Traffic Processes in Queuing Networks: A Markov Renewal Approach*, Johns Hopkins University Press.

Harrison, J.M., 1973, "Assembly-like Queues", *Journal of Applied Probability*, 10(2), 354-367.

Heidelberger, P. and Trivedi, K. S.,1983, "Queuing Network Models for Parallel Processing with Asynchronous Tasks", *IEEE Transactions on Computers*, C-31(11), 1099-1109.

Hopp, W.J., and Simon, J.T., 1989, "Bounds and Heuristics for Assembly-like Queues", *Queuing Systems*, 4,137-156.

Kamath, M., Suri, R., and Sanders, J.L., 1988, "Analytical Performance Models for Closed – Loop Flexible Assembly Systems", *International Journal of Flexible Manufacturing Systems*, 1(1), 51-84.

Kashyap, B.R.K., 1965, "A Double Ended Queuing System with Limited Waiting Space", *Proceedings of the National Institute of Science (India)*, 31, 559-570.

Krishnamurthy, A., Suri, R. and Vernon, M., 2001, "Analytical Performance Analysis of Kanban Systems Using a New Approximation for Fork/Join Stations", *Proceedings of the Industrial Engineering Research Conference, Dallas, Texas*.

Krishnamurthy, A., 2002, *Analytical Performance Models for Material Control Strategies in Manufacturing Systems*, PhD. Thesis, Department of Industrial Engineering, University of Wisconsin-Madison.

Krishnamurthy, A., Suri, R., and Vernon, M., 2002, "Two-Moment Approximations for Throughput and Mean Queue Length of a Fork/Join Station with General Inputs from Finite Populations", to appear in *Stochastic Modeling and Optimization of Manufacturing Systems and*

Supply Chains, J.G. Shanthikumar, D.D. Yao and W.H.M. Zijm (Eds.), Kluwer International Series in Operations Research and Management Science (in press).

Latouche, G., 1981, “Queues With Paired Customers”, *Journal of Applied Probability*, 18, 684-696.

Liberopoulous, G. and Dallery, Y., 2000, “A Unified Framework for Pull Control Mechanisms in Multistage Manufacturing Systems”, *Annals of Operations Research*, 93, 325-355.

Lipper, E.H., and Sengupta, B., 1986, “Assembly Like Queues with Finite Capacity: Bounds, Asymptotics and Approximations”, *Queuing Systems*, 1, 67-83.

Prabhakar, B., Bambos, N. and Mountford, T.S., 2000, “The Synchronization of Poisson Processes and Queuing Networks with Service and Synchronization Nodes”, *Advances in Applied Probability*, 32, 824-843.

Rao, P.C. and Suri, R., 1994, “Approximate Queuing Network Models of Fabrication/Assembly Systems: Part I-Single Level Systems”, *Production and Operations Management*, 3, 244-275.

Rao, P.C. and Suri, R., 2000, “Performance Analysis of An Assembly Station with Input from Multiple Fabrication Lines”, *Production and Operations Management*, 9(3), 283-302.

Som, P., Wilhelm, W.E., and Disney, R.L., 1994, “Kitting Process in a Stochastic Assembly System”, *Queuing Systems*, 17, 471-490.

Takahashi, M., Osawa, H., and Fujisawa, T., 1996, “A Stochastic Assembly System with Resume Levels”, *Asia-Pacific Journal of Operations Research*, 15, 127-146.

Takahashi, M., Osawa, H., and Fujisawa, T., 2000, “On a Synchronization Queue with Two Finite Buffers”, *Queuing Systems*, 36, 107-123.

Takahashi, M., and Takahashi, Y., T., 2000, “Synchronization Queue with Two MAP Inputs and Finite Buffers”, *Proceedings of the Third International Conference on Matrix Analytical Methods in Stochastic Models, Leuven (Belgium)*, 375-390.

Varki, E., 1999, “Mean Value Technique for Closed Fork-Join Networks”, *Proceedings of ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems, Atlanta, GA*, 103-112.