A New Approach for Analyzing Queuing Models of Material Control Strategies in Manufacturing Systems

Ananth Krishnamurthy and *Rajan Suri*,
Center for Quick Response Manufacturing, University of Wisconsin-Madison,
1513 University Avenue, Madison, WI 53706-1572,
Tel. 608-2620921, Fax: 608-2654017, Email: akrishna@cae.wisc.edu.
and
Mary Vernon,
Department of Computer Sciences, University of Wisconsin-Madison,
1210 W Dayton Street, Madison, WI-53706.

Abstract

This paper discusses a new approach for analyzing queuing models of material control strategies such as kanban, CONWIP and POLCA in manufacturing systems. In practice, these strategies regulate the flow of jobs at various stations of the manufacturing system using a combination of card loops and authorization signals. Queuing models of such systems are composed of closed queuing networks corresponding to the different card loops. Fork/join stations are used to represent the synchronization constraints imposed on jobs, cards and authorization signals. Analyzing these networks involves solving a set of multi-class closed queuing networks with fork/join synchronization stations and stations with general processing times. Since these networks cannot be solved exactly, approximate methods have been investigated. We propose a new approximate approach based on the node decomposition technique. We study the underlying semi-markov processes at each station in the network and use the two moment renewal approximations to describe arrival and departure processes at each station. One of the new contributions here is the development of a two-moment approximation for fork/join stations fed by independent arrival streams. This approximation allows us to better model the synchronization constrains typical of material control strategies. Then, we use constraints on flow of jobs in the closed network to develop closed form expressions linking the arrival and departure process parameters at the different stations. These equations are then solved using an iterative algorithm.

Key words: Closed multi-class queuing networks, Approximation methods, Performance analysis, Fork/join, Manufacturing systems

1. Introduction

The primary purpose of material control strategies such as kanban, CONWIP (CONstant WIP) and POLCA (Paired-cell Overlapping Loops of Cards with Authorization) is to provide an efficient mechanism for determining the way production is triggered and inventory is controlled on the manufacturing shop floor. In practice, these strategies regulate the flow of jobs at various stations of the manufacturing system using a combination of card loops and authorization signals. Each card loop has a finite number of cards and this restricts the total number of jobs in the system at any given time. Jobs move from one station to another only if the necessary cards and/or authorization signals are available. Kanban, CONWIP and POLCA differ in the nature of card loops and authorization signals they use to control production.

Comparisons of the performance of kanban, CONWIP and POLCA strategies can be found in [1]-[7]. These studies seem to indicate that no one system completely dominates the other. More complete exploration of the tradeoffs using analytical models could help us demarcate regions where each system dominates over the other. Queuing models seem to be the most useful for such analysis because they effectively model the stochastic nature of system capacity and its impact on system throughput. Efficient queuing models have very reasonable computational requirements which enables their use as a performance evaluation tool. Queuing models used to analyze material control systems [8]-[12] are usually composed of closed chains that correspond to the different card loops, fork/join stations that represent the synchronization constraints imposed on jobs, cards and authorization signals, and stations with general processing times (see section 2). Exact solutions for these networks are not available, and approximate methods must be used. Previous approaches have used product form approximations wherein the stations whether of the fork/join type or with general service times are replaced by equivalent stations with exponentially distributed load dependent service rates [8][13]-[15].

In this paper, we propose an approximate approach based on a node decomposition technique. We study the underlying semi-markov processes at each station in the network and use two moment renewal approximations to describe arrival and departure processes at each station. We develop a new two-moment approximation for fork/join stations and this allows us to use the node decomposition technique to effectively model material control strategies. Then, using constraints on flow of jobs in the closed network, we develop closed form expressions linking the arrival and departure process parameters at the different stations. These equations are then solved using an iterative algorithm. In this paper, we describe how this approach can be used to analyze the kanban control strategies (see section 3).

The paper is organized as follows. In section 2, we review the queuing model representations of material control strategies such as kanban, CONWIP and POLCA. We discuss the previous techniques used to analyze these queuing models and the challenges encountered in doing so. In section 3, we analyze a single stage kanban system using the node decomposition approach. In section 4 we extend our approach to analyze multi stage kanban systems. In section 5, we briefly outline how the proposed approach can be extended to analyze other control strategies such as CONWIP and POLCA in systems with multiple products. We conclude with section 6 where we provide some qualitative comparisons with a few other approaches discussed in the literature.

2. Queuing Model Representation of Material Control Strategies

Material control strategies typically have been modeled using closed queuing networks with fork/join stations and stations having general processing times. The fork/join stations are used to represent the synchronization constraints wherever present. We use the representation of kanban systems given in [8] and show how the model can be modified to represent multi product CONWIP systems. We will use the kanban representation to explain the proposed analytical solution method.

2.1 Single product kanban systems

Figure 1 models the functioning of a single product kanban system with a number of serial stages, S, equal to three. Stage i denoted by S_i could consist of a series of M_i manufacturing stations. We assume throughout this paper that each manufacturing station is a single server queue. Associated with each product in each stage are a fixed number of kanbans that circulate within a loop. In our single product system, stage S_i has K_i kanbans.

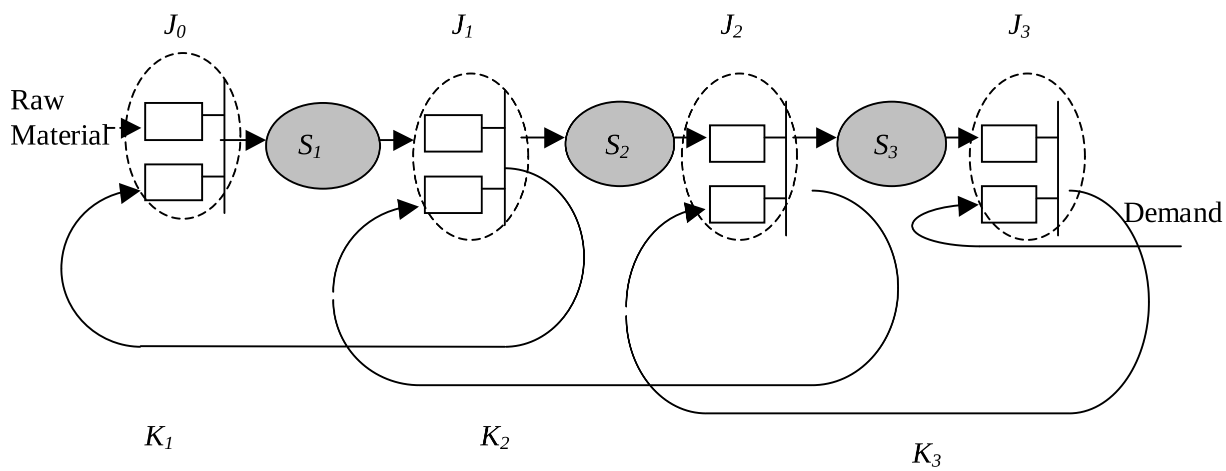


Figure 1. Queueing model representation of single product kanban systems

The kanban control mechanism controls the movement of parts from stage to stage. This control mechanism is modeled by the synchronization stations at the input to the system and at the output buffer at each stage. In the three stage system shown in Figure 1, J_0, …, J_3 are synchronization stations. The kanban control mechanism at stage S_i includes the synchronization stations $J_{i\text{-}1}$ and J_i. These synchronization stations together with the manufacturing stations in stage i will be called subsystem i (SUB_i). The station J_i represents the synchronization between finished parts of stage i and free kanbans from stage i+1. As soon as there is one entity in each of the input queues at J_i, one entity is removed from each of them and added to each of its output queues. In other words, the finished part is transferred to stage i+1 and the released kanban card is returned to the input queue of free kanban cards at the previous synchronization station $J_{i\text{-}1}$. The released kanban transmits the information of the consumption of a part to the upstream synchronization station $J_{i\text{-}1}$. Since each part, either in the manufacturing stage or in the output buffer of subsystem i, must have a stage i kanban card attached to it, the number of cards K_i is an upper bound on the number of parts in stage i.

2.2 Multiproduct kanban systems

Figure 2 shows the queuing network model of a multi-product kanban control system having 3 stages in series producing 2 types of parts. With each stage S_i and each part type r (r=1, 2), is associated $K_{i,,r}$ kanbans. Corresponding to each class r we have synchronization stations, $J_{0,r}$, …, $J_{3,r}$. For each product type r, the kanban control mechanisms between the stages operates exactly in the same way as in the case of the single product system. Therefore, although the different classes of products share the same manufacturing stations, their synchronization mechanisms are independent.

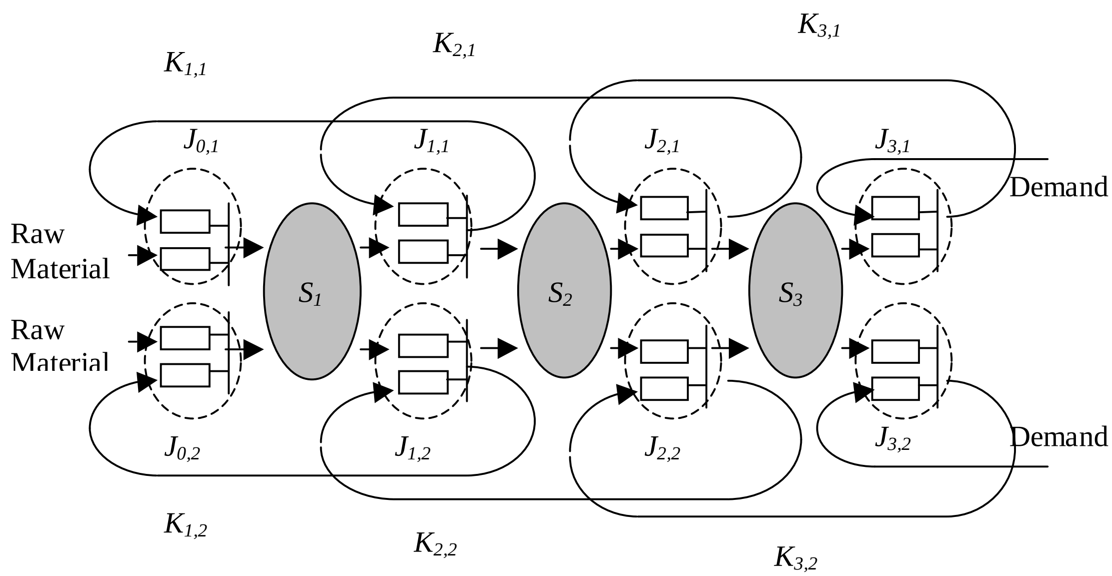


Figure 2. Queueing model representation of multiproduct kanban systems

2.3 Multiproduct CONWIP and POLCA systems

Other control mechanisms such as CONWIP and POLCA differ from the kanban system either in the way the loops are designed and/or in the way the cards are used. For a system producing a single product, the CONWIP system is similar to the kanban system, except that a single card loop controls all the manufacturing stages in the system [4]. In the POLCA system [5], a recently introduced control mechanism, the card loops extend across pairs of stages.

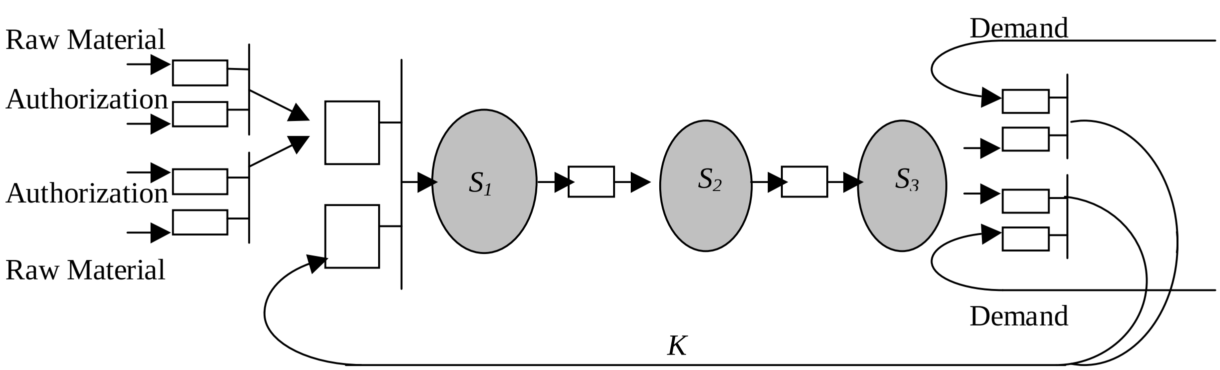


Figure 3. Queueing model representation of multiproduct CONWIP systems

Figure 3 shows the queuing network model of a multi-product CONWIP control system having 3 stages in series and producing 2 types of products. There are 3 major differences between the multiproduct kanban and the CONWIP systems. First, all the three stages are controlled by a single CONWIP loop having a total of K CONWIP cards. Second, these cards are not specific to any product. Third, it uses authorization signals to release parts into the system. At the beginning of the system, raw material of a particular product is eligible to be released into manufacturing only when the corresponding authorization signal is received. Actual release of the raw material to stage S_1 occurs only if a free CONWIP card is present before stage S_1. Once in S_1, jobs move uninterruptedly through S_2 and S_3 and wait in the output buffer of stage 3. Arrival of a customer demand releases a finished job from the output buffer of stage S_3. Simultaneously, the CONWIP card released is sent to the front of the line to allow the entry of a new job into S_1. Therefore, unlike the kanban system, the different classes of products flow through the same synchronization station.

In POLCA systems, the cards are not product specific and the different classes of products flow through the same synchronization station. However, POLCA has overlapping loops covering pairs of stages and the authorization signals are required at the beginning of each loop [5].

2.4 Challenges in performance analysis

Despite the large amount of literature on kanban systems, research in the area of analytical solutions of kanban control systems models has been limited. Several of these efforts have been confined to single product systems with limited complexity. The work of Kimura and Terada [9] uses a discrete time markov chain model to model the kanban system with stages in series. Karmarkar and Kekre [10] describe the behavior of single stage and two stage kanban systems using a continuous time markov chain (CTMC).

Real systems are more complex. They constitute multiple stages and multiple products. Insights from the simple models for kanban systems do not extend to the more complex kanban systems [4]. Additionally, it has been argued that POLCA and CONWIP outperforms kanban for these complex systems [3]-[7]. However, these arguments have not been supported by thorough analysis, the primary reason being that analyzing these systems involves solving a set of multi-class closed queuing networks with fork/join synchronization stations and stations with possibly general processing times. Obtaining the exact solutions for these networks is very difficult.

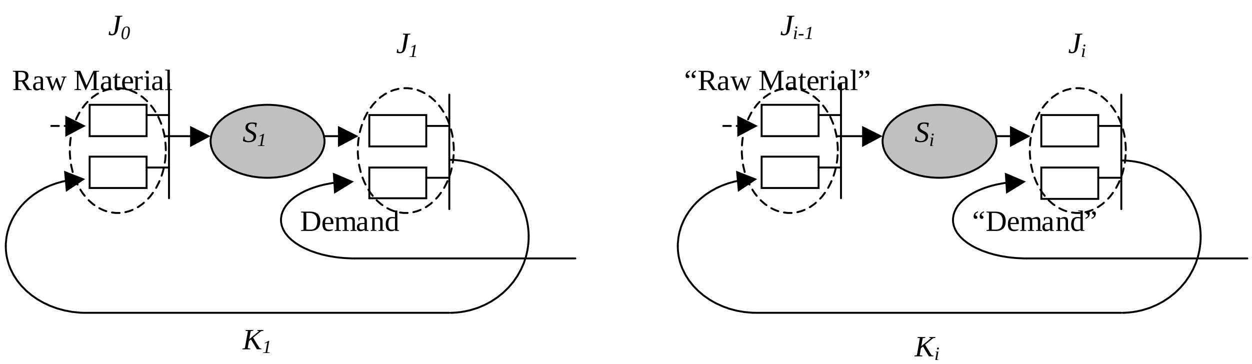
Recently some decomposition and approximation approaches have been developed for analyzing kanban systems. Mitra and Mitrani [11][12] decompose the multistage system into subsystems that are solved exactly by analyzing the underlying CTMC. They use an iterative procedure to determine the unknown parameters of each subsystem. However, their study has been confined to single product systems. Di Mascolo [8], Baynat et. al [13][14], and Duri et. al [15] present an alternative approximate approach for both single and multi product systems. They too decompose the multi stage system into subsystems, but instead of exactly solving the CTMC, they use product form approximations to solve the subsystems. We have applied this approach and our observation from that effort is that, in a general setting, the analysis can be quite challenging. The difficulty lies in 3 areas: modeling of the performance of the join stations, analyzing stations with general service times in a closed queuing network and accounting for the interaction effects of multiple classes at the various stations of the subsystem. Overcoming these difficulties is the main motivation for the node decomposition approach discussed in this paper.

3. Proposed Method

We propose an alternative decomposition approach to analyze the performance of kanban systems. The outline of the procedure is as follows. The multi stage system is first decomposed into different subsystems each of which is a closed queuing network. To solve the closed queuing network representing each subsystem, each station of the subsystem is analyzed independently by studying the underlying Markov renewal processes. An iterative algorithm is used to compute the unknown throughput and other parameters in each subsystem. Application of approaches similar to the one discussed here can be found in [16][17]. Like the method in Di Mascolo [8] and Baynat et. al [13][14], our method can be applied to analyze fairly general systems with multiple classes. Additionally, it can be easily extended to analyze fairly general classes of material control strategies. The details of the approach are described in the following sections. For ease of understanding, we first describe how the node decomposition technique can be used for analyzing a single stage kanban system. Later, we outline a procedure for analyzing multi stage systems.

3.1 Analysis of the single stage kanban system

Figure 4 represents single stage kanban systems. Figure 4 (a) represents a stand alone single stage kanban system with well-defined arrival processes for raw materials and customer demands. Figure 4 (b) represents a typical single stage system obtained as a result of decomposing a multi stage system into several single stage systems. For such a single stage system there is a "*raw material*" arrival process equal to the arrival process of finished parts from the upstream subsystem. Similarly, one can define a "*demand*" process for such a system as the departure process of the free kanbans from the downstream subsystems. For our analysis in this section, we will use the typical single stage system represented in Figure 4 (b). We assume that the arrival processes for raw materials and customer demands are defined by the estimates of their means and their squared coefficients of variation (SCV).



(a) A stand alone single stage system *(b) A typical single stage in a multistage system*

Figure 4. Queueing model representations of single stage kanban systems

The main idea behind the node decomposition approach is as follows. In the node decomposition approach the closed queuing network is broken down into its constituent workstations or nodes. The input to each queue is assumed to be a renewal process characterized by the mean and SCV of arrival times of customers. Given this information about the input process, two moment approximations are used to derive the characteristics of the output process. Specifically, we derive the mean and the variance of the inter-departure times. By treating the queues approximately as being stochastically independent, the expected equilibrium waiting time at each queue is computed. Finally, the decomposed nodes are pieced together using the fact that in a closed queuing network, the arrival process to each queue is the departure process of the preceding queue in the network. Closed form expressions are then derived for the mean and SCV of the arrival and departure processes to the individual queues. Finally, Little's law is applied to the entire system. This gives rise to a set of non-linear equations in the set of unknowns. These equations are solved to obtain performance metrics such as the system throughput, waiting times, work in progress and so on.

It is evident from the above description that the key lies in developing two moment approximations for the constituent stations in the network. It is easy to see that in our closed queuing network representation of a single stage kanban system two kinds of stations exist.

- The manufacturing stations that can be modeled as a GI/G/1 queue and
- The join stations

In the following two subsections, we analyze both these station types separately. In our description the superscript i wherever present corresponds to the stage index.

3.1.1 Two moment approximations for the manufacturing stations in subsystem S_i

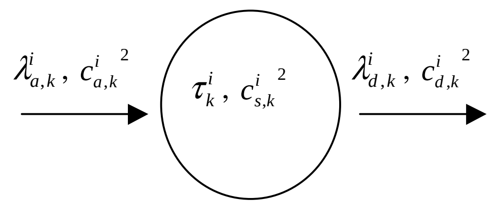


Figure 5. Station k in stage S_i

In this subsection, we use the approximate results developed for the GI/G/1 queue to develop the two moment approximations for each station within the stage S_i. Let stage S_i be composed of M_i single-server stations. Let $\lambda^i_{a,k}$ represent the mean arrival rate and τ^i_k the mean service time of station k, k=1,…,M_i. In addition, let the squared coefficients of variation ${c^i_{a,k}}^2$ and ${c^i_{s,k}}^2$ partially characterize the inter-arrival and service time distributions at station k, respectively. We assume that the utilization rate $\rho^i_k = \lambda^i_{a,k}\tau^i_k$ is less than 1. Given this information, we approximate the departure process at station k by a renewal process and partially characterize it by the mean $\lambda^i_{d,k}$ and squared coefficient of variation of the inter-departure times, ${c^i_{d,k}}^2$. Referring to Figure 5, by flow conservation, we know that the mean of the inter-departure times must equal the mean of the inter-arrival times. Thus,

$$\lambda^i_{d,k} = \lambda^i_{a,k} = \lambda^i_k \text{ for } k=1,\ldots, M_i \tag{1}$$

Further, it has been shown that the squared coefficient of the variation of the inter-departure time, ${c^i_{d,k}}^2$ is given by

$${c^i_{d,k}}^2 = {c^i_{a,k}}^2 + 2{\rho^i_k}^2{c^i_{s,k}}^2 - \frac{2\left(1-\rho^i_k\right)EW^i_k}{\lambda^i_k} \text{ for } k=1,\ldots, M_i \tag{2}$$

where EW^i_k is the expected equilibrium waiting time (delay excluding service time). This expression is originally due to Marshall [18] and, Whitt [19]. Kuehn [20] used the above formula together with an approximation for EW^i_k to approximate the departure processes in networks of queues. In Kuehn [20] and Whitt [19] EW^i_k is approximated by the simple formula

$$EW^i_k \approx \left(\frac{{c^i}_{a,k}^2 + {c^i}_{s,k}^2}{2}\right)\left(\frac{\tau^i_k\rho^i_k}{1-\rho^i_k}\right) \tag{3}$$

so that

$$c^i_{d,k}{}^2 \approx \left(1-\rho_k^{i\,2}\right)c^i_{a,k}{}^2 + \rho_k^{i\,2} c^i_{s,k}{}^2 \text{, for } k=1,\ldots, M_i \tag{4}$$

If L_k^i is the average number of jobs at station k we get:

$$L_k^i = \lambda_k^i \left(EW_k^i + \tau_k^i\right) \tag{5}$$

3.1.2 Two moment approximations for join stations J_i

The operation of the join station J_i is as follows. Let P_i and F_{i+1} denote the two upstream queues of the synchronization station J_i. The content of queue P_i is the number of finished parts in the output buffer of stage i. Each part in P_i has a stage i kanban attached. The content of queue F_{i+1} is the number of free kanbans of stage i+1. As soon as there is one entity in each of the upstream queues, one entity is removed from each of them and added to each of its downstream queues.

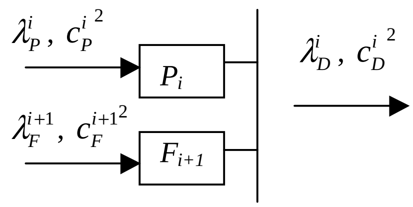


Figure 6. Join station J_i

Let the pairs $\lambda_P^i, c_P^{i\,2}$ and $\lambda_F^{i+1}, c_F^{i+1\,2}$ partially characterize the input processes at queues P_i and F_{i+1} respectively. Given this information, we wish to characterize the departure process from J_i. This involves computing the mean λ_D^i and the squared coefficient of variation $c_D^{i\,2}$ of the inter-departure times. See Figure 6.

In doing so, we build on the work done by Som et. al. [21]. In that work, they study the departure process of a kitting station in a stochastic assembly system. Their primary result was to show that the output process is a markov renewal process. They also derive an expression for the distribution of the time between kit departures assuming that the kit assembly is fed by two independent poisson processes. For our purpose we will need to consider a more general scenario. Specifically, we make the following assumptions.

1. We assume that the two input processes to the join station are independent processes.

2. We approximate the distribution of the inter-arrival times at the input queues P_i and F_{i+1} of the join stations by *balanced mean* 2-Phase Coxian distributions with density functions $\hat{g}_P^i(t)$ and $\hat{g}_F^{i+1}(t)$. Specifically we set:

$$\hat{g}_P^i(t) = c_1\mu_1 e^{-\mu_1 t} + c_2\mu_2 e^{-\mu_2 t} \tag{6}$$

Additionally,

$$\mu = \frac{2}{\lambda_P^i} \text{ and } a_1 = \frac{1}{2c_P^{i\,2}}$$

$$\frac{1}{\mu_1} = \frac{a_1}{\mu_2} = \frac{1}{\mu},$$

$$c_1 = \frac{\mu_1(1-a_1)-\mu_2}{\mu_1-\mu_2} \text{ and } c_2 = 1-c_1$$

Similar equations can be written with $\lambda_F^{i+1}, c_F^{i+1^2}$ and $\hat{g}_F^{i+1}(t)$.

3. Additionally, we assume that the capacities of the input queues P_i and F_{i+1} are bounded from above by K_i and K_{i+1} respectively. In reality, these represent the limitations on the buffer space. For our analysis, this ensures that the system reaches steady state.

We define the state of the join station J_i by the ordered pair (I_i, I_{i+1}) where I_i and I_{i+1} are the number of units in the queues P_i and F_{i+1} respectively. Note that I_i and I_{i+1} cannot simultaneously be both non zero. For each $n \in N$, we can define the random variable Z_n as the inventory position of buffer I_i or the system state of the output process immediately after the n^{th} departure epoch τ_n. The set $Z = \{Z_n : n \in N\}$ defines the output state process and the joint random variables $\{Z,\tau\} = \{Z_n, \tau_n : n \in N\}$ defines the output process. It can be shown that the output process $\{Z,\tau\}$is a Markov renewal process. It is completely characterized by its semi-markov kernel $Q^i(a,b,t)$. The semi-markov kernel of the output process $\{Z,\tau\}$ may be expressed as

$$Q^i(a,b,t) = P\{Z_{n+1} = b, \tau_{n+1} - \tau_n \le t \mid Z_n = a\} \tag{7}$$

By deriving detailed expressions for $Q^i(a,b,t)$ based on the assumed Coxian distribution, it can be shown that the semi-markov kernel is easy to generate as it has a regular structure containing only 5 types of transition equations [22].

In the limit as $t \to \infty$ the semi markov kernel $Q^i(a,b,t)$ also yields the state transition matrix P^i of the underlying Markov chain Z embedded at time τ_n. Clearly, the output process $\{Z,\tau\} = \{Z_n, \tau_n : n \in N\}$ is an irreducible, recurrent and persistent Markov renewal process for $K_i, K_{i+1} < \infty$. Under these conditions it possesses a stationary distribution defined as Π^i. The stationary probabilty vector Π^i of the underlying markov chain is obtained from the set of equations expressed in the matrix form

$$\Pi^i = \Pi^i P^i \tag{8}$$

To determine the distribution of the time between departures, we analyze the output time process $\tau = \{\tau_n : n \in N\}$. We have

$$P\{\tau_{n+1} - \tau_n \le t\} = \Pi^i Q^i(a,b,t) U \tag{9}$$

in which U is a column vector with elements equal to 1. This allows us to determine the mean λ_D^i and the squared coefficient of variation $c_D^{i\ 2}$ of the inter-departure times at station J_i. A review of the procedure is given in Figure 7.

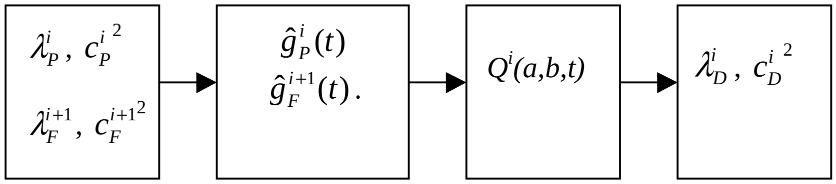


Figure 7. Analysis of join station J_i

From the analysis of the semi-markov kernel, we can also derive expressions for the average number of K_i kanban cards L_{J_i} at J_i. For simplicity, we leave these expressions in their functional form. (Details of these equations are in [22]). We get:

$$L_{J_i} = f\left(\lambda_P^i, c_P^{i\,2}, \lambda_F^{i+1}, c_F^{i+1\,2}\right) \tag{10}$$

3.1.3 Linking the nodes in a single stage kanban system

Once we have characterized the input and output processes at each node, we use known relationships among the parameters at different nodes to obtain a set of closed form non linear equations that can be solved for the unknown parameters. As mentioned earlier, subsystem SUB_i comprises of join stations J_i and J_{i+1} and the k stations in stage S_i. We give below the equations linking the different nodes in our decompositon. For the sake of simplicity, some of the linking equations have been kept in functional form.

Since SUB_i is a closed queuing network with K_i kanbans the average rate of flow of kanbans into and out of all the stations in SUB_i must be equal. Therefore:

$$\lambda_D^{i-1} = \lambda_1^i = \ldots = \lambda_k^i = \ldots = \lambda_{K^i}^i = \lambda_P^i = \lambda_D^i = \lambda_F^i = \lambda^i \tag{11}$$

where λ^i is the throughput of SUB_i.

Matching the squared coefficients of variation of the various input and output processes:

$$c_D^{i-1\,2} = f\left(\lambda_P^{i-1}, c_P^{i-1\,2}, \lambda_F^i, c_F^{i\,2}\right) \tag{12}$$

$$c_D^{i-1\,2} = c_{a,1}^{i\;\,2} \tag{13}$$

For $k = 1,\ldots, M_i$: $\quad {c^i}_{d,k}^{\;2} = \left(1-\rho_k^{i\,2}\right){c^i}_{a,k}^{\;2} + \rho_k^{i\,2}{c^i}_{s,k}^{\;2}$ as in equation (4)

For $k = 1,\ldots, M_i$: $\quad {c^i}_{a,k+1}^{\;2} = {c^i}_{d,k}^{\;2} \quad (14)$

$$c_{d,M_i}^{i\;\;2} = c_P^{i\,2} \tag{15}$$

$$c_D^{i\,2} = f\left(\lambda_P^i, c_P^{i\,2}, \lambda_F^{i+1}, c_F^{i+1\,2}\right) \tag{16}$$

$$c_D^{i\,2} = c_F^{i\,2} \tag{17}$$

Since each subsystem is a closed queuing network we have:

$$L_{J_{i-1}} + \sum_k L_k + L_{J_i} = K^i \quad (18)$$

where the values of L_{J_i} and L_k were defined in equations (5) and (10).

While analyzing subsystem SUB_i using the above equations we assume that the values of $\lambda_P^{i-1}, c_P^{i-1^2}, \lambda_F^{i+1}, c_F^{i+1^2}$ are known. The values of c_{sk}^2, τ_k for k=1,…, M_i are also inputs to the problem. Solving these equations, we obtain the values for performance metrics of the subsystem, such as throughput λ^i, waiting times and so on. For a stand alone single stage kanban system (*i.e.* i=1), the input parameters $\lambda_P^{i-1}, c_P^{i-1^2}, \lambda_F^{i+1}, c_F^{i+1^2}$ correspond to information about the arrival process of raw materials and external demands. If SUB_i is a part of a multistage system, the inputs $\lambda_P^{i-1}, c_P^{i-1^2}, \lambda_F^{i+1}, c_F^{i+1^2}$ are determined by linking the different subsystems. This is discussed in the next section.

4. Extension to Multistage Kanban Systems

The approach to solving multistage systems is described here. An S stage system is first decomposed into S different subsystems SUB_i, i=1,…,S. Individual subsystems SUB_i are analyzed using the method described in section 3. In the analysis of a typical single stage in the previous section, we assumed that $\lambda_P^{i-1}, c_P^{i-1^2}, \lambda_F^{i+1}, c_F^{i+1^2}$ at the queue P_{i-1} of the synchronization station J_{i-1} and queue F_{i+1} of the synchronization station J_{i+1} are given as inputs. In a multistage system, these parameters depend on the subsystems, SUB_{i-1} and SUB_{i+1}.

Next, we describe how these three consecutive subsystems are linked, referring to Figure 8. We make two observations. First, subsystems SUB_{i-1} and SUB_i share the same synchronization station J_{i-1}. Thus the arrival process parameters $\lambda_P^{i-1}, c_P^{i-1^2}$ at the queue P_i that are required as inputs for analyzing SUB_i are the one of the outputs of the analysis of SUB_i Second, we note that the subsystems SUB_i and SUB_{i+1} share the same synchronization station J_i. Thus the arrival process parameters $\lambda_F^{i+1}, c_F^{i+1^2}$ at the queue F_{i+1} that are required as inputs for analyzing SUB_i are the one of the outputs of the analysis of SUB_{i+1}. The above two observations suggest that we will need an iterative procedure running across all the subsystems to obtain the correct values of these parameters. Like the procedure proposed in Di Mascolo et. al [8], we use successive applications of a forward pass and a backward pass. The forward pass updates the $\lambda_P^i, c_P^{i\,2}$ parameters while the backward pass updates the $\lambda_F^i, c_F^{i\,2}$ parameters of SUB_i. We use the fact that the raw material and demand arrivals described by $\lambda_P^0, c_P^{0\,2}$ and $\lambda_F^{S+1}, c_F^{S+1^2}$ are known.

Algorithm:

Step 0. Initialize the unknown parameters $\lambda_P^i, {c_P^i}^2$ for i=1,…, S and $\lambda_F^i, {c_F^i}^2$ for i=0,…, S-1 to some value.

Step 1. *Forward Pass*:

For i=1,…, S-1

Solve SUB_i using node decomposition method using current estimates of $\lambda_P^{i-1}, {c_P^{i-1}}^2, \lambda_F^{i+1}, {c_F^{i+1}}^2$.

Obtain new estimates of $\lambda_P^i, {c_P^i}^2$.

Update if necessary.

Step 2. *Backward Pass*:

For i=S,…, 2

Solve SUB_i using node decomposition method using current estimates of $\lambda_P^{i-1}, {c_P^{i-1}}^2, \lambda_F^{i+1}, {c_F^{i+1}}^2$.

Obtain new estimates of $\lambda_F^i, {c_F^i}^2$.

Update if necessary.

Step 3. Repeat steps 1 and 2 till convergence to the desired accuracy is obtained.

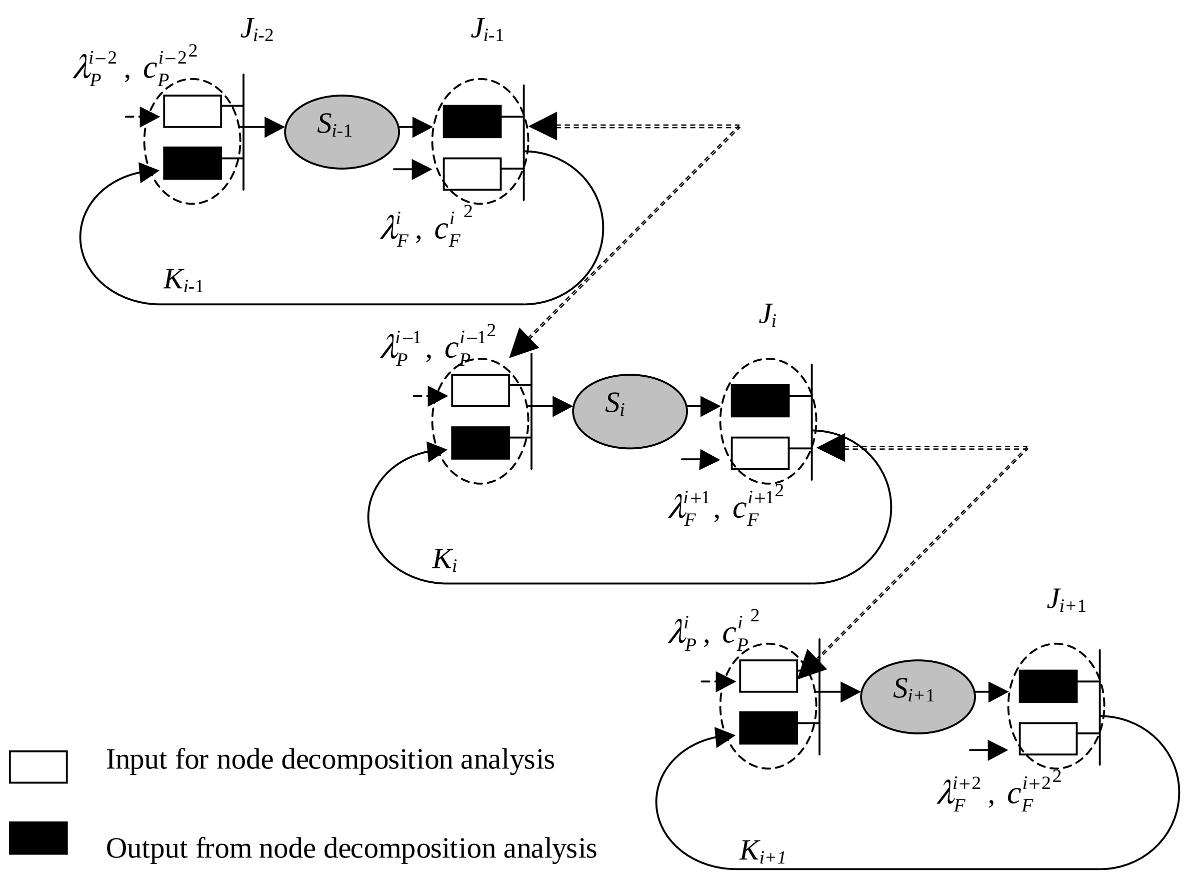


Figure 8. Linking multiple stages

Extension to single product CONWIP systems

Note that the analysis for a single product CONWIP system and the classical kanban systems are special cases of the analysis described above. For the CONWIP system, all stations in the manufacturing network are controlled by a single loop. In our analysis, we include all these stations in one single subsystem. The other extreme is the classical kanban system where each station in the serial line has its own kanban loop. In such a system, each subsystem consists of a single station, *i.e.* k=1 for all subsystems.

5. Extension to Material Control Systems with Multiple Products

In this section we briefly outline how the approach described here could be extended to systems with multiple products operating under kanban, CONWIP or POLCA. The purpose of this section is not to provide the detailed methodology, but just to motivate the idea that the proposed approach can be extended to analyze other material control strategies. One of the advantages of this approach is that the extension to more complex systems does not significantly increase the computational complexity.

From the queuing network representations shown in section 2, it can be seen that the multi product kanban, CONWIP and POLCA systems can be analyzed by the decomposition approach. The manufacturing network can be decomposed into subsystems that are closed queuing networks. The number of stages in a particular subsystem will however depend on the specific material control strategy. As before, we can solve each subsystem as a closed queuing network using the node decomposition technique. Each subsystem can be decomposed into exactly two kinds of stations, the join stations and the stations that can be modeled as a GI/G/1 queue. However, the presence of multiple products might require us to modify the analysis of the join stations. For CONWIP and POLCA systems, the input process of the join station will be either release authorizations or card arrivals. Nevertheless, it is interesting to note that we would still need only the two moment approximations and the linking equations to solve the closed queuing network corresponding to each subsystem. The forward and backward pass procedure described above could be used to link the different subsystems.

6. Conclusions

Before we conclude, in this section we conduct a brief qualitative comparison of our proposed approach to those of Di Mascolo [8], Baynat et. al [13][14], and Mitra and Mitrani [11][12]. This comparison is based on particular characteristics that are common or unique to these approaches.

All the approaches analyze multi stage systems by decomposing it into several single stage subsystems. This seems to be a convenient way of partitioning the overall state space. The approaches differ in the way the single stage systems are analyzed. Mitra and Mitrani [11][12] attempt an exact analysis of the underlying CTMC, Di Mascolo [8] and Baynat et. al [13][14], use product form approximations and the approach in this paper uses node decomposition. The computational burden of solving a CTMC can grow very rapidly in the case of multi product systems. The approach in Di Mascolo [8] and Baynat et. al [13][14], overcomes this difficulty by using product form approximations and taking advantage of the efficient methods available to solve product form networks. To obtain good product form approximation, they use exponential stations with load dependent service times. However, this implies that the number of unknown parameters grows as the number of customers in the network increases. Aggregation techniques have been proposed to simplify computation for the multi product systems. The node decomposition approach proposed here reduces the computations considerably. Unlike the approach in Di Mascolo [8] and Baynat et. al [13][14] the number of unknowns does not increase greatly with increase in the number of customers in the network.

Another property of special interest is the conservation of flow. In a multistage network in series we require that all stages have the same throughput. Through equation 11 we explicitly ensure that this condition is satisfied. In Di Mascolo [8], this constraint is implicit. In the Mitra and Mitrani [11][12] decomposition approach flow conservation is achieved through a set of equations. However Di Mascolo [8] report that these equations are non-symmetrical and this could be a draw back in terms of reversibility. Both our approach and the approach in Di Mascolo [8] satisfy the reversibility property.

It seems evident from the above that trade-offs of simplicity, efficiency and accuracy exist in the development of analytical models for material control strategies. Conducting numerical experiments and verifying the accuracy against simulation for different scenarios form an integral part of our ongoing research.

References

1. Spearman, M.L.,and Zazanis, M.A., 1992, "Push and Pull Production Systems: Issues and Comparisons", *Operations Research*, 40(3), 521-532.
2. Spearman, M.L.,Woodruff, D.L.,and Hopp, W.J., 1989, "CONWIP: A Pull Alternative to Kanban", *International Journal of Production Research*, 28(5), 879-894.
3. Gstettner, S.,and Kuhn, H., 1996, "Analysis of Production Control Systems Kanban and CONWIP", *International Journal of Production Research*, 34(11), 3253-3273.
4. Muckstadt, J.A., and Tayur, S.R., 1995, "A Comparison of Alternative Kanban Control Mechanisms. II. Experiments Results", *IIE Transactions*, 25, 151-161.
5. Suri, R., 1998, *Quick Response Manufacturing: A Company-wide Approach to Lead Time Reduction*, Productivity Press, Portland, OR.
6. Zhou, X., and Luh, P.B., 2000, " The Performance of a New Material Control and Rplenishment System: A Simulation and Comparative Study", R. Suri (Ed.*) Proceedings of the Quick Response Manufacturing 2000 Conference* (Society of Manufacturing Engineers, Dearborn, MI), 807-826.
7. Krishnamurthy, A., Suri, R., and Vernon, M., 2000, "Push Can Perform Better than Pull for Flexible Manufacturing Systems with Multiple Products", *Industrial Engineering Research 2000 Conference Proceedings*, Institute for Industrial Engineers, Norcross, GA, USA.
8. 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.
9. Kimura, O., and Terada, H., 1981, "Design and Analysis of Pull System, A Method of Multi-stage Production Control", *International Journal of Production Research*, 19(3), 241-253.
10. Karmarkar, U.S., and Kekre, S.,1987, " Batching Policy in Kanban Systems", William E. Simon Graduate School of Business Administration, *Working Paper No. QM 87-06*, University of Rochester, Rochester, NY.
11. Mitra, D., and Mitrani, I., 1990, "Analysis of a Kanban Discipline for Cell Coordination in Production Lines, I: Stochastic Demands", *Management Science*, 36, 1548-1566.
12. Mitra, D., and Mitrani, I., 1991, "Analysis of a Kanban Discipline for Cell Coordination in Production Lines, II: Stochastic Demands", *Operations Research*, 35, 807-823.
13. Baynat, B., Dallery, Y., 1996, "A Product-form Approximation Method for General Closed Queuing Networks with Several Classes of Customers", *Performance Evaluation*, 24(3), 165-88.
14. Baynat, B., Dallery, Y., 1993, "A Unified View of Product-form Approximation Techniques for General Closed Queuing Networks", *Performance-Evaluation*, 18(3), 205-24.
15. Duri, C., Frein, Y., Di Mascolo, M., 1995, "Performance Evaluation of Kanban Multiple-Product Production Systems", *Proceedings 1995 INRIA/IEEE Symposium on Emerging Technologies and Factory Automation*.ETFA'95 (Cat. No. 95TH8056), IEEE Computing Society Press, Los Alamitos, CA, USA, (3)(xi+709+603+580), 557-66.

16. Kamath, M., Suri, R., and Sanders, J.L., 1988, "Analytical Performance Models for Closed Loop Flexible Assembly Systems", *The International Journal of Flexible Manufacturing Systems*, 1, 51-84.
17. Suri, R., Sanders, J.L., and Kamath, M., 1993, "Performance Evaluation of Production Networks", in *Handbooks of OR/MS*, S.C. Graves (ed.), Volume 4, Elsevier, North-Holland, Amsterdam, 199-286.
18. Marshall, K.T., 1968, " Some Inequalities in Queuing", *Operations Research*, 16, 651-665.
19. Whitt, W., 1984, "Approximations to Departure Processes and Queues in Series", *Naval Research Logistics Quarterly*, 31, 499-521.
20. Kuehn, P.J., 1979, "Approximate Analysis of General Queuing Networks by Decomposition", *IEEE Transactions on Communications COM-27*, 113-126.
21. Som, P., Wilhelm, W.E. and Disney, R.L., 1994, "Kitting Process in a Stochastic Assembly System", *Queuing Systems*, 17, 471-490.
22. Krishnamurthy, A., Suri, R., and Vernon, M., 2000, "Node Decomposition Methods for Analyzing Queuing Models of Material Control Strategies in Manufacturing Systems", *Working Paper*, Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI.