SELECTION OF INVENTORY CONTROL POINTS IN MULTISTAGE PULL SYSTEMS

Ronald G. Askin
Shravan Krishan

Department of Systems & Industrial Engineering
University of Arizona
Tucson, AZ 85721-0020 USA
Phone: 520-621-6548
e-mail: ron@sie.arizona.edu

Abstract:
We consider multistage, stochastic, dynamic production systems using pull control for production authorization in discrete parts manufacturing. These systems have been widely implemented in recent years and constitute a significant aspect of lean manufacturing. Extensive research has appeared on the optimal sizing of buffer inventory levels in such systems. However, the issue of control points, i.e. where in the multistage sequence to locate the output buffers, has not been addressed for pull systems. Allowable container/batch sizes, optimal inventory levels, and ability of systems to automatically adjust to stochastic demand depend on the location of these control points. In this paper, necessary and sufficient conditions are derived for ensuring that a single, end-of-line accumulation point is optimal. When this is not the case, an algorithm is provided to determine the optimal control points. Results are extended to determine the optimal container size for both deterministic and stochastic systems when lead time at a stage is a function of batch size.

Keywords: Kanban, Pull Control, Buffers, Inventory

Introduction

Extensive research has been carried out on pull control based production systems in the early and mid nineties. In such systems, each inventory storage buffer is designated with a desired level of inventory for each part. This level is typically defined as a desired number of containers with each container holding a designated quantity of the product. This quantity is referred to as the container size. Following the procedure popularized by the successful Toyota Production System, “kanbans” or cards are used to authorize production. Each kanban corresponds to one container. Kanbans are kept in the output buffer attached to a full container of parts. When the first part from that container is required by the succeeding work area, the container is removed from storage and the kanban is recirculated to the production station to authorize replenishment of those items. The system automatically paces and prioritizes orders within workstations based on downstream consumption. This makes it easy to implement and self-adjusting. If the proper level of kanbans is selected, and the replenishment and demand processes are highly predictable and deterministic, then the system can be operated such that a completed container of parts reenters the output buffer just in time for its need at the downstream workstation. Hence, such systems are often considered part of the JIT (Just-in-Time) production control philosophy.

Most of the previous research focused on determining the number of kanbans, with lesser emphasis placed on container sizes and product sequence in a just-in-time (JIT) shop. However, in a multistage system, it is not necessary to include an output buffer at each stage. Within control sections (the area or sequence of workstations between buffers), a push philosophy can be incorporated. Once production is
authorized by the removal of a container from the control sections’ output buffer, a replenishment order is released to the first workstation in the control section. These orders then have authorization to flow through each stage of the control section until again reaching the output buffer, i.e. they are pushed through without waiting for a customer request. The issue of where to locate control points, although important, has not been effectively addressed in the past. In our study, we attempt to find a set of production stages in a JIT system that will act as inventory control points. These control points are the only stages that store inventory. In addition, we determine the number of kanbans and container sizes that should be used in specific conditions.

**Literature Review**

The production system that we consider in this paper is the single kanban system described by Monden (1983) as being used in the Toyota Production System for serial systems with closely located sequential stages. An overview of various extensions and refinements is given in Askin and Goldberg (2002). As indicated above, prior research has emphasized the choice of the number of kanbans to use in such systems. Philipoom et al. (1987) showed experimentally that the lead time demand distribution constitutes a major determinant of the number of kanbans needed. For modeling a kanban system with a fixed parameter specification or to determine the number of kanbans needed, stochastic analytical models with discrete time periods (Deleersnyder et al. 1989) and continuous time (Mitra and Mitrani 1990, Wang and Wang 1990, Askin et al. 1993) have also been presented. Spearman et al. (1990) describe a related pull control concept called CONWIP (CONstant Work-In-Process). CONWIP limits total inventory in the system instead of at a single workstation or stage. Within the system, the natural characteristics of the workstations will determine the inventory levels at each workstation. Askin and Goldberg (2002) discuss factors that should be considered in selecting the location of buffers in a serial line, but this issue has not been widely studied for pull systems. However, several researchers have discussed the allocation of inventory to stages in multiechelon production-distribution systems under other control strategies.

**Problem Statement**

Figure 1 below shows a single kanban Just-in-time (JIT) system. Circles represent workstations or stages and triangles represent inventory locations (output buffers). Each production-ordering kanban, either physical or as an electronic token, authorizes one container of parts of a given part type. The kanbans flow within a given control section, circulating from the output buffer where they are attached to a full container back through all production stages within the control section once they are detached upon withdrawal of the container. A proper design of system control points may improve coordination and reduce total costs. Our study aims to define a characterization of system control points for different environmental conditions and using a set of increasingly complete models.
The main problem we wish to address is which stages should serve as the control points for such a system. We assume that production batch and unit load sizes correspond to the container size.

**Model 1**: We initially consider a serial production line, producing a single part type. The batch size at each stage is assumed to be known and fixed across all stages. Demand occurs at stage \( m \) and is stochastic and independent in non-overlapping time segments. (For notational simplicity we will assume demand is Normally distributed.) There is a fixed cost associated with setting up a control point. We assume that the lead time at any stage is known. The lead time distribution is likewise therefore fixed for each stage. The number of kanbans is set to provide coverage against a defined upper percentage point (assumed to be close to 1) of the lead time demand distribution. Also, we assume a value added structure so that the holding cost at any stage is greater than the holding cost at its preceding stages. The objective is to select a set of control points that minimize the location cost plus setup cost plus inventory holding cost.

The notation used is as follows:

- \( D \) - Mean demand per time at stage \( m \)
- \( \sigma^2 \) - Variance of demand at stage \( m \) per time period
- \( f_i \) - Fixed cost per time for locating and maintaining inventory at stage \( i \)
- \( \alpha \) - Percentage of orders satisfied without delay (service rate)
- \( z_\alpha \) - Standard Normal Variate such that \( P[Z \leq Z_\alpha] = \alpha \)
- \( M \) - Set of all stages \{1, ..., \( m \}\}
- \( h_i \) - Inventory holding cost per unit at stage \( i \), \( i \in M \)
- \( L_i \) - Lead time at stage \( i \)
- \( X_i \) - \( \begin{cases} 
1 & \text{if stage } i \text{ is a control point} \\
0 & \text{otherwise} 
\end{cases} \)
- \( Y_{ij} \) - \( \begin{cases} 
1 & \text{if stage } j \text{ is a control point that serves stage } i \\
0 & \text{otherwise} 
\end{cases} \)
- \( a_i \) - Setup cost plus material handling cost per container at stage \( i \)
- \( n \) - Container size

The decision problem then becomes minimization of expected cost per period or

\[
\text{Min} \sum_{i \in M} f_i X_i + \sum_{i \in M} \frac{a_i D_i}{n} + \sum_{i \in M} X_i \cdot h_i (n-1) + z_\alpha \sigma \sum_{j \in M} h_j \sqrt{\sum_{i,j} Y_{ij} L_i} 
\]

Subject to:

\[
\sum_{j \neq i} Y_{ij} = 1 \quad \forall i \in M 
\]

\[
Y_{ij} \leq X_j \quad \forall i,j \in M, \ i \leq j 
\]

\[
Y_{ij} \geq Y_{i-1,j} \quad \forall i,j \in M 
\]

\[
X_m = 1 
\]

\[
X_j \in [0,1], \ Y_{ij} \in [0,1] 
\]
Here, the objective function, Equation (1), minimizes the sum of the total location cost, setup cost, work-in-process (WIP) inventory holding cost plus, safety stock holding cost. The first term in the objective function represents the fixed cost of locating a control point. The second term represents the total annual setup and material cost. The third term represents the WIP inventory holding cost of the partial container available as input to stage \( j+1 \). On average, there are \( \frac{n-1}{2} \) parts waiting at each first work station within a control section assuming continuous production of a single part type. The last term represents the safety stock holding cost at each control point. For the system described, expected on-hand inventory is excess of maximum inventory from expected lead time demand. Thus if the system is fully deterministic with \( \sigma = 0 \) or, in our approximation, if \( \alpha = 50\% \), the system is completely synchronized and output buffers are always empty.

Constraint (2) ensures that each stage is served only by one control point. Constraint (3) ensures that a stage can be controlled only by a control point. Constraint (4) prevents any skips. Constraint (5) ensures that stage \( m \) (the last stage) is always chosen to be a control point. Constraint (6) defines the binary restrictions on the variables.

**Solution Procedure:** This model can be reformulated as a shortest path problem as follows:

Consider stage 0 to be the input stage. Also, let \( M_{0j} \) be the cost of selecting stage \( j \) as the first inventory control point and \( M_{jk} \), the cost of having \( j \) and \( k \) as two consecutive control points. This is shown in Figure 2 below.

![Figure 2: Shortest path representation of Model 1](image)

If \( n \) is fixed, \( \sum_{i=M}^{D} \frac{a_i D}{n} = k_1 \), a constant for all feasible solutions. Also, \( \frac{h_i(n-1)}{2} \) can be added to \( f_i \) to eliminate the third term in the objective. We redefine \( f_i \) accordingly for the remainder of this section. Thus, in order to find the optimal solution to the model, we can ignore these two terms from the objective function. In general then:

\[
M_{jk} = f_k + z_a \sigma h_k \sqrt{\sum_{l=j+1}^{k} L_l} \quad \forall j = 0, \ldots, m-1; \quad k = j+1, \ldots, m
\]  

(7)

For this model with predetermined lead times, positive echelon holding costs, and constant container size, analysis leads to the following result on locating buffers:
Theorem 1: For the cost structure defined in model 1 a single control point is always optimal if for all stages \( j \), \( f_j > z_a \sigma \left( h_m \sqrt{\sum_{i=1}^{m} L_i} - \left( h_j \sqrt{\sum_{k=1}^{j} L_k} + h_m \sqrt{\sum_{r=j+1}^{m} L_r} \right) \right) \).

Proof: We will prove the theorem for a two-stage problem and then for a general case using mathematical induction. The proof is simple for a two-stage problem. This problem can be represented as shown in Figure 3.

![Figure 3: Shortest path representation for a two-stage control problem](image)

In the above Figure, it is clear that a single control point is optimal if and only if \( M_{02} < M_{01} + M_{12} \), where the costs can be calculated using Equation (7). Therefore,

\[
M_{02} = f_2 + z_a \sigma h_2 \sqrt{L_1 + L_2}, \quad M_{01} = f_1 + z_a \sigma h_1 \sqrt{L_1}, \quad \text{and} \quad M_{12} = f_2 + z_a \sigma h_2 \sqrt{L_2}.
\]

Substituting in these expressions and rearranging terms, the sufficient condition for a single control point becomes

\[
f_1 > z_a \sigma \left( h_2 \sqrt{L_1 + L_2} - h_1 \sqrt{L_1} - h_2 \sqrt{L_2} \right)
\]

Thus, the theorem is true for \( m = 2 \).

For an \( m+1 \)-stage problem, if a single control point is optimal for all \( k \)-stage problems such that \( k \leq m \), we need to consider only the two-arc solutions. This is because any path to a node \( k \leq m \) that uses two or more arcs is dominated. Therefore, for an \( m+1 \)-stage problem, a single control point is optimal if:

\[
M_{0,m+1} < M_{0,j} + M_{j,m+1} \quad \forall j \in M, \ j < m + 1
\]

Once again we can use Equation (7) to compute the costs. Therefore,

\[
M_{0,m+1} = f_{m+1} + z_a \sigma \sqrt{\sum_{i=1}^{m+1} L_i}, \quad M_{0,j} = f_j + z_a \sigma \sqrt{\sum_{i=1}^{j} L_i}, \quad \text{and} \quad M_{j,m+1} = f_{m+1} + z_a \sigma \sqrt{\sum_{i=j+1}^{m+1} L_i}.
\]

Substituting the above terms into Equation (9) and rearranging terms gives:

\[
f_j > z_a \sigma \left( h_m \sqrt{\sum_{i=1}^{m} L_i} - \left( h_j \sqrt{\sum_{k=1}^{j} L_k} + h_m \sqrt{\sum_{r=j+1}^{m} L_r} \right) \right)
\]

Thus, based on the assumption that the theorem is true for an \( m \)-stage problem, it is true for an \( m+1 \)-stage problem. Hence it is true for all \( j \in M \). This concludes the proof.

Discussion: In general, the RHS of Equation (10) is negative. It is positive only if the ratio \( h_m / h_j, \ m > j \), is large. If we assume that \( h_m = h_j = h \), Equation (10) reduces to:

\[
f_j > z_a \sigma h \left( \sqrt{\sum_{i=1}^{m+1} L_i} - \left( \sqrt{\sum_{k=1}^{j} L_k} + \sqrt{\sum_{r=j+1}^{m+1} L_r} \right) \right)
\]

The right hand side in Equation (11) is clearly negative. In this case, a single control point is always optimal even if there is no fixed cost to maintain an inventory buffer.
Model 2: In model 1 we assumed that the container size, $n$, was known. For Model 2 we assume that the container size is unknown. However, the container size is assumed to be constant across all stages. All the other assumptions of Model 1 apply. The formulation remains the same as that for model, except for one additional constraint:

$$n \leq n^\text{max}_j \quad \forall j \in M \quad (12)$$

Here, $n^\text{max}_j$ is the maximum container size that stage $j$ can handle. This would typically be dictated by the material handling technology or part form at stage $j$. Thus, we have an additional decision variable, $n$ for this model. We will consider two cases.

Case i) Fixed processing time operations: Here, the lead time, $L$, is independent of the container size. In this case the problems of container size and location of control points are readily seen to be separable in the formulation Equations (1) to (6) and (12) if we assume a single control point. The results of Askin and Goldberg (2002) can be extended to multiple stages to conclude $n^* = \sqrt{\frac{\sum_{i \in M} 2a_i D}{h_m}}$. If any constraint (12) is violated, then set $n = \min_j n^\text{max}_j$.

Case ii) Variable processing time operations: Here, lead time is a function of container size $n$ and is given by the equation: $L = k(s_i + p_i \cdot n)$

where,

- $s_i$ - setup time at stage $i$.
- $p_i$ - processing time per unit at stage $i$.
- $k$ - a constant.

In this case, the lead time at a stage is dependent on the container size. WIP cost should therefore be included in the objective function. We add the term $\sum_{i \in M} h_i \cdot L_i \cdot D$ to reflect the time each unit spends in process at each stage.

For a given $n$, the model can still be represented as a shortest path problem. Also, since the container size is fixed across all stages, Theorem 1 is valid for this case too. Thus, it is likely that a single control point will be optimal. Thus, a two stage solution approach seems reasonable. First, determine $n$. Setting $X_m = 1$, and $X_i = 0, i < m$, converts the objective function into a single variable function that can be readily solved for the optimal batch size. (The upper bounds for each stage must still be checked for feasibility and $n$ adjusted accordingly if necessary.) Given $n$, the $L_i$ can be computed and model 1 applied to ensure a single buffer.

Model 3: For this model we assume that the container sizes are not constant across stages, each control section selects a container size. All the other assumptions for model 1 apply. We consider the same two cases that we considered for model 2.

Solution Procedure:

Every stage that is served by control point $j$ will have the same container size as stage $j$. The objective function for this model will now be:
\[ \text{Min} \sum_{j \in M} f_j X_j + \sum_{i,j \in M} a_{ij} D \left( \sum_{j=1}^{n_j} Y_{ij} n_j \right) + \sum_{j \in M} h_j \cdot X_j \cdot (n_j - 1) \cdot \frac{1}{2} + \sum_{i,j \in M} h_i \cdot L_i \cdot D + z_a \sigma \sum_{j \in M} h_j \sqrt{\sum_{i,j} Y_{ij} L_i} \]

where \( n_j \) is the container size at stage \( j \).

In addition to the constraints in model 1, we have:

\[ n_j \leq Y_{ij} \cdot n_{ij}^{\text{max}} \quad ; \forall j \in M; \ i = 1, \ldots, j \]

\[ n_j = \sum_{j=1}^{n_j} Y_{ij} \quad ; \forall i \in M \]

Objective (13) incorporates the common container quantity within control sections. Constraint (14) ensures that the container size in each control section does not exceed the maximum allowable container size for the stages covered. Constraint (15) holds the container size constant with a section.

Because the container size is not constant, Theorem 1 is not valid for this model. However, we can still formulate the model as a shortest path problem, as in Figure 2. The arc costs are as follows:

**Case i)**

In general:

\[ M_{jk} = f_k + \sum_{l=0}^{j-1} a_{l,k} \cdot \frac{h_k \cdot (n_{jk}^* - 1)}{2} + \sum_{l=0}^{j} h_l L_l D + z_a \sigma h_k \sqrt{\sum_{l=0}^{j} L_l} \quad \forall j = 0, \ldots, m - 1; \ k = j + 1, \ldots, m \]

where,

\[ n_{jk}^* = \min \left\{ n_{j+1}^{\text{max}}, \ldots, n_k^{\text{max}}, \sqrt{\frac{2 \cdot D \cdot \sum_{l=0}^{j} a_l}{h_k}} \right\} \]

**Theorem 2:** For the cost structure defined in model 3 a single control point is always optimal, i.e., stage \( m \) is the only control point, if for all stages \( j \),

\[ f_j > \left( \sum_{l=0}^{m} a_{l,m} \frac{n_{l,m}}{n_{0,m}} - \sum_{l=0}^{m} a_{l,j} \frac{n_{l,j}}{n_{0,j}} \right) \cdot D + \sum_{l=0}^{j} h_l (n_{0,m}^* - n_{0,j}^*) \]

\[ \frac{\sum_{l=0}^{j} h_l (n_{0,m}^* - n_{0,j}^*)}{2} + z_a \sigma \left( \sqrt{\sum_{l=0}^{m} L_l} - \left( h_j \sqrt{\sum_{l=0}^{j} L_l} + h_m \sqrt{\sum_{l=0}^{m} L_l} \right) \right) \]

**Proof:** The proof follows the same procedure used in Theorem 1.

**Case ii)** The only difference between Case (i) and Case (ii) is that the lead time depends on the container size in Case (ii). Arguments similar to those used in Theorem 2 can be applied to Case (ii) after including WIP cost. However, the difference is in the calculation of the optimal container size at each control section. In general:
\[ M_{jk} = f_k + \sum_{l=j+1}^{k} a_l D \frac{n_{jk} - 1}{h_l} + \sum_{l=j+1}^{k} h_i \cdot L_i (n_{jk}^*) \cdot D + z_a \sigma h_i \sqrt{\sum_{l=j+1}^{k} L_i} \]

\[ \forall j = 0, 1, \ldots, m - 1; \, k = j + 1, \ldots, m \]

where, \( n_{jk}^* \) can be determined by minimizing:

\[ f(n_{jk}) = \frac{\sum_{l=j+1}^{k} a_l D \frac{n_{jk} - 1}{h_l}}{n_{jk}} + \sum_{l=j+1}^{k} h_i \cdot L_i (n_{jk}^*) \cdot D + z_a \sigma h_i \sqrt{\sum_{l=j+1}^{k} L_i} \]

\[ \forall j = 0, 1, \ldots, m - 1; \, k = j + 1, \ldots, m \]

Subject to:

\[ n_{jk} \leq n_{l}^{\max} \quad \forall l \in (j, k] \]

and \( L_i (n_{jk}^*) = k \cdot (s_j + p_j \cdot n_{jk}^*) \).

**Summary and Conclusions**

We have shown that under reasonable conditions on inventory holding costs, and a predetermined container size, a pull system should combine stages into a single control sector and use a single output buffer. Choice of container size and minimum production batch can be incorporated into the model. The shortest path model describes or serves as a useful subproblem for several extensions of the basic formulation. Future work will incorporate the stochastic effect of workstation variability, multiple products and structures, and queuing effects.

**References:**