Int J Adv Manuf Technol (2003) 21:754–759
Ownership and Copyright

The International Journal of
Advanced Manufacturing Technology

Design of a Two-Card Dynamic Kanban System Using a Simulated Annealing Algorithm

P. Shahabudeen, K. Krishnaiah and M. Thulasi Narayanan

Industrial Engineering Division, College of Engineering, Anna University, Chennai (Madras), India

The design of a kanban system addresses the selection of two important parameters, i.e. the number of kanbans and lot sizes of part types. Kanban-based operational planning and control issues have been tackled in a number of studies by means of analytical or simulation modelling. However, the estimation of these parameters becomes complicated because of issues such as variation in demand, variation in processing times, different types of products, etc. The combinatorial property of such problems warrants the development of efficient methodology or heuristics to obtain a good solution. In this paper, an attempt has been made to select the number of production and withdrawal kanbans at each workstation and the lot size for each part type required to achieve the best performance using a simulated annealing algorithm technique. An object-oriented simulation model of a two-card dynamic kanban system capable of handling different types of part with different demand requirements has been developed and used for the analysis. Each part type has its own number of production ordering kanbans and withdrawal kanbans at each workstation. The lot size can also be different for different part types. A bicriteria objective function comprising mean throughput rate and aggregate average kanban queue has been used for evaluation. Different types of problem have been tried out and the performance of the algorithm is studied.

Keywords: Kanban; simulated annealing; simulation

1. Introduction

Just in time (JIT) is the manufacturing philosophy of production systems producing what is needed at the right time and in right quantity [1]. Kanban coupled with a pull system of production is used as a means of implementing JIT. The essential elements in the design of a kanban production system are the number of kanbans needed to link the processes together, the number of machines, and the appropriate unit of lot size [2].

2. Literative Review

Kanban-based operational planning and control issues have been tackled in a number of studies by means of analytical or simulation modelling. Berkley [2] has reviewed 50 papers in the area of kanban production control and has organised them based on the type of system. He has also listed 24 vital operation design factors of a kanban system. Price et al. [3] have reviewed optimisation models of kanban-based systems. They have concluded that a worthwhile direction for future development would be the incorporation of models in decision-support systems for production control.

Several workers have attempted a mathematical modelling approach. They have assumed a deterministic demand and the material handling being carried out periodically and the periods corresponding to fixed withdrawal cycles. Kimura and Terada [4] have developed a model for a kanban system, which can be considered to be a pioneering work in the area of kanban modelling. Bitran and Chang [5] have extended the work of Kimura and Terada and offered a mathematical model for a kanban system in a multi-stage production setting. Gupta and Al-Turki [6] have described a new kanban system called a flexible kanban system. This system uses an algorithm based on a mathematical model of the system to manipulate dynamically and systematically the number of kanbans and the starvation caused by stochastic factors.

Formulating kanban-controlled lines as Markov chains has been a popular strategy for finding the optimal number of kanbans. Deleersnyder et al. [7] have modelled a line with blocking by total queue size, as a discrete time Markov chain to study the effect of kanban numbers, machine reliability, processing time and finished goods demand variability. Berkley [8] has developed a decomposition approximation using embedded Markov chains for two-card systems with periodic material handling and Erlang processing times. Nori and Sarker [9] have modelled a kanban system using Markov Chains to determine the optimum number of kanbans between adjacent

Correspondence and offprint requests to: Dr P. Shahabudeen, Industrial Engineering Division, Anna University, Chennai 6000 025, India. E-mail: psdeen@annauniv.edu; psdeen@hotmail.com
Received 30 January 2002
Accepted 13 April 2002

stations. For finding the required number of kanbans and lot sizes, simulation offers a number of advantages. In a simulation study of a four-station line with fixed order points, Berkley [10] has shown that order points have a significant effect on average inventories and production rates. In another simulation study of a two-card system with fixed withdrawal cycle times, Berkley and Kiran [11] have found that the average in-process inventories are highly sensitive to withdrawal cycle times, whereas finished goods withdrawal kanban queue time is much less sensitive. Savsar [12] has presented a simulation study of a JIT production control system and its performance under different operational conditions. Berkley [13] has simulated a two-card kanban system with multiple part types, to determine the effect of container size on average inventory and customer service levels combined, so that total in-process capacity remains constant. Results show that smaller containers lead to smaller average inventories. Andijani [14] has used stochastic system simulation to generate and construct a set of kanban allocations, which maximise the average throughput rate and minimise the average system time.

3. Present Study

3.1 Multi-Criteria Objective Function

It is found that most workers have used only a single objective function as the performance measure. In this study, percentage zero demand, mean lead time and mean total work in process are used as measures of performance.

3.1.1 Percentage Zero Demand (PZD)

Service level represents the probability of meeting the demand from the output buffer. In the JIT environment, for the customers to receive the goods just in time, almost 100% service level has to be maintained. Here, PZD is defined as the percentage of total demand immediately satisfied, to the total demand generated (similar to the percentage of zero entries in a queuing system). PZD is proportional to the service level. Hence, in this study, the objective is to maximise the PZD value.

$$PZD = \frac{\sum_{i=1}^{p} SD_i}{\sum_{i=1}^{p} TD_i} \times 100 \tag{1}$$

where,

p = number of part types

SD_i = demand immediately satisfied for part type i

TD_I = total demand generated for part type i

3.1.2 Mean Lead Time (MLT)

Whereas the external lead time is controlled by PZD, the internal lead time, consisting of various waiting times and processing times, should also be minimised to meet the overall objective of JIT. Here, MLT is the time spent by a product in the system, which is the sum of waiting times, processing times and moving times, at various work stations, averaged per station per product as follows:

$$MLT = \frac{\sum_{i=1}^{p} \left\{ \sum_{j=1}^{s} (PT_{ij} + IW_{ij} + OW_{ij}) + \sum_{j=1}^{s-1} MT_{ij} \right\}}{p \times s} \tag{2}$$

where,

s = number of work stations

PT_{ij} = processing time of part type i at work station j

IW_{ij} = waiting time of part type i at the input area of work station j

OW_{ij} = waiting time of part type i at the output area of work station j

MT_{ij} = moving time of part type i from work station j to $j + 1$

3.1.3 Mean Total Work in Process (MTW)

One of the objectives of JIT philosophy is to have no WIP. Since the total number of kanbans in the system has an impact on WIP, the number of kanbans to be made available has to be determined in such a way that WIP is maintained at a minimum level. The lots attached either to production or withdrawal kanbans represent the WIP at the output or input buffers of a station, respectively. Hence, MTW is the average number of kanbans (both production and withdrawal kanbans) waiting for each part type at each work station.

$$MTW = \frac{\sum_{i=1}^{p} \sum_{j=1}^{s} (WI_{ij} + WO_{ij})}{p} \tag{3}$$

where,

WI_{ij} = WIP of part type i at input area of work station j

WO_{ij} = WIP of part type i at output area of work station j

These measures are used to evaluate the different characteristics of the kanban model. Using the heuristic, we will try to improve and obtain the best possible value for a given single objective. In order to study the combined effect of these performance measures and design the kanban system parameters, it is required to devise a multi-criteria objective function comprising the above three measures. A simple addition or any other mathematical combination of the three measures is impossible because it is required to maximise the PZD, while the MLT and the MTW are to be minimised. Hence, the values of MLT and MTW are normalised using index values, so that they become maximisation measures. The index values MLT_INDEX and MTW_INDEX represent the maximum possible values ever attainable for the MLT and MTW, respectively. An error-trapping routine is included in the simulation model to stop the execution, in the event of either MLT or MTW exceeding these index values. In such instances, the index values are modified and all the experiments are repeated with new index values. The normalised values RMLT and RMTW are defined as:

$$RMLT = (1 - MLT/MLT\_INDEX) \times 100 \quad (4)$$

$$RMTW = (1 - MTW/MTW\_INDEX) \times 100 \quad (5)$$

Hence, the modified objective function:

$$Z = \max(\alpha_1 \, PZD + \alpha_2 \, RMLT + \alpha_3 \, RMTW) \quad (6)$$

where α_1, α_2 and α_3 are the weighting factors for each measure.

However, in this study it is assumed that $\alpha_1 = \alpha_2 = \alpha_3 = 1$. The maximum possible value for Z is 300.

3.2 Search Technique

The literature review reveals that search techniques have not been widely applied for setting the kanban system parameters. Hence, it is felt that some search heuristics may be useful for handling such problems. In the case of search techniques, the simulated annealing (SA) technique is found to have frequent application in production related problems such as group technology, scheduling, etc., for obtaining good results. In this study, the simulated annealing technique has been used. The kanban simulation model is embedded in the search heuristics. For every new solution set generated by the heuristic, simulation is performed and the objective function value Z is computed. This procedure is repeated until the search heuristic reaches the termination point.

4. The Kanban System Model

The system under study is a dynamic kanban system with stochastic demand and processing times with the following assumptions:

Multiple types of job are produced.

Each type of job has its own demand inter arrival distribution.

Each type of job has its own demand distribution.

All the jobs must be processed in all the stations.

In each station, each part type has its own number of production ordering kanbans (POK) and withdrawal kanbans (WK).

Each station has a different number of machines (NOM).

Each part type has its own lost size (LS).

Processing times of each job type are different.

The demand is satisfied on an FCFS basis.

There is an infinite supply of raw material at the input of the production system.

Any kanban detached at the output of a stage is immediately available for the upstream stage.

The distance between the workstations is not substantial.

Material handling resources are unrestricted.

A number of problems with different numbers of workstations and different numbers of part types have been used for the study.

Table 1. Levels assigned to SAA control parameters.

Factors	Level 1	Level 2
A. Initial temperature	**1.0**	1.5
B. Final temperature	**0.1**	0.5
C. Accept count	**4.0**	7.0
D. Cooling rate	0.7	**0.9**
E. Freeze count	3.0	**5.0**

Note: Values assigned after experimentation are shown bold.

5. Design of SA Based Kanban Parameter Search Model

5.1 Solution Vector

The solution vector S_n mentioned in the algorithm is a linear array containing the number of production kanbans for each part type at each work station, the number of withdrawal kanbans for each part type at each work station, the lot size of part types and the number of machines made available at different work stations.

$$S_n = [POK_{11}, POK_{12}, \ldots POK_{ps}, WK_{11}, WK_{12}, \ldots, WK_{ps}, LS_1, LS_2, \ldots, LS_p, NOM_1, NOM_2, \ldots, NOM_s]$$

The following is an example of the solution vector for a three part type, three work station problem.

$$S_n = [2\ 3\ 6\ 1\ 6\ 2\ 6\ 7\ 3\ 7\ 3\ 7\ 8\ 2\ 3\ 4\ 7\ 3\ 100\ 50\ 80\ 4\ 2\ 5]$$

5.2 Design of SA Algorithm Control Parameters

The efficiency of the SA algorithm depends upon the values set for the following control parameters:

A. Initial temperature t_i.

B. Final temperature t_f.

C. Freeze limit Φ.

D. Cooling rate α.

E. Accept limit β.

Using pilot runs, the reasonable range over which the above parameters could be varied is determined as given in Table 1.

In order to set the SAA control parameter values, an orthogonal array experiment is used. There is no interaction effect

Table 2. Response (Z^*) from the orthogonal experiment.

Experiment	A	B	C	D	E	E_1	E_2	Z^*	Number of perturbations
1	1	1	1	1	1	1	1	287.82	83
2	1	1	1	2	2	2	2	292.92	247
3	1	2	2	1	1	2	2	283.53	41
4	1	2	2	2	2	1	1	290.38	84
5	2	1	2	1	2	1	2	292.14	117
6	2	1	2	2	1	2	1	295.00	425
7	2	2	1	1	2	2	1	288.80	34
8	2	2	1	2	1	1	2	286.53	59

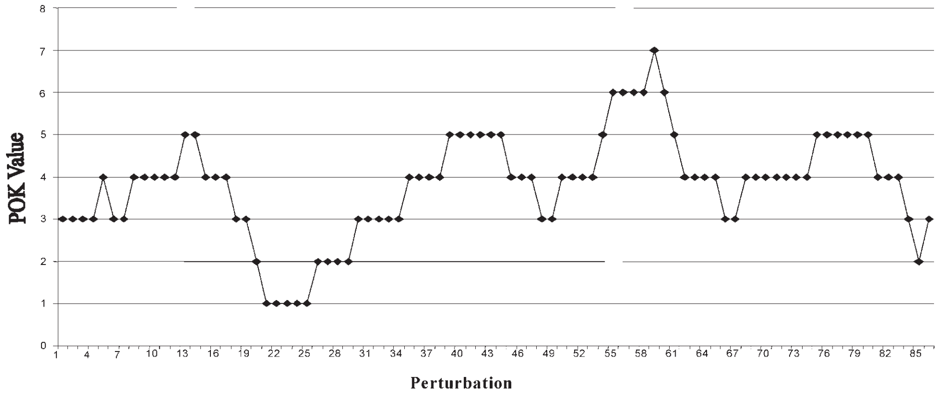

Fig. 1. SAA exploration.

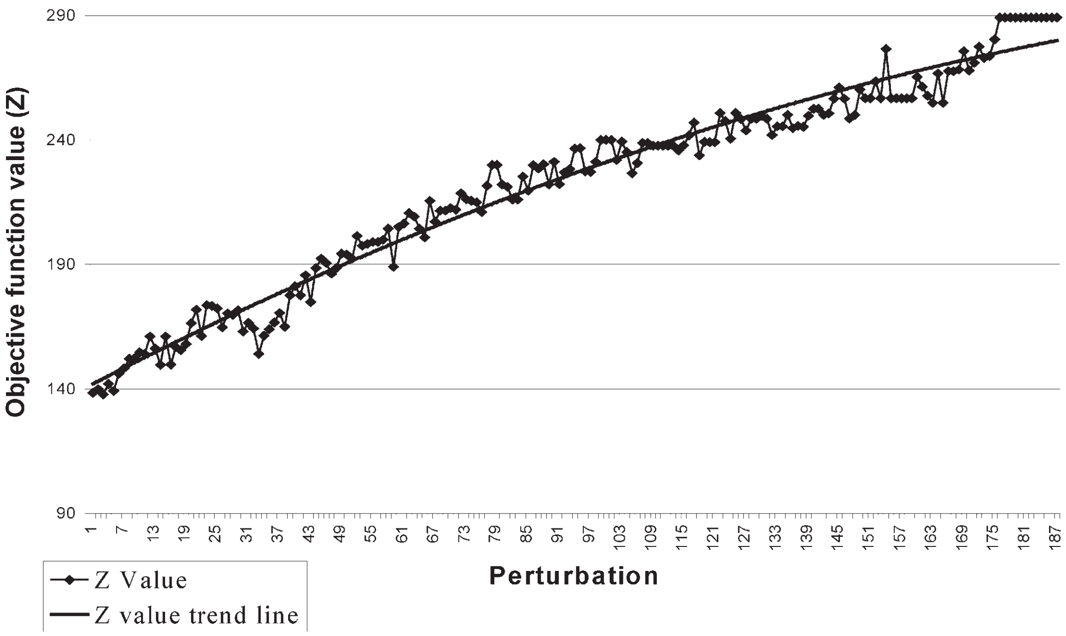

Fig. 2. SAA convergence.

Table 3. Modified ANOVA of orthogonal experiment.

Factor	D_f	SS	MS	F	At α = 5% significant/not significant
A	1	7.64	7.64	2.19	NS
B	1	43.43	43.43	9.61	S
D	1	19.66	19.66	4.35	NS
E	1	16.13	16.13	3.57	NS
Error	4	18.09	4.52		
Total	7	97.31			

and the degree of freedom required is 5. Hence, an L8 orthogonal array is selected. The first five columns are assigned the above five parameters, A, B, C, D, and E. The remaining columns represent the primary errors E1 and E2. Eight experiments are conducted as specified by the L8 table. For each experiment, SAA control parameter values are set as given in the L8 table and the Z^* values are determined using the simulation model. The Z^* value is the response variable. These values are shown in Table 2.

The analysis of variance is performed on these data and it is found that the F-statistic of factor A is reduced, its sum of squares is pooled with the error sum of squares, so that the number of factors become approximately equal to half of the columns. The modified ANOVA is given in Table 3.

The F-statistic for final temperature t_f is greater than $F_{0.05}$ with 1 and 4 degrees of freedom. Hence, it is significant and it is fixed at level 1. By considering the mean values, level 2 is fixed for the parameters cooling rate α and accept limit β. Initial temperature t_i and freeze limit Φ are fixed at level 1 by considering the number of perturbations which are proportional to the CPU time. The values thus fixed are given in Table 1 in bold letters.

Table 4. SAA based solution for a 2 part type and 5 workstation problem.

Station	Part type				NOM	PZD	RMTW	RMLT	Z^*	CPU time (time units)
	1		2							
	POK	WK	POK	WK						
1	1	1	4	2	4	99.8	97.52	96.82	294.16	53.23
2	1	1	4	3	2					
3	2	3	2	2	3					
4	3	1	2	1	2					
5	3	4	9	2	3					
LS		60		60						

5.3 Perturbation

Perturbation is the process of obtaining the neighbour of the current solution vector. From the current solution vector, an element (POK or WK or LS or NOM) is chosen at random. In the chosen element, one of the values is modified at random within its feasible range. Perturbation may also be applied to more than one element at a time in the solution vector. Larger lot sizes along with high POK values are not desirable since they lead to high WIP values. Hence, a check is made to avoid increase in both LS and POK for a part type within a perturbation.

6. SAA Based Kanban Search Procedure

Step 1. Input SAA parameters : Set t_i, t_f, Φ, β, α; $n = 0$

Step 2. Input kanban model parameters:
p = number of part types
s = number of work stations
demand distribution parameters
inter arrival distribution parameters for each of the p part types, process time distributions for each of the s workstations and material handling time distribution parameters.

Step 3. Generate initial solution S_0

Step 4. Call kanban system simulation module.
Perform simulation for the duration of 75 000 hours.
Compute Z^*.

Step 5. Set $f_0 = Z^*$; $f^* = f_0$; $S^* = S_0$

Step 6. Set $n = n + 1$

Step 7. Using perturbation determine S_n.

Step 8. Call kanban system simulation module.
Perform simulation for the duration of 75 000 hours.
Compute Z^*.

Step 9. Set $f_n = Z^*$; $S^* = S_0$

Step 10. Compute change in Z^* value
$\delta = f_n - f_{n-1}$

Step 11. If $\delta > 0$ then (accept the solution) goto step 12
else goto step 13.

Step 12. If $f_n > f^*$ then
(new global value) set $f^* = f_n$
$S^* = S_n$; $fz = 0$
goto step 16
else goto step 16.

Step 13. Compute probability of accepting inferior solution at step n
$p_{an} = e^{(-\delta/tn)}$

Step 14. If r (0,1) $< p_{ai}$ then (accept inferior solution) goto step 17
else goto step 15.

Step 15. $tot = tot + 1$;
if $t_n < = t_f$ then goto step 23
else goto step 7.

Step 16. If $t_n < = t_f$ then goto step 23
else goto step 17.

Step 17. $acc = acc + 1$; $tot = tot + 1$.

Step 18. If ($tot > 4 \times \beta$) or ($acc > \beta$) then goto step 19
else goto step 6.

Step 19. Compute $atr = acc/tot$.

Step 20. If $atr <= 0.15$ then $fz = fz + 1$

Step 21. If $fz >= \Phi$ then goto step 23.

Step 22. $t_{n+1} = t_n * \alpha$; $acc = 0$; $tot = 0$;
goto 6.

Step 23. End of SAA based kanban design parameters search algorithm.
Report Z^* and S^* .
STOP

7. Conclusion

An SA based search is used to set the parameters of a kanban-based production system, which is a pioneering work in this field. The exploration of SAA in the search space is shown in Fig. 1 production kanbans (POK) for part type 1 at workstation 2 is taken as an example, and the values assigned to it as the search progresses are plotted. This depicts the way in which the SA based search advances through the search space. The perturbation effect can be seen clearly seen by the manner in which the value of this parameter is explored. The variation is by only one unit on either side. The manner by which the SA based search converges is given in Fig. 2. Considering the convergence property, it can be seen that variation in the Z

value is rapid at the beginning of the search, and the rate of variation diminishes as the search progresses. This is mainly due to the higher probability of accepting inferior solutions at the early stages of the search, which become fixed as the search progresses.

A sample output of kanban system parameter selection using SAA is given in Table 4. The example considered is a moderate-sized problem with five workstations and two part types. The number of production kanbans (POK) and withdrawal kanbans (WK) to be provided in each station for each part type, is given in the table. The results show that the lot size for both the part types to be 60. Regarding performance of the model, the overall value is 294.16/300, which is equal to 98.72%. The PZD is 99.8% which means that 99.8% of the time, orders are satisfied form the output buffer without any waiting. The normalized values of lead time and WIP are also near to 100%, which in turn shows that they have very small values, which is desirable.

Acknowledgement

The authors thank the anonymous referees for their valuable comments, which when addressed, has made the paper complete.

References

1. D. Hutchins, Just in Time, Productivity Press (India), 1993.
2. B. J. Berkley, "A review of the kanban production control research literature", Production and Operations Management, 1, pp. 393–411, 1992.
3. W. Price, M. Gravel and A. L. A. Nsakanda, "Review of optimisation models of kanban based production systems", European Journal of Operational and Research, 75, pp. 1–12, 1994.
4. O. Kimura and H. Terada, "Design and analysis of pull system, a method of multi stage production control", International Journal of Production Research, 19, pp. 241–253, 1981.
5. G. R. Bitran and C. Chang, "A mathematical programming approach to a deterministic kanban system", Management Sciences, 33, pp. 427–441, 1987.
6. S. M. Gupta and Y. A. Y. Al-Turki, "Adapting just-in-time manufacturing systems to preventive maintenance interrupts", Production Planning and Control, 9, pp. 349–359, 1998.
7. J. Deleersnyder, T. J. Hodgson, H. Muller and P. J. O'Grady, "Kanban controlled pull systems: an analytic approach", Management Science, 35, pp. 1079–1091, 1989.
8. B. J. Berkley, "A decomposition approximation for periodic kanban controlled flow shops", Decision Sciences, 23, pp. 291–311, 1992.
9. V. S. Nori and B. R. Sarker, "Optimum number of kanbans between two adjacent stations", Production Planning and Control, 9, pp. 60–65, 1998.
10. B. J. Berkley, "Simulation tests of FCFS and SPT sequencing in kanban systems", Decision Sciences, 24, pp. 218–226, 1993.
11. B. J. Berkley and A. S. Kiran, "A simulation study of sequencing rules in a kanban – controlled flow shop", Decision Sciences, 22, pp. 559–582, 1991.
12. M. Savsar, "Effects of kanban withdrawal policies and other factors on the performance of JIT system-a simulation study", International Journal of Production Research, 34, pp. 2879–2899, 1996.
13. B. J. Berkley, "A simulation study of container size in two-card kanban system", International Journal of Production Research, 34, pp. 3417–3445, 1996.
14. A. Andijani, "Trade off between maximizing throughput rate and minimizing system time in kanban systems", International Journal of Operations and Productions Management, 17, pp. 429–445, 1997.