Available online at www.sciencedirect.com
ScienceDirect
Journal of the Franklin Institute 343 (2006) 521–531
Journal of The Franklin Institute
www.elsevier.com/locate/jfranklin

Design and analysis of production control scheme for Kanban-based JIT environment

Mohammad D. Al-Tahat[a,*], Adnan M. Mukattash[b]

[a] Industrial Engineering Department, University of Jordan, Amman 11942, Jordan
[b] Industrial Engineering Department, The Hashemite University, Zarka 13115, Jordan

Received 6 December 2005; accepted 27 January 2006

Abstract

This paper aims at designing production control scheme for Kanban-based just-in-time (JIT) environment. Synchronization mechanism for a single-stage single-product Kanban-controlled production line is developed, in such a way that yields a feasible operating cost as well as a feasible average of work-in-progress (WIP). The Kanban production line is formulated as $(M/M/s : GD/\infty/\infty)$ queuing model. A new approach for analyzing the queue model is discussed. An operating cost model is then developed to determine the unknown parameters of the system. Numerical examples are used to demonstrate the computations of different system parameters. Avenue for future research is also indicated.

Keywords: Kanban; Pull production systems; Shop floor control; Material control policies

1. Introducing the research problem

Partially completed products that are currently in the shop are referred to as work-in-progress (WIP). The WIP may be in a queue awaiting the availability of their next workstation, being loaded or processed on a machine, or being moved between workstations. In addition, WIP may exit because of parts being held in temporary storage pending approval from customer or instructions from engineering due to a design change,

*Corresponding author. Tel.: +962 6 535 5000x2793; fax: +962 6 535 5588.
E-mail addresses: altahat@ju.edu.jo, altahat@yahoo.com (M.D. Al-Tahat), adnanm63@yahoo.com (A.M. Mukattash).

0016-0032/$30.00
doi:10.1016/j.jfranklin.2006.01.001

or a decision on disposition resulting from a quality problem, or arrival of other parts/materials to be mated with these parts in a next production operation.

Design of a single-stage Kanban-controlled production line that produces a single-product type is considered as the problem of this work. This Kanban station contains WIP parts that are currently being processed. The raw material enters the workstation after receiving them from the suppliers, and the WIP, as flows through the line. The process of transporting WIP continues until the finished product departs the workstation and then delivered to the customer in time. The output rate of the workstation is generally dictated by the demand of the final products. An important executive concern is how to control and synchronize the flow of materials among the line; so as to create a consistent integrated shop floor control system that successfully meets the customer demand just-in-time (JIT). Several researches have contributed to the development of many approaches for solving such dilemma, among which the Kanban control policy that was invented in Toyota in the 1970s, and has since been widely used in industry [1]. Flexible Kanban system is discussed by [2]. The CONWIP control policy is described and analyzed by Gstettner et al. [3] and Wen et al. [4]. The base stock control policy is also described and analyzed by Scott et al. [5]. The generalized Kanban control system is introduced by Buzacott [6]. The generic Kanban systems for dynamic environment is modified by Chang et al. [7]. The extended Kanban control system is proposed and discussed by Claudine et al. [8], and many others. A comparison of these control policies as well as other policies can be found in Liberopoulos and Dallery [9] and Zhao et al. [10]. Comparisons of the performance of these policies indicate that no one strategy completely dominates the other [11]. Deep exploration of the tradeoffs using analytical models could help differentiate regions where each policy dominates over the other. Ananth et al. [11] stated that queuing models seem to be the most useful tool for such analysis. They modeled Kanban policy as a closed cyclic queuing network with manufacturing stations and fork/join synchronization workstations. Consequently, they proposed parametric decomposition approach and followed two-moment approximations to analyze the system. Stefan et al. [12] present a generalized analysis technique for queuing networks with mixed priority strategy and class switching; they show how to transform a queuing network that cannot be solved into a network model that can be solved using a standard analysis technique. A general purpose analytical method for performance evaluation of multistage single part Kanban-controlled production line is developed by Dallery [13]. The system is modeled as a queuing network. Original Kanban system is decomposed into a set of subsystems, then each subsystem is associated with a particular workstation and analyzed in isolation using a product from approximation technique.

Akturk et al. [14] present a literature review on the classification techniques of determining both the design parameters and Kanban sequences for JIT manufacturing systems. They also summarized the model structures, decision variables, performance measures and assumptions in a tabular format. A key result on the throughput equivalence has been derived by Paik et al. [15] for a general fork/join queuing network with finite buffers and GI service time distribution. Bhaba et al. [16] constructed a cost function that is developed based on the cost incurred due to the raw material, WIP between workstations, and the finished goods. They obtained an optimal number of raw material orders that minimize the total cost function, which used to find the optimal number of Kanban.

A minimum level of WIP yields lowers total inventory cost as well as lowers unit production cost. Therefore, minimizing WIP inventory level is supposed to be considered as an important strategy when design, operate, and/or manage any production system. Vijay et al. [17] estimate the total cost of WIP in a Kanban-controlled production network based on the shortage cost per Kanban per unit time and the unit holding cost. Another approach for estimating cost of WIP in a Kanban-controlled production is modeled by Bhaba et al. [16]; the expected total cost of WIP in this model consists of the cost of having a Kanban at the workstation and of holding the cost only.

2. Objective

The primary objective of this paper is to design a synchronized single-stage single-product Kanban-controlled production line in such a way that sustains a feasible WIP level in the system. Workstation production capacity and processing rate, utilization factor of the system, number of servers in the system, and the ordering rate of raw material are considered as the main design parameters.

3. Kanban philosophy

Kanban is a control mechanism that links production activities and transmitted demand information from final product buffers to the preceding workstation by using Kanban cards. As shown in Fig. 1, withdrawal of a product from the finished parts inventory by the demand triggers the system. As a full container is delivered to customer, the production ordering Kanban is detached from it and is posted at the production ordering Kanban post of the station to signal the need for production of equivalent numbers of units, which are withdrawn by customer.

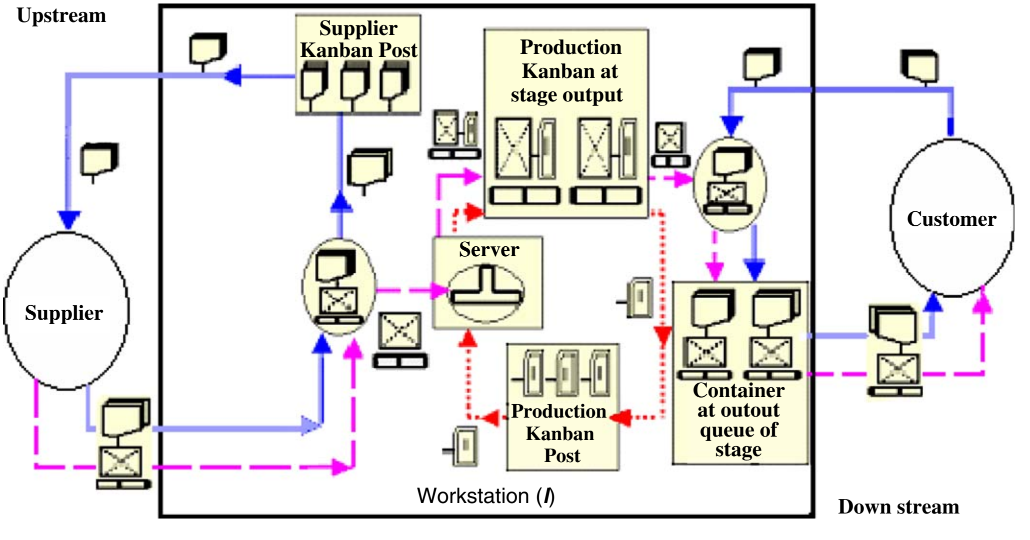

Fig. 1. Circulation of information in a single-stage single-product Kanban-controlled production line. Withdrawal and supplier Kanban circulate as shown by the solid line, production Kanban circulates as shown by the dotted line, and WIP flows as shown by the dashed line.

The workstation can start operation when (1) it is free, (2) there is a free withdrawal Kanban at the output queue, and (3) there is a supplier Kanban attached to a full container at the input queue (coming from supplier).

As the full container is withdrawn into the station for processing, the supplier Kanban is detached from it and is posted at the supplier Kanban post at the input queue of the workstation. Part carrier collects the supplier Kanban from its post at the workstation and moves it back to the supplier. If there is a full container at the output queue of the supplier, the supplier Kanban from the workstation is attached to the full container and is moved to the input queue of the workstation. As a result, products are processed according to the final demand coming to the line from the customer.

4. Notations

To model the different interactions of the system variables the following notation is used:

τ	lead-time demand or processing time for one container (unit time)
D_{C}	outside demand rate received from customer (units/unit time)
K_C	number of withdrawal Kanban and is equal to number of production Kanban
n_{K}	size of each withdrawal Kanban (units)
μ_{w}	processing rate for workstation (units/unit time)
s	number of servers installed in the workstation
ρ	utilization percentage of the overall workstation
λ_{w}	arrival rate of WIP into the workstation (units/unit time)
WIP	average unit of WIP exit in the production line at any time (units)
π_0	probability that all servers are idle
ℓ	lead-time, or average time product spends in the production line, or flow time
C_{sw}	cost of purchasing and ordering Raw material needed to meet demand rate (\$),
b	unit purchasing cost (\$/unit)
o	cost of having and transporting one supplier Kanban (\$/supplier Kanban)
K_S	number of supplier Kanban
C_{w}	processing cost of the overall demand (\$)
p	unit processing cost (\$/unit)
i	fixed installation cost per one server (\$/server)
C_{wc}	transportation cost of the demanded quantity to customer (\$)
c	cost of having and shipping one Kanban to customer (\$/withdrawal Kanban)
C_{wip}	holding cost of all WIP units in the system (\$)
h	unit holding cost (\$/unit/unit time)
Toc	total operating cost of the system to produce the desired demand (\$)

5. Assumptions

The approach introduced here has been presented according to the following assumptions: at any time, there will be no demand shortage in the system. The shortage cost of finished product is zero; since the line considered produces only one type of product, the set-up cost and changeover time are neglected; zero safety lead-time factor; there is no safety stock in the considered line; and system workstation queue is modeled as $(\mathrm{M/M/s: GD}/\infty/\infty)$.

6. Queuing formulation of the system

The workstation of the line is modeled as $(\mathrm{M/M/s: GD}/\infty/\infty)$ queue and analyzed using the formulas of $(\mathrm{M/M/s: GD}/\infty/\infty)$ queue.

As shown in Fig. 2, servers receive demand from customer; consequently, supplier receives demand from servers in a rate equivalent to the total servicing rate of the workstation. Generally, we can express the processing rate of the workstation as the demand rate coming from outside customer. The servicing rate of the workstation and customer demand rate can be compared with the arrival and service rates of a typical queuing system. In the queuing system, customers arrive at the arrival rate and wait to be serviced, after which they depart. Similarly, for the system proposed here materials are produced at a production rate of (R_i, production rate of producing product on the workstation i) and depart with demand rate of customer. The conditions of a pull system require that materials exit production line only at the rate desired by customers. Hence, demand rate for customer serves as the rate at which WIP departs the workstation and so the demand rate is equivalent to the service rate in a queuing system (production rate in Kanban system R_i).

WIP results from the flow of each unit through the system during processing. Although we can try to place bounds on this level of WIP, it is difficult to estimate it precisely. For example, at a minimum, each unit remains at the workstation long enough to produce its container quantity to the extent that queuing time exists at workstation. This time will be longer, if the lead-time demand (processing time of one container) for producing of a specific product on the system is τ, and then shortage will be avoided if we set the target inventory to equal the lead-time demand (τ, D_C) i.e.

$$(K_C)(n_\mathrm{K}) \geqslant (\tau D_\mathrm{C}). \tag{1}$$

Number of Kanban implemented in the line can be expressed as

$$K_C \geqslant \frac{\tau . D_\mathrm{C}}{n_\mathrm{K}}.$$

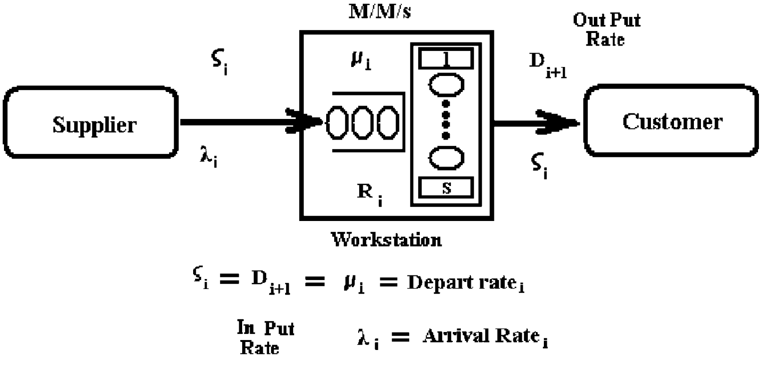

Fig. 2. A single queue model of single-stage Kanban system.

And hence, the minimum number of Kanban to ensure replenishment before lead-time demand is consumed as

$$K_C = \frac{\tau . D_C}{n_K}, \tag{2}$$

$$\tau = \frac{n_K}{\mu_w}. \tag{3}$$

To synchronize the production in the single Kanban workstation shown in Fig. 2, the workstation service rate must achieve the output demand with the minimum inventory level, i.e. service rate must equal customer demand rate

$$\mu_w = D_C. \tag{4}$$

With utilization rate for the workstation of ρ and S identical parallel servers, and to maintain the synchronization feature of the system, the arrival rate to the servers (the processing rate of the supplier) must be

$$\lambda_w = (s)(\mu_w)(\rho). \tag{5}$$

Based on queuing theory principles, the workstation is modeled as for (M/M/s : GD/∞/∞) queue. Average WIP units exit in the production line at any time, and lead-time ℓ (average time product spends in the *system* or flow production time) are computed as in Eqs. (6) and (8), respectively,

$$WIP = \left[\frac{(s\rho)^2(\pi_0)}{(s)!(1-\rho)} \cdot \frac{1}{\lambda_w(1-\rho)} + \frac{1}{\mu_w}\right], \tag{6}$$

$$\pi_0 = \frac{1}{\sum_{n=0}^{s-1} (s\rho)^n/n + (s\rho)^s/s!(1-\rho)}, \tag{7}$$

$$\ell = \frac{(s\rho)^2 \pi_0}{s!(1-\rho)} \cdot \frac{1}{s\mu_w - \lambda_w} + \frac{1}{\mu_w}. \tag{8}$$

As a result in Kanban-controlled production system, customer demand must be equivalent to the maximum products that can be produced (production capacity) by the workstation. The above formulas are used to find the minimum WIP of a product flow in a single Kanban workstation under no safety stock, the case that has been assumed by this work.

7. Operating cost of the system

The operating cost consists of four main components parts: the first one, C_{sw}, which is the cost elements constituting raw material cost and ordering cost; these are unit purchasing cost of raw material and is given by ***b*** ($/purchased unit) and cost of having and transporting one supplier Kanban to supplier (ordering cost) and is given by ***o*** ($/supplier Kanban).

The second cost component, C_w, is the cost elements of manufacturing. This implies unit processing cost and is given by ***p*** ($/unit) and a fixed amount of cost equal to the cost of installing one server in the workstation given by ***i*** ($/server). The next cost component, C_{wc} is the transportation cost of the demanded quantity to customer ($). Transportation cost depends on the cost of having and transporting one Kanban to customer, ***c*** ($/Kanban).

The last cost component is, C_{wip}, represents the total holding cost of WIP in the system and *is* estimated by the average unit holding cost ***h*** (\$/units/unit time).

It is necessary to stress in this section of the work that the system has been designed to avoid shortage of finished product (shortage cost is zero) and a single item is assumed (set-up cost is neglected).

The total operating cost *Toc* can be written as

$$Toc = C_{\mathrm{sw}} + C_{\mathrm{w}} + C_{\mathrm{wc}} + C_{\mathrm{wip}}, \tag{9}$$

where C_{wip}, C_{wc}, C_{w}, and C_{sw} are expressed in Eqs. (10), (11), (12), and (13), respectively,

$$C_{\mathrm{wip}} = (h).(WIP).(\ell), \tag{10}$$

$$C_{\mathrm{wc}} = (c).(K_C), \tag{11}$$

$$C_{\mathrm{w}} = (p).(D_{\mathrm{C}}) + s(i), \tag{12}$$

$$C_{\mathrm{sw}} = \lambda_{\mathrm{w}}.(b) + K_S.(o), \tag{13}$$

$$K_S = \frac{\lambda_{\mathrm{w}}}{n_s}. \tag{14}$$

Values of λ_{w}, WIP, and ℓ are obtained by Eqs. (5), (6) and (8), respectively, and n_s can be assumed as n_{K}. After substitution, the overall cost model can be presented as

$$Toc = [(h).(WIP).(\ell)] + [(c).(K_C)] + [(p).(D_C) + s(i)] + [\lambda_{\mathrm{w}}.(b) + K_S.(o)]. \tag{15}$$

Numerical example 1: A Kanban-controlled workstation receives an outside customer demand rate of 12 units per day. If the workstation own a material handling equipment of maximum load size equal to 6 parts per day and if the workstation includes 2 servers with utilization rate approach to 100% (ideal case), then

$$\mu_{\mathrm{w}} = D_{\mathrm{C}} = 12 \text{ parts per day},$$

$$\tau = \frac{n_{\mathrm{K}}}{\mu_{\mathrm{w}}} = \frac{6}{12} = 0.5 \text{ days},$$

$$K_C = \text{first integer value greater than } \frac{\tau.D_{\mathrm{C}}}{n_{\mathrm{K}}} = 2.$$

Same result, as shown in Fig. 3, is obtained by OM Explorer which is a set of Microsoft Excel templates accompany with operation management textbook by Krajewski et al. [18] for solving common problems in Operations Management

$$\lambda_{\mathrm{w}} = (s)(\mu_{\mathrm{w}})(\rho) = (2)(2 \times 6)(1) = 24 \text{ parts per day},$$

$$\pi_0 = 0 \text{ (probability that all servers are idle is zero)},$$

$$WIP = \left[\frac{1}{\mu}\right] = \frac{1}{12} = 0.0833 \text{ units},$$

$$\ell = \tfrac{1}{12} = 0.0833 \text{ day}.$$

The above values indicate that the workstation is working full capacity all the time with 100% utilization rate. In real life this is so difficult and theoretical because of many

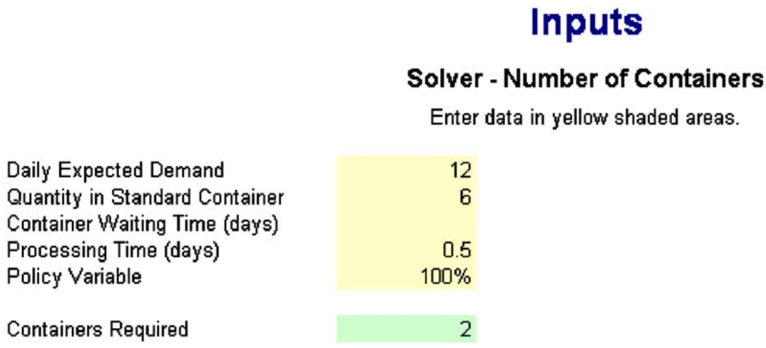

Fig. 3. Computing number of containers by OM explorer [18].

Table 1
Analysis of single-stage Kanban model at constant utilization rate

Scenario	s	λ_w	μ_w	λ_{eff}	p_O	WIP	Lq	ℓ	Wq
Design of single Kanban stage comparative analysis									
1	3	24.000	12.000	24.000	0.111	2.889	0.889	0.120	0.037
2	4	24.000	12.000	24.000	0.130	2.174	0.174	0.091	0.007
3	5	24.000	12.000	24.000	0.134	2.040	0.040	0.085	0.002
4	6	24.000	12.000	24.000	0.135	2.009	0.009	0.084	0.000
5	7	24.000	12.000	24.000	0.135	2.002	0.002	0.083	0.000
6	8	24.000	12.000	24.000	0.135	2.000	0.000	0.083	0.000
7	9	24.000	12.000	24.000	0.135	2.000	0.000	0.083	0.000

constraints. Therefore, system design must be modified either by considering a low utilization or increasing the number of servers.

The impact of increasing the number of servers for a constant utilization rate on the average WIP units for the previous example is shown in Table 1 and Fig. 4, respectively.

Results obtained by TORA optimization system, version 1, H.A. Taha [19]. On the other side, the impact of changing utilization rate on the average number of WIP for a fixed number of servers (variable arrival rates) is shown in Table 2 and Fig. 5, respectively.

Obviously one can conclude that the best design is a comprise solution that combine a good number of server (s) with a proper utilization rate.

Numerical example 2: For $\boldsymbol{b} = \$3/\text{unit}$, $\boldsymbol{o} = \$15/\text{supplier}$ Kanban, $\boldsymbol{p} = \$20/\text{unit}$, $\boldsymbol{i} = \$200/\text{server}$, $\boldsymbol{c} = \$20/\text{withdrawal}$ Kanban, and $\boldsymbol{h} = \$5/\text{units/unit time}$, total operating cost can be computed by Eq. (15) for many scenarios as in Table 3. The solutions above are obtained by "E-Z solver™", version 1, an engineering equation solving and analysis tools, developed by intellipro, Inc., and published in 1998 by John Wiley and Sons. The data in Table 3 can be figured out as shown in Fig. 6.

8. Conclusions

The Japanese Kanban technique to achieve the goal of lean production environment has been observed to minimize the work-in-progress (WIP) as well as to minimize inventory

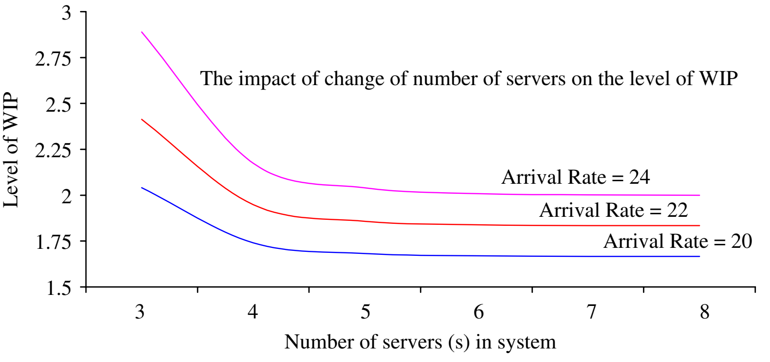

Fig. 4. Impact of numbers of servers on average WIP in the system.

Table 2
Analysis of single-stage Kanban model for different utilization factor (when $s = 3$)

Scenario	s	λ_w	μ_w	λ_{eff}	p_O	*WIP*	*Lq*	ℓ	*Wq*
Design of single Kanban stage comparative analysis									
1	3	35.280	12.000	35.280	0.005	50.100	47.160	1.420	1.337
2	3	30.600	12.000	30.600	0.040	06.689	4.139	0.219	0.135
3	3	25.200	12.000	25.200	0.096	03.249	1.149	0.129	0.046
4	3	19.800	12.000	19.800	0.176	02.008	0.358	0.101	0.018
5	3	14.400	12.000	14.400	0.294	01.294	0.094	0.090	0.007

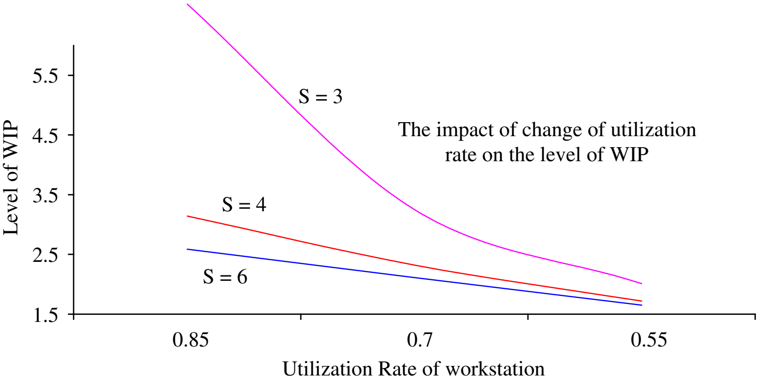

Fig. 5. Impact of utilization factor on average WIP.

cost. The research issue considered here is one of the problems faced today in Kanban-controlled production lines. The research has considered a queuing synchronizing mechanism that simulates the interaction of the different system parameters together. The research has proved that customer demand rate, and available container size dictate

Table 3
Total operating cost for different system scenarios

Scenario	*Toc*	*WIP*	ℓ	K_C	D_C	λ_w	K_S	s
1	2071	50	1.4	5	21	35	6	5
2	1657	40	1.2	4	18	34	5	4
3	1260	30	1.0	3	15	30	4	3
4	880	20	0.8	2	12	25	3	2
5	540	10	0.6	1	10	20	2	1

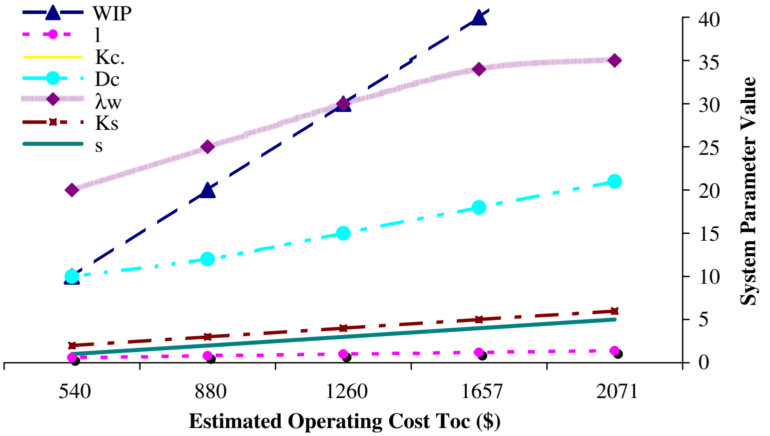

Fig. 6. Impact of system parameter on total operating cost *Toc*.

the values of the system parameters, such as, the value of the number of Kanban needed in the system, system capacity, servers needed at a specific utilization rate, number and sizes for orders of raw material needed from out side supplier. All of these parameters are dependent on one another and they dictate the system operating cost together. These results have been raised through the analytical analysis of two numerical examples.

The present study is limited to only a single-stage Kanban-controlled production line; it may be worthwhile to focus future research on the issues of multi-stage serial production line for single-product.

References

[1] Y. Monden, Toyota Production System: Practical Approach to Production Management, Industrial Engineering and Management Press, Narcross, GA, 1983.
[2] S.M. Gupta, Y.A.Y. Al-Turki, R.F. Perry, Flexible Kanban system, Int. J. Oper. Prod. Manage. 19 (10) (1999) 1065–1093.
[3] S. Gstettner, H. Kuhn, Analysis of production control systems Kanban and CONWIP, Int. J. Prod. Res. 34 (11) (1996) 3253–3273.
[4] W. Zhang, M. Chen, A mathematical programming model for production planning using CONWIP, Int. J. Prod. Res. 39 (12) (2001) 2723–2734.
[5] S.E. Grasman, T. Lennon Olsen, J.R. Birge, Setting base stock levels in multi-product systems with setups and random yield, University of Missouri-Rolla, Rolla, Missouri 65409-0370, 2003.

[6] J.A. Buzacott, Queuing models of Kanban and MRP controlled production systems, Eng. Cost Prod. Econ. 17 (1989) 3–20.
[7] T.M. Chang, Y. Yih, Generic Kanban systems for dynamic environments, Int. J. Prod. Res. 32 (4) (1994) 889–902.
[8] C. Chaouiya, G. Liberopoulos, Y. Dallery, The extended Kanban control system for production coordination of assembly manufacturing systems, IIE Trans. 32 (2000) 999–1012.
[9] G. Liberopoulos, Y. Dallery, A unified framework for pull control mechanisms in multi-stage manufacturing systems, in: J. MacGregor Smith, S.B. Gershwin, C.T. Papadopoulos (Eds.), Annals of Operations Research, vol. 93: Performance Evaluation and Optimization of Production Lines, Baltzer Science, Bussum, The Netherlands, 2000, pp. 325–355.
[10] Z. Xiaobo, Q. Gong, J. Wang, On control strategies in a multi-stage production system, Int. J. Prod. Res. 40 (5) (2002) 1155–1171.
[11] K. Ananth, S. Rajan, V. Mary, Two moment approximations for fork/join stations with applications to modeling Kanban system. Proceeding of the 12th Conference on Production and Operations Management Society, POM-2001, March 30–April 2, 2001, Orlando, FI.
[12] S. Greiner, G. Bolch, K. Begain, A generalized analysis technique for queuing networks with mixed priority strategy and class switching, Comput. Commun. 21 (1998) 819–832.
[13] Y. Dallery, An analytical method for performance evaluation of Kanban controlled production system, Oper. Res. 44 (1) (1996).
[14] M.S. Akturk, F. Erhun, An overview of design and operational issues of Kanban systems, Int. J. Prod. Res. 37 (17) (1999) 3859–3881.
[15] C.H. Paik, D.W. Tcha, Throughput equivalences in fork/join queuing networks with finite buffers and general service times, Int. J. Prod. Res. 33 (3) (1995) 695–703.
[16] B.R. Sarker, C.V. Balan, Operations planning for a multi-stage Kanban system, Eur. J. Oper. Res. 112 (2) (1999) 284–303.
[17] V.S. Nori, B.R. Sarker, Optimum number of Kanban between two adjacent stations, Production Plann. Control 9 (1) (1998) 60–65.
[18] L.J. Krajewski, L.P. Ritzman, A CD ROM attached to Operation Management Strategy and Analysis Textbook, sixth ed., Prentice-Hall, Englewood Cliffs, NJ, 2002 ISBN 0-03-042356-4.
[19] H.A. Taha, A CD ROM attached to Operations Research An Introduction, seventh ed., Prentice-Hall, Englewood Cliffs, NJ, 2003.