

Optimal Work-In-Process control for a closed multistage production system with machine preference

Bacem Samet, Florent Couffin, Marc Zolghadri, Maher Barkalla, Mohamed Haddar

► To cite this version:

Bacem Samet, Florent Couffin, Marc Zolghadri, Maher Barkalla, Mohamed Haddar. Optimal Work-In-Process control for a closed multistage production system with machine preference. Design and Modeling of Mechanical Systems-III, pp.1051-1059, 2018, 7th Conference on Design and Modeling of Mechanical Systems (CMSM 2017), 2195-4356. 10.1007/978-3-319-66697-6_103 . hal-01644071

HAL Id: hal-01644071

<https://hal.archives-ouvertes.fr/hal-01644071>

Submitted on 21 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Work-In-Process control for a closed multistage production system with machine preference

Bacem Samet^{1/2}, Florent Couffin¹, Marc Zolghadri¹, Maher Barkalla², Mohamed Haddar²

¹ Laboratoire QUARTZ, Institut supérieur de mécanique de Paris

sametbacem@gmail.com

florent.couffin@supmeca.fr

marc.zolghadri@supmeca.fr

² Laboratoire de Recherche de Mécanique, Modélisation et Productique (LA2MP), Ecole Nationale d'Ingénieur de Sfax

bark_maher@yahoo.fr

mohamed.haddar@enis.rnu.tn

Abstract. In this work, we expose a closed loop production system which have great similarity with the Kanban system as the Work In-Process (WIP) is constant (CONWIP). This system is a multistage single-product and it support more than one machine in every stage. These machines can have different characteristics and they have limited capacity of buffering. We consider also that there is a preference for a machine usage in each stage. We use a Closed Queueing Network (CQN) with a repetitive blocking mechanism to model this system. The used resolution approach is the Maximum Entropy Method (MEM). In a first place, we look for the optimal WIP to keep in this system with respecting a production rate. In a second place, the routings to machines in the same stage are evaluated regarding the buffering capacity and the machine characteristic. We present some numerical results on a simple two stage production system which holds the same number of batches looping. Interesting perspectives for the development of the presented model are looked over.

Keywords: closed queueing network, work in process, conwip, entropy maximization, multi-stage production system.

1 Introduction

Manufacturing industry tends to be highly competitive. Firms are faced with the strict requirements of the customers. There are preference for some performance criteria such as quality of the product, respect of the delivery time and the competitive price of the goods. Since the manufacturing system is expected to respond to all these demands, it's an obligation to respect the required quality and to supply the quantity demanded and to reduce the cost of production.

Complexity of the production system rises from two main factors. In one hand there is a stochastic nature of the attributes of the system which can be internal variables as the service time of the machine or the breakdown periods of the machine, and external variables as the quantity demanded by the customers and raw materials prices. In the other hand, the operations in this system are interconnected; for example when the inventory is increased the production rate is impacted and the cost of the production too. So, there are compromises to set up to respect the customer demand with minimum cost.

Many studies have made a scope on the production system as a field of improvement and optimization. A particular type of production system; the multi-stage production system which presents a wide range of manufacturing system, was analyzed in (Muckstadt & Roundy, 1993). In such system many actions can be carried out such as allocation of service load (or service rate) which is looking for optimal service rate in different machines in the production line system. For instance, in (S . Hillier & W . Boling, 1979), the optimal allocation of work is searched with respect to the number of work stations in the line, the limit on the amount of work-in-progress and the variance of station operation times. The arrangement of stages which is ordering the stages so that the throughput is maximized is studied in (Yamazaki, Sakasegawa, & J . George, 1992). It is noted that in every stage there are a number of identical servers. Another problem addressing the assignment of servers was studied in (Ding & Greenberg, 1991), in this problem the single servers are optimally arranged for a tandem queueing system. The allocation of buffers problem which deals with searching the optimal buffering space in a production line to achieve a specific objective, is surveyed in (Yamazaki et al., 1992).

Reducing the quantity of WIP contributes to cost reduction by virtue of suppressing the costs of keeping useless inventories. The production control system is a very important leverage that should be well chosen to meet these performances. The CONWIP, a pull production control system, is defined as a system maintaining constant the maximum amount of WIP (Framinan, González, & Ruiz-Usano, 2003). It was first time defined in (Spearman, Woodruff, & Hopp, 1990) to minimize the WIP in a single product flow shop subject to a given throughput level. This same purpose was also conducted in (Asbjør M Bonvik & Gershwin, 1996)

and (A. M. Bonvik, Couch, & Gershwin, 1997). These aforementioned works are simulation based models. The queueing theory was exploited for analysis and performance evaluation. The application of the queueing theory was surveyed in (Govil & Fu, 1999). This analytical approach of modeling was considered in (Gstettner & Kuhn, 1996), particularly a closed queueing network with unlimited queueing buffers was established for a Conwip flow shop single product system and the optimal WIP with respect to a certain throughput level is sought.

We have noticed a limitation in the literature, lying in the study of a particular case which is the machine preference in every stage. This constraint was integrated in a scheduling architecture made in (Fox & Smith, 1984). It was also studied in (Baskar & Anthony Xavier, 2014) with the purpose to improve the percentage utilization of a priority machine. When there are differences in the machines in a stage, there would be a preference to totally occupy a machine than another. This preference can be due to accuracy of the machine by making less defected products. Ordering machines for priority of usage can be made for greater service rate or broadly for effectiveness of the particular machine.

The direction of this work is to determine the optimal quantity of WIP in the closed loop multi-stage production system with taking in consideration a production level rate as a target and taking into account the limited capacity buffers. Furthermore, the machine preference in every stage is considered. We propose a CQN model with limited buffering and blocking for a flow shop multistage production where the number of batch in the system remains constant. This closed loop production system can be ruled by a CONWIP control system since the quantity of WIP is the same. The remainder of this paper is organized as follows: in section 2, we describe the under scope type of production system. In section 3, we develop the CQN model for this system. In section 4, we present the approximate resolution approach based on Entropy Maximisation. In section 5, numerical results are exposed for a simple case. Finally, some interesting future developments for this model are revealed.

2 The multi-stage production system

The production system under study is a multi-stage flow, supporting multiple machines in every stage. We consider the case when there is an order of preference for the machines in the same stage. This system is a closed loop and so the number of batches (WIP) circulating in the system remains the same. For instance in figure 1, we expose a two-stage production system. In every stage, two parallel machines doing the same process and we presume a priority machine constraint for processing the job (i.e. priority of using M1 rather than M2 in stage 1 and priority of using M3 rather than M4 in stage 2). If the priority machine will be fully occupied and when its buffer is overloaded, the jobs are redirected to the other

machine. We suppose having an Unloading/Loading (U/L) process closing the cycle of production. This process receives the full batches, empties them and returns them empty to the flow line.

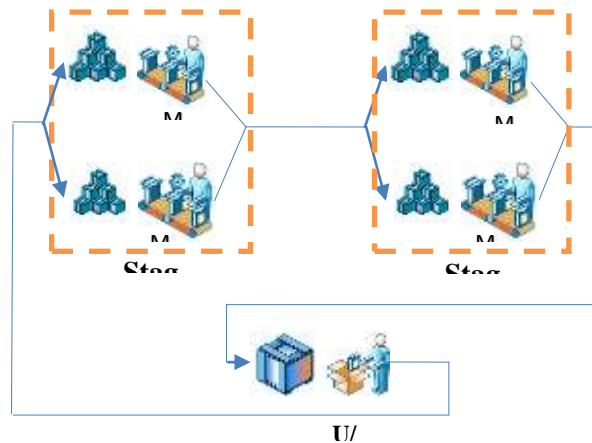


Figure 1. Two-stage production system

We seek to occupy maximally the priority machines (M1 and M3). In counterpart, there are constraints on the buffering capacity of such machines. As it is targeted to rout the maximum amount of jobs to M1 than to M2 for instance, the percentage of the quantity routed that fits in its buffering capacity should be found. The probability that a job is rejected by a full buffer of M1, indicates the convenience of the routing percentage. Keeping a production level with the minimum quantity of WIP is monitored by the throughput of the production system as a function of the WIP. So further in this work, the probability of being rejected in M1 and the throughput of the system are calculated thanks to the CQN. We present in the next section, this model with a blocking mechanism for this production architecture.

3 Closed queueing network model with RS-RD blocking

We consider L as the quantity of WIP. We model this system by a CQN with L jobs/customers presented in figure 2. Since every machine have a limited capacity buffer space, the queueing system support a blocking mechanism defined as Re-

petitive Service-Random Destination (RS-RD). This blocking defines the mechanism of a job finished being served in a queue i and intending to enter a full buffer of a downstream queue j . In the RS-RD the job loopback to queue i to undergo a new service time and this dynamics is repeated until the job finds a free buffer space in a downstream queue to which it is routed (Bose, 2002).

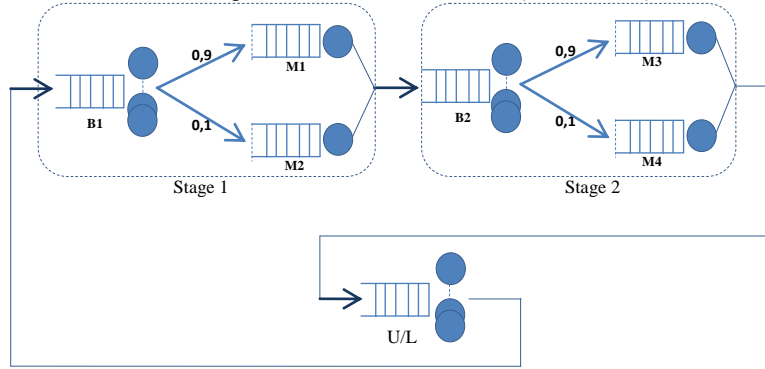


Figure 2: Closed Queuing Network model for two-stage production system

Every machine is modeled by a single server node (M_i) (for example, M_1 and M_2 in stage 1 in figure 1) with limited capacity buffering. We introduce in the upstream of every station a multi-server node (B_i) (for example, B_1 in stage 1 and B_2 in stage 2) which will play the role of the router of the job when the buffer of the downstream machine is overloaded. The number of the servers in (B_i) is L . In this manner, if all the jobs are directed to this type of node, they can be held there. In this type of node, the service time is taken very small compared to the service time of the machines. The service times of all the nodes have a general distribution which can be defined by a mean time (μ_i) and a squared coefficient of variation (sqv_i) in node i . The routing matrix is a key element which defines the probabilities between the nodes of the network (e.g. see figure 2 for the routing between B_1 and M_1). This CQN have the underlying state space:

$$S = \{(n_1, \dots, n_M) : \sum_{i=1}^M n_i = L, 0 \leq n_i \leq L_i, i = 1, \dots, M\} \quad (1)$$

With n_i is the number of jobs in the node i and L_i is the capacity of the node i . The resolution of a queuing network is the establishment of the steady state probability of occurrence of this state. In the underlying section we describe the resolution approach.

We exploit the MEM for solving such a CQN. This method seek to attribute the same chances of occurrence for every state, with respect to normalization and some marginal constraints (Kouvatsos & Xenios, 1989). The Entropy function

(eq.2) is maximized subject to the related constraints. We refer to (Kouvatsos & Xenios, 1989) for the details of this resolution.

$$H_p(n) = -\sum_n p(n) \log p(n) \quad (2)$$

The Lagrange multipliers resolution method gives a product form solution of the CQN. This product form establishment allows facilitation to the network resolution. In the case of an open network with Repetitive-Service Blocking (Kouvatsos & Xenios, 1989), the product form is established and afterward the probability of a state of the network is calculated through the marginal probability of the queues in isolation.

The resolution of the CQN goes through two main steps. The first step is about the resolution of the pseudo-open network corresponding to the CQN. It is an open network of queues with no arrival nor departure of jobs and it has the same parameters (service time characteristics and routings) of the CQN. The network is decomposed to deduce the mean rates and the squared coefficient of variations of the general distribution of both the service times and the inter-arrival times in the nodes. The MEM is applied then individually to these queues. The second step of resolution consists of injecting the Lagrange coefficients of the ME solution for individual queues deduced from the first step to the CQN. An iterative convolution algorithm is used to find marginal probabilities of the CQN. And finally, a correction of several parameters (Lagrange coefficients of the pseudo-open network) is done to insure that the throughputs respect the flow balance equations (eq.3).

$$X_i = \sum_{j=1}^M \alpha_{ji} X_j \quad (3)$$

With X_i is the expected throughput of node i and α_{ji} the routing probability from node i to node j .

The outputs of this resolution are probabilities of the states (of the set S). In our case we seek two performance indicators as outputs of this algorithm: the production rate of the production system which is the throughput of the Unloading/Loading machine and secondly, the blocking probability entering to a particular machine.

4 Numerical Results

In this section, the system presented in the queueing model in figure 1 is evaluated. The service times are generally distributed with rates and squared coefficient of variation indicated in table 1. The production rates of the priority machines (M1, M3) are higher than the other machines to imply the quality of rapidity in these machines. The squared coefficient of variation (sqv=1) characterize an exponential distribution for the service time.

Table 1. Service time characteristics.

Nodes	General Distribution	
	Rate of the service time	sqv
M1, M3	0.5	1
M2, M4	0.4	1
B1, B2	60	1
Unload/Load	60	1

Buffer capacity of M_2 and M_4 is taken sufficiently larger (capacity $M_2=14$) than those of the priority machines. From the resolution, we sketch firstly the production rate of the system which is the throughput of the (Unload/Load) process (figure 3). And secondly, we exploit the steady state blocking probability entering to M_1 (figure 4) which is an indicator of the choice of the routing of jobs in the machines of the stage 1.

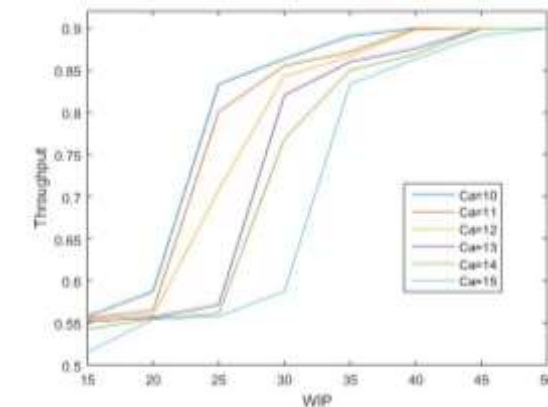


Figure 3. Throughput as a function of the WIP

In figure 3, we present the evolution of the throughput of the system as the circulating WIP increases. This evolution is sketched for different capacities of M1 and M3 (with capacity M1= capacity M3= ca). So every curve corresponds to a certain capacity. To have a particular throughput level (e.g. 0.8 (job/h)), the quantity of WIP that should be kept in the system is deduced (e.g. it's sufficient to keep 23 jobs in the network for ca=10).

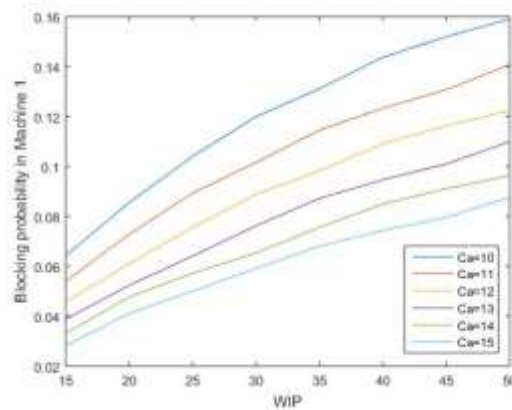


Figure 4. Blocking probability in M1 as a function of the WIP

From figure 4, the amount of jobs that can be blocked trying to enter M1 is deduced, for e.g. we know that almost 10% of jobs entering to M1 (when the WIP=23 and for ca=10) are blocked so with such buffer size of M1 it is advised to rout only 90% of jobs to M1 and the rest to M2. The same reasoning can be conducted to machines of stage 2.

These presented curves show a coherence of the results of this model. Then the presented model and the used analyses indicators tend to be a useful decision tool for the under-study production system.

5 Conclusion

A simple multi-stage CONWIP flow shop production system for a single product is modeled by a closed queueing network supposing RS-RD blocking mecha-

nism. This model gives an insight into the system dynamics and performance analysis regarding control and design changes are possible. The underlying queueing framework was exploited to look for the routings to machines when a preference for a machine use is present in a stage. The optimal number of WIP is sketched regarding a target throughput or the maximum reachable throughput.

Other performance indicators can also be calculated, tanks to this model, as the mean response time being enqueued in a machine or the mean number of jobs in a station.

The generalization to a multiproduct closed system is interesting as a further development for more general cases. Introducing the batch processing, breakdown delays for the service and set-up delays between different products workloads can also be valuable. Another perspective is to add the fork/join nodes in the network for more generalization to study assembly lines.

References

- Baskar, A., & Anthony Xavier, M. (2014). Optimization of makespan in job and machine priority environment. *Procedia Engineering*, 97, 22–28. <http://doi.org/10.1016/j.proeng.2014.12.220>
- Bonvik, A. M., Couch, C. E., & Gershwin, S. B. (1997). A comparison of production-line control mechanisms. *International Journal of Production Research*, 35(3), 789–804. <http://doi.org/10.1080/002075497195713>
- Bonvik, A. M., & Gershwin, S. B. (1996). Beyond Kanban: Creating and analyzing lean shop floor control policies. *Manufacturing and Service Operations Management Conference Proceeding*. Retrieved from <http://w3.mit.edu/manuf-sys/www/amb.msom040496.pdf>
- Bose, S. K. (2002). *An Introduction To Queueing Systems* (1st editio). New York: Kluwer Academic/ Plenum Publishers. <http://doi.org/10.1007/978-1-4615-0001-8> ISBN
- Ding, J., & Greenberg, B. S. (1991). Bowl Shapes Are Better with Buffers—Sometimes. *Probability in the Engineering and Informational Sciences*, 5(2), 159. <http://doi.org/10.1017/S0269964800001996>
- Fox, M. S., & Smith, S. F. (1984). ISIS--A knowledge-based system for factory scheduling. *Expert Systems*, 1(1), 25–49. <http://doi.org/10.1111/j.1468-0394.1984.tb00424.x>
- Framinan, J. M., González, P. L., & Ruiz-Usano, R. (2003). The CONWIP production control system: Review and research issues. *Production Planning & Control*, 14(3), 255–265. <http://doi.org/10.1080/0953728031000102595>
- Govil, M. K., & Fu, M. C. (1999). Queueing theory in manufacturing: A survey. *Journal of Manufacturing Systems*, 18(3), 214–240. [http://doi.org/10.1016/S0278-6125\(99\)80033-8](http://doi.org/10.1016/S0278-6125(99)80033-8)
- Gstettner, S., & Kuhn, H. (1996). Analysis of production control systems kanban and CONWIP. *International Journal of Production Research*, 34(11), 3253.

<http://doi.org/10.1080/00207549608905087>

- Kouvatsos, D. D., & Xenios, N. P. (1989). MEM for Arbitrary Queueing Networks with Multiple General Servers and Repetitive-service Blocking *, *10*, 169–195.
- Muckstadt, J. a, & Roundy, R. O. (1993). *Handbooks in Operations Research and Management Science. Logistics of Production and Inventory*. (S. C. Graves, A. H. G. Rinnooy Kan, & P. H. Zipkin, Eds.) *Logistics of Production and Inventory* (Vol. 4). North Holland. [http://doi.org/http://dx.doi.org/10.1016/S0927-0507\(05\)80182-3](http://doi.org/http://dx.doi.org/10.1016/S0927-0507(05)80182-3)
- S . Hillier, F., & W . Boling, R. (1979). On the Optimal Allocation of Work in Symmetrically Unbalanced Production Line Systems with Variable Operation Times. *Management Science*, *25*(8), 721–728.
- Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP, a pull alternative to kanban. *International Journal of Production Research*. <http://doi.org/10.1080/00207549008942761>
- Yamazaki, G., Sakasegawa, H., & J . George, S. (1992). On Optimal Arrangement of Stations in a Tandem Queueing System with Blocking. *Management Science*, *38*(1), 137–153. Retrieved from <http://www.jstor.org>