The Effect of the Workload on Due Date Performance in Job Shop Scheduling

Miguel Cezar Santoro
Departamento de Engenharia de Produção,
Escola Politécnica, Universidade de São Paulo – USP, São Paulo, SP, Brazil
E-mail: santoro@usp.br

Marco Aurélio de Mesquita
Departamento de Engenharia de Produção,
Escola Politécnica, Universidade de São Paulo – USP, São Paulo, SP, Brazil
E-mail: marco.mesquita@poli.usp.br

Abstract
This paper provides a simulation model to study the effect of the work-in-process control on due date performance in job shop environment. The due date performance is measured by both the number of tardy jobs and the total tardiness. The simulation runs include different shop configurations (flow shop and general job shop), workloads and sequencing rules. As expected, the results reveal that due date performance is highly dependent on the work-in-process, particularly after the system reaches saturation. Nevertheless, the model is very useful to show job shop managers the effect of the work-in-process control in the due date meeting performance.

Keywords: scheduling, job shop, CONWIP, dispatching rules, simulation

Introduction
Since the emergence of The Japanese Production System, a massive inventory reduction effort is underway. Manufacturing companies worldwide have workout the competitive priorities of Cost, Quality and Speed.

These priorities are somehow conflicting. To achieve low unit cost, the factory should produce high volume of low mix products, but high volume would increase lead times and inventory cost. Meanwhile, high mix and low volume orders are the current demand pattern in the manufacturing environment.

The Toyota Production System was successful in proving that an automotive company can be profitable producing in small lots. The well known practice of Just-in-Time (JIT) supports production processes with lower inventory. The JIT systems are called pull production, in contrast to the traditional push production. In push systems, job orders are

Volume 5, Number 1, 2008, pp. 75-88

released to shop floor to meet due dates. The Material Requirements Planning (MRP) and its successors operate according to this logic. Pushing orders to shop may increase factory congestion and causes efficiency loss.

In a pull system, orders are released according to the factory workload. New orders are authorized to enter the shop once the total shop workload is below the predetermined maximum level. As a consequence, the maximum work-in-process (WIP) is held constant.

The mechanism to control the WIP in JIT manufacturing systems is the Kanban. Basically, the Kanban cards limit the stock between to subsequent work stations in a repetitive low mix, high volume production line.

Spearman et al. (1990) formulated another pull mechanism that seams to be more adequate to intermittent manufacturing (high mix, low volume production). The mechanism ensures that just after an open order is completely finished, a new order can be released to the shop floor. This method was called constant work-in-process (CONWIP).

In fact, the CONWIP is a hybrid strategy since the job flows in a push fashion inside the plant (usually under the FIFO rule), although new orders are pulled only when the WIP drops the predetermined WIP limit. Framinan et al. (2003) present a complete review of CONWIP production control system. Figure 1 provides a visual comparison of the three basic mechanisms discussed above.

In general, the decisions to be taken when implementing CONWIP control mechanism are to determine: (i) the production quota; (ii) the maximum amount of workload; (iii) the

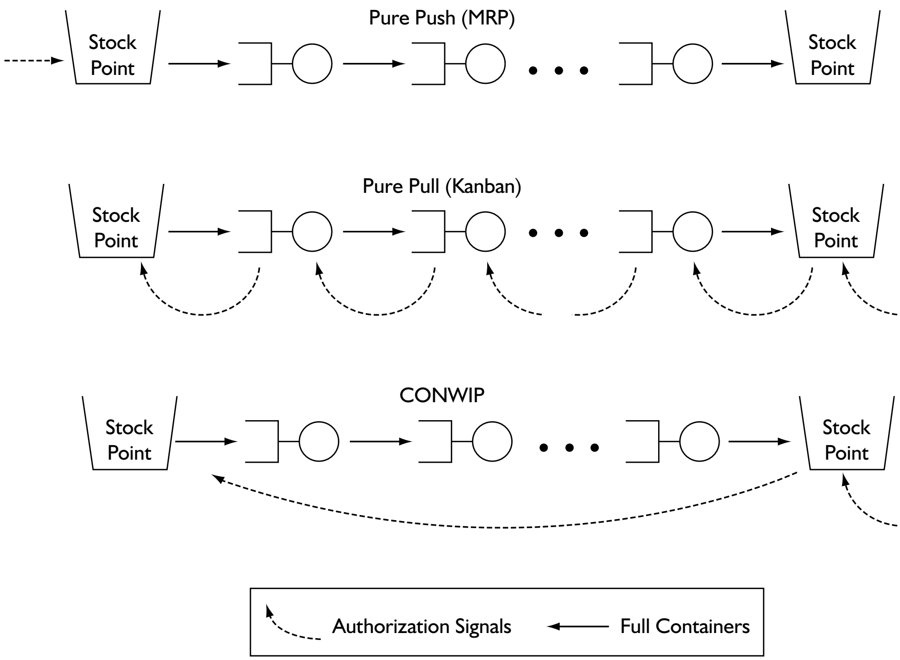

Figure 1 - Push and Pull Production Systems.

capacity shortage trigger; (iv) how to forecast the backlog list; (v) the number of cards into the system; and (vi) how to sequence the orders in the system (Spearman et al., 1990).

This paper provides a simulation model to evaluate the effect of the work-in-process control on due date performance in both flow and general job shop environments. The following section discusses the simulation-based approach for job shop scheduling. Section 3 presents the simulation model used for the evaluation purposes. The scenarios considered are presented in section 4 and the results achieved follow in section 5. The section 6 concludes the paper with an overall analysis and the concluding remarks.

Simulation-Based Scheduling

The classical scheduling theory considers three main groups of goals: (i) on-time delivery; (ii) high throughput; and (iii) maximum machine utilization. For each group, there are different performance measures such as: the maximum tardiness, the number of tardy job, the *makespan* (total completion time), the mean flow time etc. These goals are conflicting and most of the results from scheduling theory "optimize" just one given performance measure (Conway, 1965; Morton and Pentico, 1993; Baker, 1995; Pinedo, 2005).

A scheduling problem can be characterized by a set of jobs, each of them with one or more operations which must be performed in a fixed sequence on different machines. The purpose of scheduling is to determine the schedule that optimizes some performance measure.

The problems considered by the traditional scheduling theory are roughly classified into four main classes: (i) single machine; (ii) parallel machines; (iii) flow shop; and (iv) general job shop. Most of these problems are well known by their combinatorial nature. In particular, the general job shop scheduling problems are included into a large class of intractable numerical problems known as NP-hard (Jain and Meeran, 1999).

Tardiness criterion is of great significance in manufacturing systems since this is one of the most important measures of customer service in a high competitive market. However, very little work is reported on the tardiness problem. A specific review on the flow shop problem is presented by Kim (1995) and an extensive review on scheduling problems with tardiness criterion can be found in Koulamas (1994).

On the other hand, an alternative approach to scheduling problems is the simulation-based scheduling. Simulation has become a widely used tool for operations management. There are many simulation software packages available today that can be used to model and evaluate real-scale system under different performance measures and operational conditions (Law and Kelton, 1991).

The main advantage of using simulation is that one can handle larger problems in reasonable computation times. Complex dispatching rules that integrate management policies and technological constraints can be incorporated into the simulation model. Specially, one can include the work-in-process constraints, which is the main concern of this paper.

Volume 5, Number 1, 2008, pp. 75-88

In the operational level, where the operations scheduling are accomplished, the inventory issue concerns basically to the work-in-process level. Then, two research questions arise:

1. How the work-in-process constraint affects the due date performance?
2. What is the ideal work-in-process level to achieve an efficient operation?

In order to evaluate the CONWIP effect over due date performance, one should, for each WIP constraint and performance measure, solve the corresponding job-shop scheduling problem. As mentioned earlier, the job-shop scheduling problem presents high computational complexity and the optimization algorithms based on mathematical programming (e.g., the branch-and-bound method) would not solve real problems. The alternative approaches would be the search-based meta-heuristics and dispatching rules simulation.

A number of papers have been published over the years dealing with different sequencing rules, using both flow-time and due-date based performance measures (Panwalkar and Iskander, 1977; Vepsalainen and Morton, 1987; Baker, 1995; Chiang and Fu, 2007).

In this paper, we choose the dispatching rules simulation because of the ease of implementation, flexibility, low computational time and satisfactory performance in providing solutions to the job-shop scheduling problem.

The simulation model presented considers four dispatching rules (shortest processing time, earliest due date, least dynamic slack and least work in next queue), two due date related performance measure (total tardiness and number of tardy job) and two shop configurations (flow and job shop).

Simulation Modeling

A simulation model was developed to study the effect of the WIP constraint on due date performance in a job shop environment. The model performs the scheduling of N jobs through a shop of M machines, based on some available sequencing rules. Each job comprises a set of operations to be executed sequentially, each operation in one machine, with a predetermined setup and process times. These jobs are grouped according to their routing into R possible routes, each route corresponds to a particular product and it is characterized by the same sequence of operations with specified setups and processes times.

The shop configuration is determined by the number of machines and the flow pattern. The user should specify how many routes (R) to consider and, for each route, the respective sequence of machine. This may be done by manually or automatically inputting the sequences, using a random route generator. In this case, the user inputs a transition matrix like the one showed in Table 1. In the transition matrix, each entry p_{ij} corresponds to the probability of an order leaving station *i* proceeds to station *j*. Moreover, station 0 is the entrance and the N+1 is the exit corresponding station.

Table 1 - Job Shop Transition Matrix.

	1	2	3	4	5	6	7	8	9
0	50	30	10	10					
1		50	30	10	10				
2	5		50	30	10	5			
3	5	10		40	20	10	10	5	
4		5	10		35	20	20	10	
5			5	10		40	25	15	5
6			5	5	10		40	30	10
7				5	5	10		40	40
8				5	5	5	5		80

In the general job shop problem, orders can move from one station to any other. The flow shop is a particular case of the job shop where there is an implicit machine sequence such that orders can only proceeds forward, that is, p_{ij} will be zero for all pairs (i, j) with $i > j$.

In addition to the routes, the user should specify the due date and the total process time (including setup) for each order. Again, to simplify the data input, the process time of each operation can be randomly generated using some usual probability distribution.

After generating each operation time, a due date is assigned to each order by sum up the process time of all operations and multiplying this total to a factor k greater than one. This value plus a random deviation will determine the specific due date. By varying the k factor, one can achieve different workloads. Higher values of k will produce orders with grater slacks, which mean that orders can wait more time in queue without being late. As k gets close to one, the total queuing time should be reduced in order to complete order on time, and the scheduling problem become much harder.

Finally, we assume that all jobs are available for scheduling at time zero, which it is a common assumption in job-shop scheduling research. Since all time parameters are known in advanced to the simulation, the problem just formulated is classified as a static deterministic job shop scheduling problem that is hard to solve optimally even for a low number of machines and jobs (Pinedo, 2005).

In this paper, it is applied the dispatching rule scheduling approach with CONWIP, that is, a maximum number of orders allowed in the shop is fixed. The orders on the backlog list are released to shop according to a selected sequencing rule.

Inside the shop, each machine has also one sequencing rule chosen from a set of sequencing rules available. The simulation model consists of a discrete event continuous time model and works as follow.

The main event is the completion of an operation in one machine. If this operation is the last one in the process routing, the order is considered finished, otherwise it proceeds to the next station. If this station is idle, operation starts immediately, or, if not, the order joins the queue.

The end of an operation in one machine turns it available to another job. If there is any in queue, the next job will be chosen according to the machine sequencing rule, otherwise, the machine becomes idle.

If the WIP fall down the limit when an order leave the system (last operation finished), a new order is pick up from the backlog list also obeying a dispatching rule assigned. The simulation process proceeds until all jobs are processed. The model was implemented using Visual Basic for Application and Microsoft Excel™.

Experimental Scenarios

In the simulation were considered two shop configurations (flow and general job shop), with high and medium workloads and four sequencing rules. For both shop configurations, it was considered 8 machines and 10 routes with at most 10 operations each route.

Two transition matrices are used for the random generation of the routes, one for each configuration, as shown in Table 1 and Table 2. In the flow shop transition matrix (Table 2), the lower diagonal cells are null since the jobs can not return to any previous machine.

The job shop and the flow shop routes parameters are presented in Table 3. The operations times were generated from a normal distribution with mean 4 and standard deviation 0.4 for all operations. The columns labeled "time" are equal to the sum of all operations times, that is, the total processing time (not included the queuing time) of the orders in that route.

The amount of 30 orders was generated for each of the 10 routes, all of them with ready times equal to zero. The due dates were also randomly generated from a normal distribution with mean $k \cdot t_0$ and standard deviation $0.1 \cdot k \cdot t_0$, where t_0 is the total route time from Table 3 and k is the factor that determine the scenario workload. In the job shop configuration, k assumes the values 1.25 and 1.30 for high and medium workloads, respectively, and in the flow shop configuration the corresponding values are 1.30 and 1.35.

The sequencing rules considered were:

1. SPT - Shortest Process Time,
2. EDD - Earliest Due Date,
3. SLA - Dynamic Slack,
4. LWQ - Least Work in Next Queue.

Table 2 - Flow Shop Transition Matrix.

	1	2	3	4	5	6	7	8	9
0	70	20	10						
1		70	20	10					
2			70	20	10				
3				60	30	10			
4					60	30	10		
5						60	30	10	
6							70	20	10
7								70	30
8									100

Table 3 - Job Shop and Flow Shop Routes.

Job Shop Routes			Flow Shop Routes		
#	**Route**	**Time**	**#**	**Route**	**Time**
1	1, 2, 3, 6	16.9	1	1, 3, 5, 6	15.5
2	1, 3, 7, 8	16.4	2	2, 3, 4, 5, 6, 8	25.6
3	1, 5, 8	11.6	3	1, 3, 4, 7, 8	20.3
4	1, 2, 3, 5, 6, 7, 8	27.3	4	3, 4, 5, 7, 8	20.9
5	3, 4, 6, 7	16.3	5	1, 2, 3, 5, 6, 7, 8	26.5
6	1, 2, 4, 3, 1, 2, 4, 5, 7	36.0	6	1, 2, 3, 5, 7	19.8
7	2, 6, 5, 6, 8	20.7	7	1, 2, 3, 4, 5, 7	24.5
8	1, 2, 4, 5, 7	19.7	8	1, 3, 5, 6, 8	19.8
9	2, 3, 4, 6, 7	21.0	9	2, 4, 6, 7, 8	19.1
10	1, 2, 4, 5, 8	20.6	10	3, 4, 5, 6, 7	20.1

The SPT and EDD are the most usual dispatching rules considered in job shop scheduling simulation. The former is usually associated with higher throughput, since it speeds up the smaller operations and reduces queuing. The second seeks to reduce tardiness by prioritizing the most urgent jobs.

In the third rule, the job with the minimum slack time has higher priority. Slack time is obtained by subtracting the current time and the total processing time of the remaining operations from the due date. Finally, the fourth rule will select the order that has a subsequent operation on the machine with the current minimum work in queue. This rule tends to minimize the chance of machine idleness and to achieve a more continuous flow.

In total, sixteen instances of the problem were considered, each of that corresponding to one configuration, one charge and one sequencing rule. Each instance was initially simulated with no constraint on the work-in-process (WIP) and the highest WIP observed becomes the upper limit for it. (Setting the WIP above that upper limit will not change any performance measure.) Then, the WIP was gradually reduced to verify the effect of this constraint on the performance of the system, which was evaluated by the following performance measures:

TTA = total tardiness,

NTO = number of tardy orders.

In the job shop configuration, each instance comprises 300 orders passing through the shop. Each instance is replicated with 10 different levels of work-in-process, raging from 10 to 100 jobs. For the flow shop case, it was tested 14 upper limits, raging from 10 to 140 jobs.

The number of machines and jobs considered are assumed to be representative of the scheduling problems found in small and medium-sized enterprises.

The software was codified in VBA for Excel and the simulation was run in a microcomputer with Intel Pentium 4 3.0 GHz processor and 512 Mb RAM. The largest computational time did not exceed 90 seconds.

Volume 5, Number 1, 2008, pp. 75-88

Results

This section presents the analysis of the results obtained from the simulation model. These results are graphically shown in the next four subsections, considering the two due-date related performance measures (Total Tardiness and Number of Tardy Orders) and the two shop configurations (job shop and flow shop). Each of the following subsections presents a short analysis of the results achieved.

Total Tardiness - Job Shop

The first scenario considered is the Job Shop with the Total Tardiness measure. For this scenario, the High and Medium load are presented respectively in Figures 2 and 3.

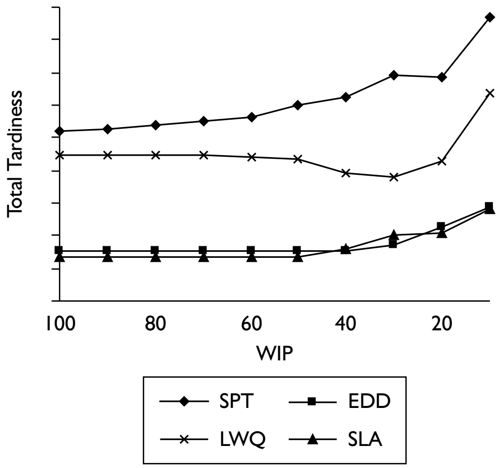

Figure 2 - Total Tardiness for Job Shop – High Load.

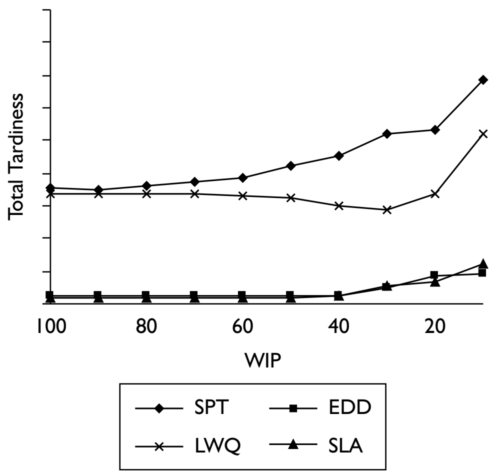

Figure 3 - Total Tardiness for Job Shop – Medium Load.

First of all, one can realize that the workload has changed the values but not the general shape of the plot. By the way, the analysis forward is valid for both loads. From Figures 2 and 3, it is clear that sequencing rules EDD and SLA have an equivalent and better response than the other two rules considered (SPT and LWQ).

Both EDD and SLA present a more regular response, with a uniform increase on Total Tardiness for WIP lower than a turning point of 50 units. For higher values of WIP, there is no change in the total tardiness. Values lower but close to limit level, like 40 units, would promote a less congestion system with almost the same performance on due date criteria. As the WIP decreases to 10 units, the total tardiness increases considerably. The response shape is similar for high and medium workload.

Other results from the simulation reveal that this turning point corresponds to the point beyond that the additional reduction on WIP will cause the increase of the *makespan*, another usual performance measure that correspond to the total time need to complete all operations of all orders in the backlog list. Lower *makespan* is also associated with higher machine utilization.

The rules SPT and LWQ, in comparison to the previous two, always yield higher values of Total Tardiness and a less predictable behavior. The SPT rule had the higher value of the WIP limit (80 units) and below this point, the tardiness performance deteriorates. This fact indicates that SPT rule works better under high congestion, although it is outperformed by the due date related rules (EDD and SLA) in both medium and high workload.

The LWQ rule presents a peculiar pattern if compared to the others. The response curve seems to be not strictly increasing function. It shows a local minimum achieved for a WIP around 30 units. If one continues reducing WIP under this level, the Total Tardiness will increase. It is worth to mention that this pattern was not observed for the *makespan* for the same scenario.

One possible reason to the existence of a local minimum in the total tardiness curve is that, as the system become more relaxed, the rule succeeds in speeding up the flow by choosing stations with shorter queue. This effect stops when the global load continues decreasing. Again, both LWQ and SPT were much worse than due date related rules, named EDD and SLA, in rather medium and high load.

Total Tardiness - Flow Shop

A similar analysis was done for the flow shop configuration. Figures 4 e 5 present the corresponding flow shop results, respectively for the High Load and Medium Load instances.

Like the job shop configuration, the graphics on total tardiness showed the same pattern in the medium and high load case, just changing the values, as expected. Once again, it is possible to group pairwise SPT / LWQ rules and EDD / SLA. The formers were clearly out performed by the others.

In the flow shop, that comprises less subsequent alternative operations, the LWQ lost their advantage in relation to the SPT rule. In addition, the local minimum was not

Volume 5, Number 1, 2008, pp. 75-88

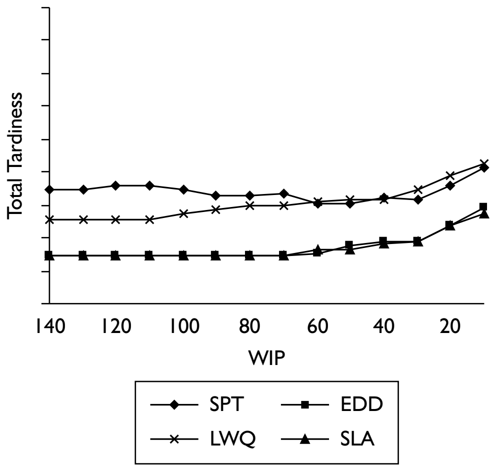

Figure 4 - Total Tardiness for Flow Shop – High Load.

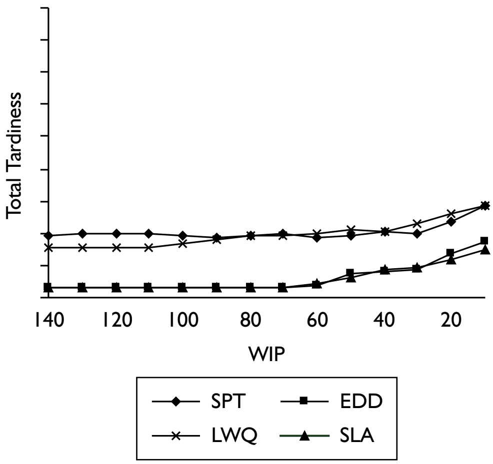

Figure 5 - Total Tardiness for Flow Shop – Medium Load.

detected. In fact, for lower values of WIP, the results for LWQ and SPT become similar. The same happened in the pair EDD / SLA.

The results in Figures 4 and 5 are less regular when compared with those for the job shop case. In spite of the representative sample of jobs simulated, the results do not clearly indicate the existence of an optimum level of WIP in Flow Shop Scheduling with SPT rule and Total Tardiness goal. Additional tests should be conducted in order to get better reasoning on it.

Number of Tardy Orders - Job Shop

What follows is the equivalent analysis replacing the performance measure of Total tardiness by the Number of Tardy Orders. Figures 6 and 7 show the Number of Tardy Orders for the Job Shop, under respectively High and Medium Load.

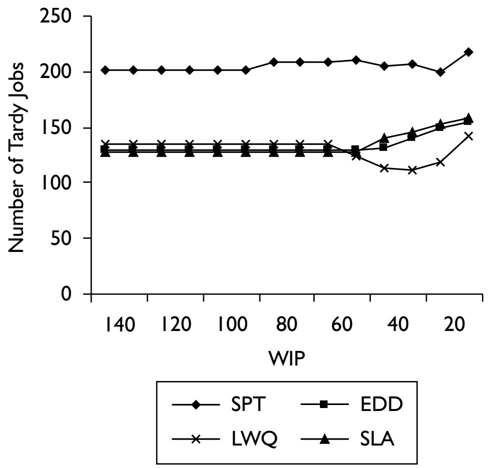

Figure 6 - Number of Tardy Jobs for Job Shop – High Load.

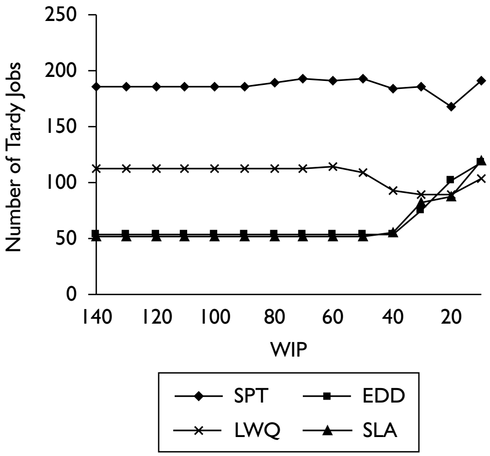

Figure 7 - Number of Tardy Jobs for Job Shop – Medium Load.

Once again, the High and Medium Load are different only in the absolute values, being the response curve of a similar shape for each rule.

The EDD and SLA present results very similar to those for Total Tardiness, that is, the increase in the Number of Tardy Job for WIP under the 50 units limit level. The SPT rule display results with little variation in the WIP range considered. The best performance is reached in the WIP level of 20 units considered.

Considering the LWQ rule, it results in lower Number of Tardy Orders than EDD and SLA for WIP levels below 50 units. In the Medium Load condition, this just happened for values below 20 units. This would be a promising result but the high levels of Total Tardiness achieved before is not. This apparent contradiction suggests that the rule yield a few number of orders late but those late orders with a higher tardiness, that is, a greater variance in orders lateness.

Volume 5, Number 1, 2008, pp. 75-88

Additional results from simulation indicates that the mean tardiness grow up with reduction on WIP levels for all scheduling rules and that due date relate rules outperformed the other rules considered in that performance measure.

Number of Tardy Jobs - Flow Shop

Concluding the scenarios studied, Figures 8 and 9 present the Number of Tardy Jobs in the Flow Shop configuration.

The rules EDD and SLA provided better and closed results if compared to the Job Shop case. There is an increase in the Number of Tardy Jobs for values of WIP below 30 units and around 50. The LWQ rule, on the other hand, revealed an increase in the Number of Tardy

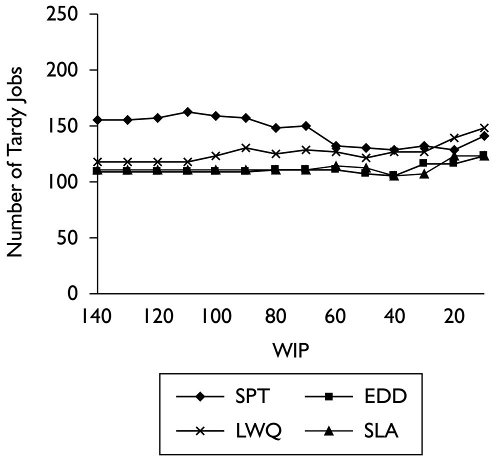

Figure 8 - Number of Tardy Jobs for Flow Shop – High Load.

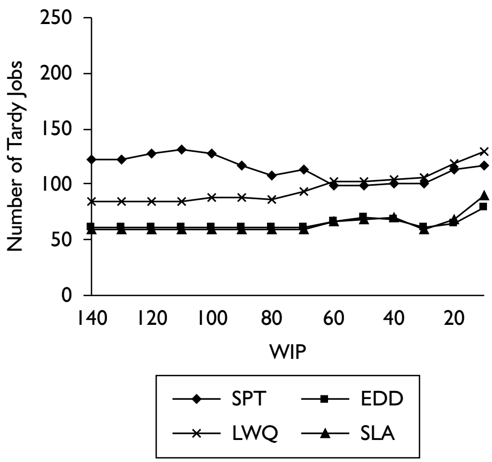

Figure 9 - Number of Tardy Jobs for Flow Shop – Medium Load.

Orders before, when WIP reached 70 units. This probably happened because higher levels of WIP will cause longer queuing time and greater Number of Tardy Jobs. From the above results, it is not possible to make sure inferences about the optimum value of WIP for the flow shop configuration to minimize the Number of Tardy Orders.

Conclusions

Firstly, the Total Tardiness may be considered more important than the Number of Tardy Jobs. In the current manufacturing scene, for example, there is no effect to reduce the Number of Tardy Orders in 50% meanwhile increasing the total tardiness (and consequently the mean tardiness) three or four times.

The analysis considering the Total Tardiness proved that the rules EDD and SLA are consistently better and more regular than the others. Furthermore, in a Job Shop configuration, the results reveal the existence of a WIP limit level that should be evaluated and used for production control purposes. This limit is considerable lower than the limit level obtained with other rules. This suggests that the performance on Total tardiness can be considerably improved in a Job Shop with the use of CONWIP and the sequencing rules like EDD and SLA.

In fact, the level of WIP affect other productivity measures not explicit consider herein, the WIP optimal level could be even lower than that limit, depending on the trade off between inventory reduction and the capacity utilization.

Manufacturing companies with high stock out cost and low inventory cost should operate with WIP near that limit, meanwhile those with low stock out cost and high inventory cost, the WIP should be even lower.

The previous analysis fails in the case of SPT and LWQ since the results reveal a strong correlation between performance measures and both, the shop and load configuration. Although it is possible to evaluate the WIP optimum for a specific case under determined configuration and load, the response curve exhibit a very irregular pattern, it make difficult to find a limit with the same property of that encountered in the last case. This probably happens because both the EDD and SLA rules consider due dates to set the sequencing priorities, in opposite of the other two.

In case one considers the Number of Tardy Orders more important than Total Tardiness, the results suggests a careful analysis of the LWQ rule in the Job Shop with High Load configuration, since it had a superior performance and a clear optimum (minimum level).

References

Baker, K.R. (1995), Elements of Sequencing and Scheduling (Amos Tuck School, Hanover).

Chiang, T.C. and Fu, L.C. (2007), Using dispatching rule for job shop scheduling with due date-based objectives. International Journal of Production Research, Vol. 45, No. 14, pp. 3245-3262.

Conway, R.W. (1965), "Priority Dispatching and Work-In-Process Inventory in a Job Shop", Journal of Industrial Engineering, Vol. 16, No. 2, pp. 123-130.

Framinan, J.M. & González P.L. & Ruiz-Usano R. (2003), "The CONWIP Production Control System: review and research issues", Production Planning and Control, Vol. 14, No. 3, pp. 255-265.

Jain, A.S. and Meeran, S. (1999), "Deterministic Job Shop Scheduling: Past, Present and Future", European Journal of Operational Research, Vol. 113, No. 2, pp. 390-434.

Kim, Y.D. (1995), "Minimizing Total Tardiness in Permutation Flowshops", European Journal of Operations Research , Vol. 85, No. 3, pp. 541-555.

Koulamas, C. (1994), "The Total Tardiness Problem: Review and Extension", Operations Research, Vol. 42, No. 6, pp. 1025-1041.

Law, A.M. and Kelton, W.D. (1991), Simulation Modeling and Analysis (McGraw Hill, New York).

Morton T.E. and Pentico D.W. (1993), Heuristic Scheduling Systems (Wiley, New York).

Panwalkar, S.S. and Iskander W. (1977), "A Survey of Scheduling Rules", Operations Research, Vol. 25, No. 1, pp. 45-61.

Pinedo, M. (2005), Planning and Scheduling in Manufacturing and Services (Prentice Hall, Englewood Cliffs).

Spearman, M.L. & Woodruff, D.L. & Hopp, W.J. (1990), "Conwip: a pull alternative to kanban", International Journal of Production Research, Vol. 28, No. 5, pp. 879-894.

Vepsalainen, A.P.J. and Morton, T. E. (1987), "Priority Rules for Job Shops with Weighted Tardiness Cost", Management Science, Vol. 33, No. 8, pp. 1035-1047.

Biography

Miguel Cezar Santoro is Associate Professor of Operations and Logistics Management at the Polytechnic School, University of São Paulo (USP). He earned his PhD in Industrial Engineering from the University of São Paulo. He teaches undergraduate and graduate courses in Operations and Logistics Management. His research areas of interest include production scheduling, inventory and distribution.

Marco Aurélio de Mesquita is Assistant Professor of Operations and Logistics Management at the Polytechnic School, University of São Paulo (USP). He holds a BS and MS degree in Naval Engineering and a PhD degree in Production Engineering from the Polytechnic School (USP). His research interests include modeling and simulation of production systems, inventory management and production scheduling. He is a member of the ABEPRO (The Brazilian Association for Production Engineering) and of the INFORMS.

Submitted: 09th October, 2008.
Accepted: 19th December, 2008.