Workload Control under Continuous Order Release

N. O. Fernandes¹, S. Carmo-Silva²

¹ Dept de Engenharia Industrial, Instituto Politécnico de Castelo Branco, 6000 Castelo Branco, Portugal, <u>nogf@est.ipcb.pt</u>

² Dept de Produção e Sistemas, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal, <u>scarmo@dps.uminho.pt</u>

Abstract

Past research on workload control (WLC) has been essentially focused on discrete order release. This means that release of orders to the shop floor takes places on a periodic basis. Continuous order release has been somehow neglected, in spite of its apparent potential for improving system performance, including the reduction of order flow times. This paper addresses a simulation study of this order release approach. The study contributes for improving the basis for setting workload norms and selecting the workload control strategy under continuous order release. Additionally, it gives insights on the choice of routing alternatives, which facilitate the linkage of continuous order release mechanisms with the specific characteristics of the shop floor.

Keywords: Workload Control, Order Release, Simulation.

¹Corresponding author: N. O. Fernandes

Escola Superior de Tecnologia, Instituto Politécnico de Castelo branco, Av. do Empresário, 6000, Castelo Branco, Portugal *Phone*: +351272339300, Fax: +351272339309, *E-mail*: nogf@est.ipcb.pt

1. Introduction

For manufacturing enterprises to stay competitive in the global market of today, manufacturing strategies have to be focused on speed of response to customer requirements, which means short delivery times, on time deliveries and flexibility to customer requirements. As a result there has been a large growth in the number of enterprises that operate in the MTO sector (Stevenson, 2005).

Workload control (WLC) is a production planning and control (PPC) concept designed specifically for complex manufacturing environments, with particular relevance to the make-to-order (MTO) sector and small and medium sized enterprises (SMEs) (Stevenson, 2009). Its main principle is to keep the length of queues on the shop floor at appropriate levels to meet promised deliver dates, taking into account the system capacity and capabilities. If these queues are kept short and stable, then waiting times and therefore flow times, will be controlled and reliable (Kingsman, 2000). It is possible to identify three hierarchical levels, related with stages in the order flow, at which the control of these queues can be attempted, namely order entry, order release and dispatching (Breithaupt et al., 2002). At each level, a decision must be made relatively to the orders allowed to proceed to the next stage and whether this requires capacity adjustments.

Order release is described as an essential decision function and a core parte of WLC (Missbauer, 2009). It occurs when orders (jobs) are released into the shop floor for processing. The concept behind controlled release is to release orders selectively, at the right moments in time, to improve shop performance. An order release mechanism is used, in combination with a pre-shop pool, to determine the moment and the orders to release into the shop floor. Releasing a set of orders is only feasible if it does not violate workload norms. Orders generated by the planning system or arriving directly from customers over time, are gathered in the pre-

shop pool and are only released if they fit workload norms, usually defined in hours of work, of the required capacity groups (e.g. work centres). This means that the decision to release an order is based on its influence to the current shop floor situation.

Within WLC, releases may take place at periodic, i.e. discrete time intervals (e.g. at the beginning of each working shift, day or week), or on a continuous basis, i.e. at any time during the system's operation.

Continuous order release is based on the continuous monitoring of workload in the shop floor, in order to determine whenever a workload falls below its norm. At this moment the feasibility of order release is checked for the orders in the pool. This has implications for the effort required to manage the pre-shop pool and the order release activities, but allows for a more up to date control of the shop floor and stabilises workload on capacity groups.

Discrete order release, on the other hand, is based on periodic observations of workload in the shop floor. At fixed periods of time, workload on the shop floor is computed, and the decision to release orders is taken. When orders are released periodically, release mechanisms have to set a release period. This must be less than the smallest slack of the orders in the pool, in order to avoid lateness. Past research (e.g. Land, 2006) has shown, however, that the choice of an appropriate period between releases is a delicate decision, greatly influencing shop performance and that, non-periodic release methods must be emphasised within future research. Land concluded that a long release period delays orders in the pool and increases the time they spend in the entire system. A short release period, on the other hand, hinders the release of large jobs, in terms of processing time and routing sequence, and thus may impede the right timing of release of these jobs. The reduced costs of feedback information and the grater simplicity of discrete order release are the most suitable explanation why it is suggested by practitioners and most of the researchers (Bergamaschi et al 1997).

3

In fact, most of the past research on the WLC concept has been focused on discrete order release mechanisms (e.g. Oosterman et al., 2000, Kingsman and Hendry, 2002, Cigolini and Portioli, 2002, Land, 2006 and Henrich et al., 2007). Exceptions, include Land and Gaalman (1998), Sabuncuoglu and Karapinar (1999), the Generic POLCA mechanism (Fernandes and Carmo-Silva, 2006) and (Ebadian et al., 2009). Recently, Stevenson and Silva (2007) reported refinements made to a particular WLC methodology, the LUMS approach, in two independent empirical research projects undertaken in Portugal and in the U.K. In both projects job release typically takes palace daily (or even one a shift) rather than weekly (as proposed in the original methodology).

Continuous order release reflects the increasing shorter lead time expectations of customers. It is likely to improve flow times (since continuously reviewing jobs for release minimizes their time in the pool) and seems to be in line with the current competitive manufacturing environment where enterprises have to be focused on speed of response to customer requirements. These market pressures are common to a large number of companies and thus can be considered, as noted by Stevenson and Silva (2007), to be a relatively generic change that is likely to be required in future implementations of the WLC concept.

Based on these expectations, a simulation study of the performance of three WLC strategies under continuous order release, in a job-shop environment, was carried out and here presented. The study aims at improving the basis for setting workload norms, selecting the workload control strategy and deciding upon routing alternatives, which facilitate the linkage of continuous order release procedures with the specific characteristics of the shop floor, especially machine characteristics. In particular, answer to the following questions is sought:

- How workload control strategies perform under continuous order release?
- How machine interchangeability affects the routing decisions under continuous order release?

Note that machine interchangeability is possible when to two or more machines have identical processing capabilities.

The remainder of this paper is organized as follows. Section 2 addresses the overall research methodology, detailing the simulation model and the experimental setup. Section 3 presents, analyses and discusses the results of the simulation study. Finally, in section 4 concluding remarks are made and directions for future research work are presented.

2. Research Methodology

A simulation study was carried out using the Arena[®] software (Kelton et al. 2004). During simulation runs, data were collected under system steady-state. The length of each simulation run is 30,000 time units, in which the first 9,600 time units are considered as the warm-up period. The average values of 90 replications are presented as results. The statistical analysis was performed using the paired Student *t*-test with a 95% confidence level.

2.1 Simulation Model

A job shop without an explicit bottleneck has been considered. The structure of the shop consists of six work centres, each with a single multi-purpose machine, except for work centre 1. This work centre is modelled as having two machines with a given degree interchangeability (i.e. processing overlapping capabilities related to the technical ability of the machines to perform similar operations).

Order entry

Jobs (orders) arrive from the customer at the order entry stage, where the process plan is defined. This includes the routing, the operation processing times and the promised due date. Arriving jobs are accumulated in a pre-shop pool ahead of the release stage.

The job's route in this study is determined by 20 discrete routeing patterns, each having an equal probability of occurrence. Table 1 shows the required work centres for each routing pattern. The average number of operations per routing pattern is 3.6.

[Insert Table 1]

The mean job inter-arrival time is 0.666 time units and follows an exponential distribution. Processing times at each machine follow a 2-Erlang distribution with a mean of 1 time unit, except for the machines of work centre 1. For these machines, the mean processing time is 2 time units. This result in a 90% average planned utilization in all machines. Due dates are modelled as a random variable and are determined by the order arrival time plus a uniformly distributed time allowance. The minimum allowance value equals the planned shop flow time for the maximum of 6 operations. The maximum allowance value was chosen to ensure that, under immediate release, the number of tardy jobs is about 5%.

Order release

In this stage, a release mechanism is used to determine the moment and the jobs that are actually released into the shop floor. When a job arrives at the pre-shop pool or when a job completes an operation, workload is updated and the pool is inspected in order to select a new job for release (see Figure 1). Only those jobs for which the planned release time falls within a specific time limit are candidates for release. In this simulation study the *time limit* was set to "infinite". This should improve the possibilities for load balancing within the release procedure (Land, 2006).

The release mechanism used comprises both, timing and balancing functions. This means that, on one hand, jobs waiting in the pool are considered for release according to the earliest planned release time, which determines their relative urgency (*timing function*). Planned release times are determined by backward scheduling from the promised due date, using a

constant lead time per operation, which was established through some pilot simulation runs under immediate release. On the other hand, jobs are only released if they fit workload norms for required capacity groups (*balancing function*). Capacity groups in our study refer to the smallest production units to be controlled during release.

Workload accounting for capacity groups is based on the *adjusted aggregated load* method (Oosterman et al., 2000). The underlying idea of this method is that the accounted load should be corrected for the variable position of the capacity group in the routings of the released jobs. This means that just a fraction (p_{ij}/n_{ij}) of the theoretical operation processing time (p_{ij}) of a job *j* on a machine from capacity group *i*, is accounted in case of downstream operations, with n_{ij} being the position of *i* in the job's routing. In this case each job being released will increase workload with p_{ij}/n_{ij} for each capacity group in its routing and after the respective operation at a capacity group is completed, workload will decrease accordingly.

[Insert Figure 1]

Dispatching

Once a job is released, its progress through the shop is controlled by priority dispatching rules. It is assumed that jobs follow a first-in-first-out (FIFO) dispatching rule in all machines. Setup times have been considered sequence independent and assumed as part of the operation processing time. Transportation times between work centres were neglected.

2.2 Experimental Design

The following experimental factors were evaluated: the workload norms level; the workload control strategy; the routing decision; and the degree of machine interchangeability. Table 2 summarises the experimental factors and the associated levels within the simulation study.

[Insert Table 2]

Workload control was studied at three levels or strategies, namely *upper workload bound*, *lower workload bound* and *workload balancing*.

- The first strategy, *upper workload bound*, is a typical approach to workload control allowing the release of a job into the shop floor, only if workload in all capacity groups of the job's routing does not exceeds an upper limit or workload norm.
- The second strategy, *lower workload bound*, seeks to avoid 'starving' of work centres by ensuring that workload in all capacity groups is above a lower limit or workload norm. This means that a job will be released if, in one or more capacity groups of the job's routing, workload is lower than the lower limit. However, only those jobs within the *pool* whose first operation is at one of the under-loaded capacity groups are released.
- The third strategy, *workload balancing*, releases jobs only if they contribute for a better load balancing among capacity groups. The balancing index considered is given by:

$$BI = \sum_{j} \sum_{i} \left| F_{ij} - rw_i \right| \tag{1}$$

 F_{ij} represents the accounted workload on capacity group *i* resulting from releasing job *j* into the shop floor, and rw_i is the reference workload level set for *i*. The best balancing situation is obtained by minimizing *BI*. This strategy assumes that relaxing workload norms, by using workload balancing, do not necessarily leads to a higher average workload on the shop floor (Land and Gaalman 1998). The strategy tends to compensate for the fact that, when workload across capacity groups is poorly balanced, rigid upper workload limits for the heavily loaded capacity groups may block the release of work to the under-loaded capacity groups.

The routing decision refers to the strategy for choosing one of the machines of work centre 1 for the job routing. This was studied at two levels, namely, at *release* and at *dispatching*.

• Making the routing decision at release may support the balancing function based on a detailed balancing of workloads across machines, by fitting jobs from the pool into

workload norms. This requires defining independent capacity groups, one per each of the machines of the work centre 1. Balanced loads are expected to improve the accurateness of flow times, which, in turn, enable establishing more accurate lead times and therefore release times. The routing decision at release is based on a *smallest load* rule.

• Postponing the routing decision to the shop floor, i.e. to dispatching, may have a favourable effect on waiting times, by grouping the machines of work centre 1 in single capacity group and collecting jobs in a common queue for both machines. This prevents jobs from waiting on a machine while the other is idle. In fact, reducing the sources of variability, due to machines sharing of a common buffer, reduces the total amount of buffering required to achieve a given level of performance. This is known as a form of pooling that involves sharing inventory buffers to cover variability in multiple sources of demand. Jobs are selected to be processed on machines accordingly a FIFO (*frist-in-frist-out*) rule.

Because a company often uses machines over time that may vary in age, specification, processing speeds and set-up times, it is unlikely that they are (completely) interchangeable. As a result, it can be necessary to group semi-interchangeable machines. To evaluate the influence of this on shop performance the degree of machine interchangeability was studied at three levels, namely: 1, 0.2 and 0. These values are related with the percentage of jobs that can be performed in either of the two machines of the work centre 1, respectively 100% (interchangeable machines), 20% (semi-interchangeable) and 0% (no interchangeable).

Workload norms are deterministic parameters setting the *maximum*, the *minimum* or the *reference* workload level, accordingly to the workload control strategy used, on each capacity group. To determine the best performing workload norms levels it is common practice in simulation studies (e.g. Thuerer et al., 2009, Henrich, 2007, Land, 2006, Oosterman et al., 2000) to define it as an experimental variable. This variable is varied stepwise down from

infinity, which means immediate release under the continuous timing convention. Since machines show identical characteristics, i.e. utilization, operation processing times, stream of arriving orders and average flow times, norm values were set identical for all capacity groups. Only when work centre 1 is spitted in two capacity groups, different workload norm values have to be set. In this study they are related by a fixed percentage to the norm values of the other capacity groups. This percentage was determined through some pilot simulation runs, under immediate release.

3. **Results and Discussion**

This section presents and discusses the results of the simulation study described in the preceding section. The performance measures recorded are the time in system, the shop flow time, the percentage of tardy jobs and the standard deviation of job lateness. Time in system is used as an indicator of balancing performance of the order release mechanism and refers to the time a job spends waiting in the pre-shop pool plus the shop flow time. The benefits of reducing time in system are related with reducing the overall response time to customers. The shop flow time refers to the time that elapses between job release and job completion. Reducing the shop flow time has also intrinsic benefits, which implies a smaller WIP (*work in process*) and therefore reduced capital tied up. The percentage of tardy jobs refers to the fraction of jobs that is completed after the promised due date. The standard deviation of job lateness is a measure of how spread out a lateness distribution is. It is used as an indicator of timing performance, i.e. it indicates how close the jobs are completed near their due dates.

Selecting the Workload Control Strategy

Figure 2 shows time in system behaviour for each one of the workload control strategies. These are the results of keeping the value of each factor in the reference level, underlined in Table 2, when the workload control factor is varied. Only workload norms have been fully varied in combination with the workload control strategy.

In this figure, the weighted job-average value of the time in system is plotted against the weighted job-average value of the shop flow time. Superior strategies yield lower time in system for a given shop flow time, i.e. will have a curve which is shifted down and to the left. A point on the curve is the result of simulating a workload control strategy at a specific workload norm level. Note that norms are equally tight if they result in the same shop flow time.

As can be seen, curves converge at a shop flow time of 32.2 time units, which is also the value of the time in system. This is the result of unrestricted workload norm levels, meaning that jobs do not wait in the pre-shop pool of orders, i.e. release is immediate. As could be expected, in these circumstances, all workload control strategies give the same results. Tighter norms lead to lower values of shop flow times and, up to a point, also to time in system decreases, Figure 2. The smallest value of time in system, 28.3 time units, is achieved for a shop flow times of 21.8 time units, under the balancing strategy. This represents about 12% reduction in time in system and about 32% reduction in the shop flow time, time in system tends to increase substantially. This means that waiting time in the shop floor is partially replaced by waiting time in the pre-shop pool of orders. Thus, since the time in system is the sum of the pool time and the shop floor time, we may conclude that waiting times in the pool increase more than waiting times on the shop floor decrease. This means that to avoid deterioration of time in system, norms cannot be set excessively tight.

[Insert Figure 2]

Table 3 resumes the obtained results for the performance measures considered. In this table, workload control strategies are compared at a norm level that results in about 32% reduction

in the shop flow time, as indicated in Figure 2. Table 4 lists the percentage improvements of controlled release strategies over immediate release. The paired Student t-test is used to perform the statistical analysis with a 95% confidence level.

[Insert Table 3]

[Insert Table 4]

The appropriate choice of workload norms under controlled release can significantly improve the system performance in terms of the average time in system, average shop flow time and percentage of tardy jobs, relatively to immediate release. This is mainly because controlled release is able to adjust the release decision, responding to stochastic events such as processing time variability. It is also expected that the improvement introduced by the controlled release process will becomes more significant with the increase of the system utilization. However, controlled release deteriorates system performance in terms of the standard deviation of the lateness (StDev lateness). This essentially results from the introduction of variability by the release process, due to the use of workload norms that may disturb the planned release sequence by holding back the release of some urgent jobs.

Among the evaluated controlled release strategies, workload balancing is the best performer in terms of time in system and percentage of tardy jobs. The upper bound strategy performs best in terms of the standard deviation of lateness. This may be attributed to the fact that less variability is introduced in the release process by upper bound strategy. The lower bound strategy does not seem to provide an effective form of control. It shows the worst results for all the performance measures studied.

Deciding on Routing Alternatives

Figures 3 to 5, show time in system behaviour for the routing decision at release and dispatching, for different degrees of machine interchangeability, under the upper bound control strategy.

[Insert Figure 3] [Insert Figure 4] [Insert Figure 5]

Figure 3 refers to a situation of (fully) interchangeable machines. As can be seen, the dispatching curve stays below the release curve for each level of norm tightness, i.e. curves do not cross each other. This shows that, with interchangeable machines, making the routing decision at shop floor, i.e. at the dispatching level, is preferable.

Figure 4 refers to the situation of no interchangeable machines. Curves do not cross each other, as in the situation of interchangeable machines. However, in this case, the release curve stays below the dispatching curve for each level of norm tightness. This shows that making the routing decision at order release is preferable.

Figure 5 refers to a 20% interchangeable situation, which means that 20% of the jobs can be carried out on both machines. As can be seen, the two performance curves cross each other, meaning that the level of norm tightness influences the routing decision.

We can conclude, from results, that the routing decision is influenced by the degree of machine interchangeability under continuous order release. This is consistent with the research of Henrich et al. (2007) in a similar manufacturing environment, for discrete order release. This suggests that the routing decision is not influenced by the timing convention, continuous or discrete.

By studying Figures 3, 4 and 5 it is possible to conclude that: (1) a higher degree of interchangeability leads to a considerable improvement on the system performance in terms of

time in system; and (2) routing at dispatching seems to be less robust to the degree of interchangeability than routing at release.

[Insert Figure 6]

Figures 6 shows time in system behaviour for the routing decision at release and dispatching, for the 20% interchangeable situation, under the workload balancing strategy. For less restrictive workload norms performance curves become closer to each other. This means that an increase in the balancing capabilities offer by the workload balancing strategy leads to a decrease in the shop load, and therefore the pooling synergies partially looses its effect. Pooling synergies results from collecting jobs in a common queue for both machines of work centre 1 as discussed in section 2.2.

In general, it can be concluded that routing at dispatching partially looses its effectiveness with low degrees of interchangeability and with improved balancing capabilities of the order release mechanisms.

4. Conclusions

Continuous order release control has a significant impact on the manufacturing system performance. In this paper, we report a simulation study of several issues concerning WLC under continuous order release. Specifically, we discuss the impact of the workload control strategy and the implications of machine interchangeability on the system performance.

Results show that no single workload control strategy performs best for all performance measures. Workload balancing performs better for the time in system and for the percentage of tardy jobs, while the upper bound performs better for the standard deviation of lateness.

Results also show that machine interchangeability has a major influence on the routing decision. Routing at dispatching performs better for high degrees of interchangeability and for

less restrictive workload norms. However, the effectiveness of this control option may be limited if order release mechanisms have good balancing capabilities.

Considering the attractive results offered by the workload balancing strategy on time in system, future research work should explore ways of reducing the standard deviation of lateness of this strategy. This would make workload balancing a particularly recommended strategy for job shop operation under continuous order release.

5. References

- Breithaupt, J.W., Land, J. Nyhuis, P. 2002. The workload control concept: theory and practical extensions of Load Oriented Order Release. *Production Planning and Control*, 13 (7), 625-638.
- Cigolini, R. Portioli, A. 2002. An experimental investigation on workload limiting methods within ORR policies in a job shop environment. *International Journal of Production Research*, 13(7), 602-613.
- Ebadian, M., Rabbani, M., Torabi, S.A., Jolai, F., 2009. Hierarchical production planning and scheduling in make-to-order environments: reaching short and reliable delivery dates. *International Journal of Production Research*, 47, (20), 5761–5789.
- Fernandes, N.O., Carmo-Silva, S. 2006. Generic POLCA A production and materials flow control mechanism for quick response manufacturing. *International Journal of Production Economics*. 104 (1), 74-84.
- Henrich, P., Land, M., Gaalman, G. 2004. Exploring applicability of the workload control concept. *International Journal of Production Economics*, 90 (2), 187-198.
- Henrich, P., Land, M., Gaalman, G. 2007. Semi-interchangeable machines: implications for workload control. *Production Planning and Control*, 18 (2), 91-104.
- Kelton, W., Sadowski, R., Sturrock, D., 2004. Simulation with arena, 3rd Ed. New York: McGraw-Hill.
- Kingsman, B.G., Hendry, L. 2002. The relative contributions of input and output controls on the performance of a workload control system in make to order companies. *Production Planning and Control.* 13 (7), 579–590.
- Land, M. 2006. Parameters and sensitivity in workload control. *International Journal of Production Economics*, 104 (2), 625-638.

- Land, M., Gaalman, G. 1998. The performance of workload control concepts in job shops: Improving the release method. *International Journal of Production Economics*, 56–57 (20), 347–364.
- Missbauer, H. 2009. Models of the transient behaviour of production units to optimize the aggregate material flow. *International Journal of Production Economics*, 118 (2) 387-397.
- Oosterman, B., Land, M., Gaalman, G. 2000. The influence of shop characteristics on workload control. *International Journal of Production Economics*, 68 (1), 107-119.
- Sabuncuoglu, I., Karapinar, H.Y. 1999. Analysis of order review/release problems in production systems. *International Journal of Production Economics* 62 (3), 259-279.
- Stevenson, M. 2006. Refining a Workload Control (WLC) concept: a case study. *International Journal of Production Research*, 44, (4), 767–790.
- Stevenson, M., Huang, Y., Hendry, L.C. 2009. The development and application of an interactive end-user training tool: part of an implementation strategy for workload control. *Production Planning and Control*, 20 (7), 622-635.
- Thuerer, M., Silva, C., Stevenson, M. 2009. Workload control release mechanisms: from practice back to theory. *International Journal of Production Research*, In Press.

Routing	Operation number					
pattern	#1	#2	#3	#4	#5	#6
1	2	4	6	1	5	3
2	1	3	5			
3	2	3	5	4		
4	5					
5	4	2	5	6	1	
6	2	5	4	6	1	3
7	1	3	2	6		
8	2	6				
9	2	5	4			
10	3	1	5	4	6	2
11	6	2	3			
12	2	6	1	3		
13	2	3	6			
14	4	1	2	5	3	
15	1					
16	4	3	6	5	1	
17	4					
18	3	4	6	5		
19	4	1	6	5		
20	4	1				

Table 1. Work centre job routing matrix.

Factor	Level 1	Level 2	Level 3
Workload control strategy	<u>upper</u>	lower	balancing
Routing decision	<u>dispatching</u>	release	
Degree of machine interchangeability	<u>1</u>	0,2	0
Workload norms	stepwise down from infinite		

Table 2. Experimental factors and levels.

Performance measures	upper	lower	balancing
Shop flow time	21.84 (±0.10)	21.84 (±0.13)	21.83 (±0.10)
Time in system	30.60 (±0.46)	30.65 (±0.52)	28.31 (±0.35)
Percent tardy	9.20 (±0.38)	10.28 (±0.44)	7.70 (±0.30)
StDev lateness	21.36 (±0.55)	25.67 (±0.83)	23.35 (±0.96)

Table 3. Workload control strategies performance results.

Performance measures	upper	lower	balancing
Shop flow time	-32.1%	-32.1%	-32.1%
Time in system	-4.9%	-4.7%	-12.1%
Percent tardy	-9.2%	+1.5% ⊕	-25.3%
StDev lateness	+12.2%	+34.9%	+20.2%

[®] not significant at a 95% confidence level

Table 4. Percentage improvement over immediate release.



Figure 1. The release decision making process.



Figure 2. Performance curves under different workload control strategies.



Figure 3. Performance curves for routing decision with interchangeable machines under upper workload bound control.



Figure 4. Performance curves for routing decision with no interchangeable machines under upper workload bound control.



Figure 5. Performance curves for routing decision with semi-interchangeable machines under upper workload bound control.



Figure 6. Performance curves for routing decision with semi-interchangeable machines under workload balancing.