Workload control in job shops, grasping the tap

Land, Martin

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2004

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Land, M. J. (2004). Workload control in job shops, grasping the tap s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Summary

The term *job shops* is used to indicate companies that produce customer-specific components in small batches. *Jobs* (production orders) in a job shop are characterised by a large variety of *routings* and *operation processing times*. This variety, combined with irregular order arrivals, generally leads to long waiting times on the shop floor.

The *Workload Control (WLC) concept* is specifically designed to create a match between capacity requirements and capacity availability in such an environment. The key decision within this concept is the *release* of jobs to the shop floor. This decision is often illustrated by the metaphor of a tap controlling the flow into a bathtub (the shop floor). This thesis focuses on grasping the tap, with the double meaning of holding it firmly and understanding it. Although WLC can be typified as a simple and practical concept, the exact influences of the release decision are difficult to understand. Nevertheless, a detailed understanding of these influences is essential for the application and development of WLC. Therefore, a twofold objective has been formulated for this research project:

1. To deepen knowledge on the functioning of release methods within the WLC concept;
2. To improve release methods within the WLC concept.

The research project has been subdivided in four phases: a comparison and analysis of release methods, a simulation study to test the methods and to provide an in-depth analysis, the redesign of methods, and finally an evaluation of the redesign.

In the first phase the WLC concept is analysed in detail and a comparison of the classical WLC release methods is carried out. A typical characteristic of the WLC concept is that it collects the accepted jobs in a pre-shop *job pool*. Periodically, a release procedure determines the set of jobs that can be released to the shop floor. After their release the jobs stay on the shop floor. *Priority rules* then determine the sequence in which the jobs are processed at each (work)station. The release procedure consists of two components: Firstly the jobs are sequenced in order of their planned release date and then the jobs are tested against a set of station-specific *workload norms*. A job that fits is selected and included in the workloads. A job that would cause a norm to be exceeded will have to wait until the next release moment. In that situation a job with a later planned release date that obeys the norms might be selected instead.

The release methods fulfil both a *timing function* and a (*load*) *balancing function*. A method is supposed to have good timing qualities if it creates a low variance of lateness among jobs. Considering the jobs in order of their planned release dates should contribute to the timing function of the described release procedure. Selecting
jobs that obey the norms must ensure the load balancing qualities. Load balancing should be interpreted as releasing a balanced mix of jobs to the shop floor. A balanced mix is a mix that guarantees a smooth arrival pattern of work at each station. This in turn speeds up the throughput of jobs, which results in a low average throughput time. Furthermore, it creates low and steady levels of direct load. The direct load of a station is measured as the sum of the processing times (in hours of work for the considered station) of those jobs for which preceding operations have been completed while the operation at the considered station is not yet completed. A steady direct load level contributes to the predictability of the station throughput times. This is important for the reliability of the planned release dates since these are based on planned station throughput times. Thus, the balancing function contributes indirectly to the timing function.

Each job has its own routing that determines in which sequence the stations are visited. Only the direct load of the first station in this routing can be influenced immediately by the release of a job. The direct load of a station that is visited downstream is influenced after a certain delay. As such it does not make sense to relate the workload norms straight to the direct loads. The two classical WLC release methods that are investigated in this thesis differ in their way of dealing with this indirect influence. This has resulted in two types of workload norms: (A) norms for the estimated direct load level at the next release moment, and (B) norms for the aggregate load. The aggregate load of a station includes the work that is still upstream so it relates to all hours of work that have been released but that have not been completed at the station. In order to estimate the direct loads at the next release moment the so-called Load conversion method has been developed in Hanover, Germany. Along with the choice of the norm type each method requires setting a large number of parameters. Workload norms and planned station throughput times must be specified for each station and also the length of the period between releases (release period) must be chosen. Additionally, a time limit can be included, which determines how far the jobs are first considered for release before their planned release dates.

A detailed analysis of those aspects that influence the timing and balancing function triggers a number of considerations that necessitate an in-depth assessment in the second phase. An important consideration relates to the synergy between timing and balancing. On the one hand load balancing contributes to timing by creating predictable station throughput times. On the other hand it may disturb the planned release sequence. If the most urgent jobs do not fit the norms, less urgent jobs could be released instead. This may especially delay jobs with a large number of operations and jobs with large processing times. The simulation study is used to evaluate the consequences of this dual interaction between the timing and balancing...
function in more detail. The balancing qualities themselves give rise to two questions: 1) Do the release methods succeed in keeping loads close to the norm levels? 2) Will fulfilment of norm levels for converted loads (method A) or aggregate loads (method B) also guarantee steady direct loads? The first question is driven by the fact that the release procedure considers the jobs in a prearranged sequence to fill the norms combined with limited theory on the determination of suitable norm levels. Several considerations play a role in the second question. With respect to method A it must be doubted whether it is sufficient to focus the load estimate on the next release moment and whether the Load conversion method is adequate. As regards method B it can be shown that a constant aggregate load level will lead to direct load fluctuations in case of high job mix variety. The above considerations lead to detailed research questions that drive the simulation study in the second research phase. In addition, they provide some guidelines for the redesign of the release methods in the third phase.

The simulation study in the second phase of the research project is focused on three questions in particular:

1. How do the classical WLC release methods perform with respect to load balancing and the timing of job release?
2. What is the influence of the norm type and the norm levels on the performance and how does this influence relate to routing variety?
3. How sensitive is the performance to the other parameters of the release methods?

The simulation model is deliberately kept elementary to enable a clear identification and analysis of performance determinants. The model includes two types of routings. The first type leads to directed flows of jobs on the shop floor and the second type leads to undirected or criss-cross flows of jobs. In both types the visited set of stations varies among the jobs. Other experimental variables are the norm types, priority rules, utilisation levels and the parameters required by the release methods. The performance measures concentrate on indicating the load balancing and timing qualities of the release methods. The main indicators are the average total throughput time of jobs and the standard deviation of job lateness respectively. Moreover, a large number of additional indicators are included for in-depth analyses of certain performance aspects.

The simulation results clarify the functioning of the WLC release methods. Lowering the workload norms reduces the average total throughput time within certain limits. This confirms the presence of load balancing qualities. Even when the norm reductions lead to a more than 40% decrease of the average throughput time on the shop floor, this decrease may still compensate the increased waiting time in the pre-shop pool, i.e. the total throughput time still decreases. An analysis of the direct
load patterns creates a more complete picture of norm influences. It shows that the direct load fluctuates around the level implied by the norm, but also that the direct load approaches the zero-level for a large part of time. The latter is caused by momentary starvation of stations which is inherent in the characteristics of job shops. Utilisation levels of 100% cannot be realised.

The timing qualities can mainly be attributed to the control of station throughput times. The sequence of job releases appears to be a minor factor. The realised moment of job release strongly differs from the planned release date when workload norms are tight. Nevertheless, the final standard deviation of lateness decreases for a large range of workload norm reductions, while jobs are prioritised on a first-come-first-served basis on the shop floor. This can only be explained from a more regular time-phased flow of jobs through the shop.

The observed performance relationship between norm type and routing type confirms the expectations regarding the strengths and weaknesses of converted and aggregate load norms respectively. Aggregate load norms are less suitable for the routing type that leads to undirected flows. In that case the position of a station in the job routing strongly varies among jobs. On the contrary, converted load norms are less suitable for direct flows where typical starting and finishing operations can be identified. The results support the conjecture that it is less sensible to restrict the estimate of direct load influences for a typical finishing operation to the next release moment: a restriction that is implied by the Load conversion method.

The classical WLC release methods only show good performance when they are combined with first-come-first-served priorities on the shop floor. The release methods reduce the effectiveness of due date oriented priority rules, while in turn due date oriented priority rules disturb the intended influence of the release methods on the arrival patterns of jobs at workstations.

The sensitivity analysis shows that the choice of planned station throughput time levels has surprisingly little influence on timing performance. It is not even beneficial to tune this parameter with the workload norm. Low workload norms levels result in a better performance when they are combined with high levels of planned station throughput times. This can be explained from the problems to release jobs with a large number of operations in time when norms are tight. Long planned station throughput times increase the priority of particularly jobs with long routings in the release procedure. Short release periods have a remarkably strong influence on the timing function. The realised sequence of order releases increasingly deviates from the planned sequence when the period between release moments is reduced. Specifically jobs with long routings and processing times are delayed before they have sufficient priority to fit the workload norms.
The third phase of this research project is devoted to the development of improved release methods, using the knowledge built up in the previous two phases. The approach followed to develop the first method is characterised by a minimal set of design constraints. This serves the purpose of avoiding the general weaknesses related to the use of workload norms for timing and balancing functions. This leads to the method SLAR (Superfluous Load Avoidance Release) which is completely free of workload norms. SLAR releases jobs on a continuous basis using two triggers for new job releases. Each trigger is accompanied by its own selection mechanism:

1) The direct load of a station becomes zero:
From the set of jobs in the pool with the first operation to be performed on the station, the job with the earliest planned start time is selected. If this set is empty, no job is released.

2) A station has completed the operation of a job and all jobs waiting for the station on the shop floor are non-urgent:
From the set of urgent jobs in the pool with a first operation on the station, the job with the shortest operation processing time is selected. If this set is empty, no job is released.

SLAR is based on a number of principles, which have been deduced from the results of the previous research phases.
1. SLAR keeps the workloads as close to zero as possible, instead of targeting on a certain norm level.
2. SLAR reacts with direct inputs from release to correct the consequences of upstream inputs.
3. SLAR requires only one parameter, to which performance is not sensitive.
4. SLAR maintains the planned release sequence between non-urgent jobs requiring the same first station.
5. SLAR switches from a timing orientation to a throughput orientation when multiple jobs become urgent.
6. SLAR is designed for compatibility with a due date oriented dispatching rule.

The approach followed in the development of the second method only allows for changes in the norm type keeping the basic classical release procedure intact. The strengths of both classical norm types are combined and only the calculation of the workload is adjusted. The new calculation can be viewed as a correction to the aggregate workload. It is similar in that each job has a constant contribution to the load from release to completion at the considered station. It is different in that the contribution is just a fraction of the operation processing time; the fraction depends on the position of the station in the order routing. The correction ensures that the load can also be viewed as an estimate of the future direct load, similar to Load conversion. However, the estimate no longer relates to the next release moment but
depends on the momentary job mix on the shop floor. It estimates the average direct load that would result if the mix of routings of released jobs remained unchanged in the long term.

The new methods are evaluated in the fourth and final phase of this research project. The opportunities for comparing SLAR directly with the periodic classical release methods are limited because of the continuous approach followed in SLAR. Therefore, the results have been related to performance with unrestricted release. It shows that SLAR enables larger reductions of throughput times on the shop floor before performance starts to deteriorate vis-a-vis the classical methods. More significantly, SLAR is able to create the same relative performance improvements in combination with a due date oriented priority rule on the shop floor. The results allow for the conclusion that SLAR is a method that avoids the weaknesses regarding timing and balancing that could be attributed to the use of workload norms.

The second redesign, the newly developed norm type, has been tested for both directed and undirected flow of jobs. The new norm type is able to create a similar or better performance than the classical norm type that performed best in the considered environment. Thus, it is shown that the observed strengths of the classical norm types can be combined in a more robust norm type.

As has been intended, new knowledge on the functioning of WLC release methods has been developed by exposing the opportunities for improvement and by the actual development of improved methods. As a result, this knowledge is being utilised in practical applications of the WLC concept.