Load-Based Generic Polca: Performance Assessement Using Simulation

Nuno O. Fernandes[1(✉)] and Sílvio Carmo-Silva[2]

[1] Instituto Politécnico de Castelo Branco, Av. do Empresário, 6000-767 Castelo Branco, Portugal
nogf@ipcb.pt

[2] ALGORITMI Research Unit, Department of Production and Systems, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Abstract. POLCA (i.e. Paired-cell Overlapping Loops of Cards with Authorization) is a card-based decision support system for production control, developed to support the adoption of Quick Response Manufacturing. Two variants of POLCA have been proposed in the literature to improve POLCA performance: Load Based POLCA and Generic POLCA. In this paper, we combine these two variants into a single production control system and analyse its performance for different backlog-sequencing rules. The results of a simulation study carried out for a make-to-order flow shop, support the strategy of combining these two POLCA variants and show that capacity-slack backlog sequencing based on corrected aggregate load have the potential for improving performance.

Keywords: Production control · Generic POLCA · Simulation

1 Introduction

POLCA (i.e. Paired-cell Overlapping Loops of Cards with Authorization) is a card-based decision support system for production control that was developed to support the adoption of Quick Response Manufacturing. POLCA was designed specifically for low-volume, high-variety environments and controls the flow of jobs through the production system by a combination of release authorisations and a WIP cap [1]. It makes use of overlapping loops of cards between pairs of successive work cells in the routing of a job, imposing a WIP cap in every loop.

POLCA has remained largely unchanged since its introduction [2]. Two modifications proposed in the literature that resulted in increased effectiveness are Generic POLCA [3] and Load-Based POLCA (LB-POLCA) [4]. Generic POLCA changed the original loops structure of POLCA: a job cannot start processing on the first workstation (or cell) until all the cards for entire job processing are available to be attached to it. This means that, before processing can start at the first workstation, production capacity must be reserved at all downstream workstations in the job routing. LB-POLCA changed the original unit-based system into a load-based version. It was thought for providing a more adequate and robust representation of available capacity in production environments in which the operation times of jobs vary significantly and

G.I. Gheorghe (ed.), *Proceedings of the International Conference of Mechatronics and Cyber-MixMechatronics - 2017*, Lecture Notes in Networks and Systems 20,
DOI 10.1007/978-3-319-63091-5_10

product mix changes occur [4]. It also avoids the problem of defining the quantum of the POLCA cards, i.e. the amount of work (e.g. hours) each card represents, and the problem of constantly fine-tuning the number of cards if the demand and/or mix of products changes [4].

Little research has been published on these two POLCA variants. An exception is [5], where the LB-POLCA performance is assessed dependent on load accounting approaches. In analyzing POLCA variants a major question emerges: what performance could we expect from combining both, LB-POLCA and Generic POLCA, i.e. from a Load-Based Generic POLCA (LB-GPOLCA) system? Once the backlog-sequencing decision may influence load balancing across work cells and thus system performance [6–8], a second question arises: how can the backlog-sequencing rule be used to increase the effectiveness of LB-GPOLCA system?

An exploratory study based on controlled simulation experiments is used to answer these questions. We will show that capacity-slack backlog sequencing based on the corrected aggregate load have the potential to improve system performance.

The remainder of this paper is structured as follows. The simulation model used to evaluate performance is described in Sect. 2, and in Sect. 3 results are presented, discussed and analysed. Finally, conclusions are drawn in Sect. 4, where managerial implications and future research directions are also outlined.

2 Simulation Study

In this section, the simulation model considered in the study, the experimental design and the measures used to evaluate system performance are detailed.

- **Simulation Model**
  A simulation model of a pure flow shop has been implemented using ARENA software. In the pure flow shop, each job visits all stations in the same sequence in order of increasing station number. Our model is stochastic, whereby job routings, operations times, inter-arrival times and due dates are random variables. The shop contains six stations, where each station is a single constant capacity resource. A station is required at most once in the routing of a job.
  Operation times follow a truncated 2-Erlang distribution with a maximum of 4 h and a mean of 1 time unit after truncation. Set-up times are considered as part of the operation time. Meanwhile, the inter-arrival time of jobs follows an exponential distribution with a mean of 1.111 h, which deliberately results in a utilization level of 90%. Due dates are set exogenously by adding a random allowance factor, uniformly distributed between 40 and 60 h, to the job entry time. The minimum value will be sufficient to cover a minimum shop floor throughput time corresponding to the maximum operation time (4 h) for the maximum number of possible operations (6) plus an arbitrarily set allowance for the waiting or queuing times.
- **Job Release and Dispatching**
  As in previous simulation studies, e.g. [3, 9], it is assumed that all materials are available and all necessary information regarding shop floor routing and processing

times is known upon the arrival of an order to the shop. Orders flow into a pre-shop pool (or backlog) to await release according to the LB-GPOLCA method.

Seven workload limits are applied, ranging from 4 to 10 h, and infinity. These limits apply to all stations, as the corrected aggregated load [10] was used for load accounting and the shop is balanced. They have been chosen based on preliminary simulation runs, allowing a better insight into of the performance impact of the experiment factors. In LB-GPOLCA, workload is accounted for all stations in the routing of the job (except the gateway) from release to moment the corresponding operation is completed at the station.

LB-GPOLCA uses a backlog-sequencing rule to determine the sequence in which jobs (or orders) are considered for release. Four sequencing rules have been considered in this study:

- Earliest Release Date (ERD), this is the rule advocated in generic POLCA [3]. In our study, the earliest release date of a job is calculated by backward scheduling from the job due date the estimated throughput time for each operation in the routing of the job. The allowances are given by the running average of the realized operation throughput times. Note that once all jobs have the same routing across stations ERD transforms into earliest due date (EDD).
- Shortest Total Work Content (STWK), a load-oriented rule that sequences jobs according to the sum of all processing times in the routing of an order.
- Capacity Slack CORrected (CScor) prioritizes jobs using a capacity slack ratio Sj as given by Eq. (1). The lower the capacity slack ratio of job j, the higher the priority. The rule integrates two elements into one priority measure: the load contribution of a job to a station s, LSj, in time units; and the load gap, i.e., the difference between a load norm NS and the current corrected aggregate load at station WS corresponding to operation i: $N_S - W_S$.

$$S_j = \sum_{s \in R_j} \frac{L_{sj}}{(N_s - W_s)} \tag{1}$$

where: Rj is the set of workstations in the remaining routing of job j.

- Capacity Slack number of jobs in the direct load (CSjobdir), which replaces corrected aggregate load at station W_s in Formula 1 by the direct load queuing at the station measured in terms of the number of jobs, i.e. the load that queues and is in processing at a station.

Concerning job release, two input control strategies are considered in the study, namely: (1) job release is determined by the workstations load and; (2) is determined by the workstations load and by the number of jobs at the input buffer of the first workstation in the routing of the job, i.e. the gateway queue. In the latter situation job release is only allowed if the gateway queue is empty. This means that most of the jobs instead of waiting at the buffer of the gateway workstation they will wait in the backlog.

Concerning job dispatching, i.e. the decision on which job in queue to process next, the Earliest Operation Due Date (EODD) rule [11] is used at all machines of the shop floor.

- **Experimental Design and Performance Measures**
  The experimental factors and the levels at which they were tested in the study are (see Table 1): (i) the backlog sequencing rule (ERD, STWK, CScor, CSjobdir); (ii) the input control strategy (workstations load, workstations load and gateway queue); and (iii) seven load norms for the workload that is allowed at each loop. A full factorial design was used with 56 scenarios, where each scenario was replicated 100 h. All results were collected over 13,000 h following a warm-up period of 3,000 h. These parameters allow us to obtain stable results while keeping the simulation run time to a reasonable level.

Table 1. Experimental factors and levels

Experimental factor	Levels
Backlog sequencing rule	ERD|STWK|CScor|CSjobdir
Input control strategy	Workstations load|workstations load and gateway queue
Load norm (h)	4, 5, 6, 7, 8, 10 and infinity

Four main performance measures are considered in this study as follows: (1) mean total throughput time, i.e., the mean of the completion date minus the arrival time date across jobs; (2) percentage tardy, i.e., the percentage of jobs completed after the due date; (3) mean tardiness; and (4) the standard deviation of lateness. The total throughput time is used as the main indicator of the balancing capabilities of the approaches being tested.
The main indicator of delivery performance is the percentage of tardy jobs, which is influenced by both the average lateness and the dispersion of lateness across jobs. In addition to the four main performance measures, we also measure the average shop floor throughput time as an instrumental performance variable. While the total throughput time includes the time that an order waits before being released, the shop floor throughput time only measures the time after an order is released to the shop floor.

3 Simulation Results

This section presents and discusses the results of the simulation study. To aid interpretation, results are presented in the form of performance curves. The left-hand starting mark of the curves represents the tightest load norm (4 h). The load norm used increases step-wise by moving from left to right in each graph, with each data mark representing one load norm. The right-hand mark represents an infinite load norm, meaning unrestrictive release of jobs to the shop floor for the continuous line curves and release based only on the gateway queue for dashed line curves. Loosening the load norm increases the level of work-in-process and, thus increases the shop floor throughput times.

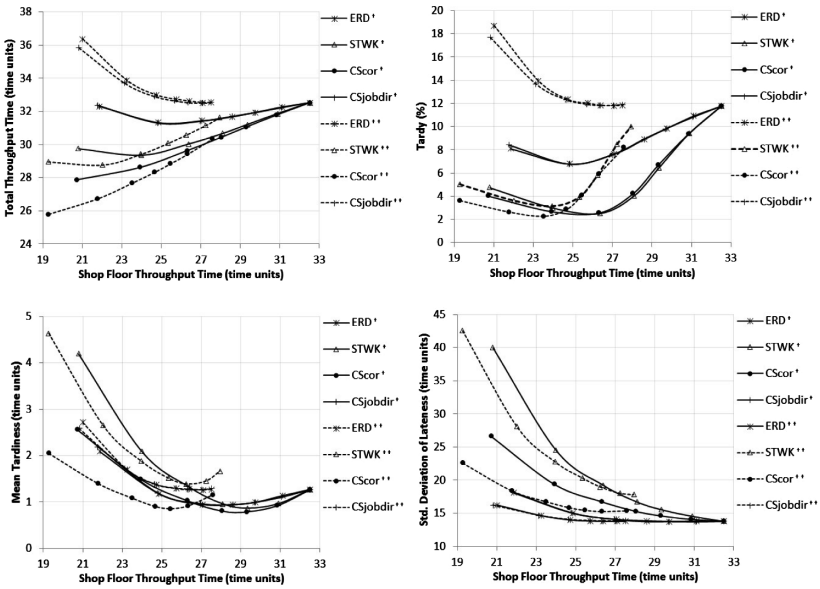

Fig. 1. Performance results for different backlog sequencing rules and input control strategies. $^{+}$results for input control based on workstations load;$^{++}$results for input control based on workstations load and the gateway queue.

Figure 1(a–d) shows the total throughput time, percentage tardy, mean tardiness, and standard deviation of lateness results over the shop floor throughput time, respectively. Analysing results, we can see that restricting the workload that is released to the shop floor by using LB-GPOLCA improves performance, i.e., results in the lower values of total throughput time, percentage tardy and mean tardiness, if load norms are not set to tight. This means that LB-GPOLCA outperforms immediate release.

We also can observe from Fig. 1, that job release subject to the queue size of the input buffer of the gateway workstation, shifts curves left, i.e., reduces the shop floor throughput time. This is due to fewer jobs that are released to the shop floor, as they stay in the backlog until the input buffer of the first machine falls to zero.

Concerning the behaviour of the backlog sequencing rules, CScor leads to the best performance, while the ERD and CSjobdir leads to the worst. This is independent of the input buffer strategy used. Meanwhile, the performance ERD and CSjobdir strongly deteriorates if input control is exercised based on both, the workstations load and the queue size of input buffer of the gateway workstation. Once LB-GPOLCA does not impose a load limit on the gateway workstation, capacity slack rules, such as CScor, are particularly relevant to balance workload balancing across workstations. Concerning load balancing, our study shows that CSjobdir seems to be inadequate.

Finally, STWK results in a relatively low percentage of tardy jobs. However, this is obtained at the cost of a higher standard deviation of lateness and mean tardiness. This

means that few jobs (particularly large jobs in terms of the work content) are being delayed at the pool, but for a long time.

4 Conclusions

This paper discusses a load-based version of the Generic POLCA system (LB-GPOLCA). Two major research questions emerged in our study: how LB-GPOLCA performs in the context of make-to-order flow shop? Which rule must be used for backlog sequencing? Based on insights from the Workload Control literature four different backlog sequencing rules were proposed and tested using simulation. Our results indicate that LB-GPOLCA outperforms immediate release and that a capacity slack rule based on the corrected workload for sequencing jobs before they are release to the shop floor is likely to perform well and is, therefore, recommended.

By pointing out the influence of backlog sequencing decision in LB-GPOLCA, this study has obvious managerial implications, if we practitioners need to deal with this for controlled job release. Future research work should extend the study to other shop configurations and production settings to enlarge the scope of recommendations suggested by the results.

Acknowledgements. This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013 and by Instituto Politécnico de Castelo Branco.

References

1. Suri, R.: Quick Response Manufacturing: A Companywide Approach to Reducing Leadtimes. Productivity Press, New York (1998)
2. Riezebos, J.: Design of POLCA material control systems. Int. J. Prod. Res. **48**(5), 1455–1477 (2010)
3. Fernandes, N.O., Carmo-Silva, S.: Generic POLCA—a production and materials flow control mechanism for quick response manufacturing. Int. J. Prod. Econ. **104**(1), 74–84 (2006)
4. Vandaele, N., Van Nieuwenhuyse, I., Claerhout, D., Cremmery, R.: Load-based POLCA: an integrated material control system for multiproduct. Multimach. Job Shops Manuf. Serv. Oper. Manag. **10**(2), 181–197 (2008)
5. Fernandes, N.O., Thürer, M., Stevenson, M., Carmo Silva, S.: Load-based POLCA: an assessment of the load accounting approach. In: Rocha, Á., Correia, A., Adeli, H., Reis, L., Costanzo, S. (eds.) Recent Advances in Information Systems and Technologies. WorldCIST 2017. Advances in Intelligent Systems and Computing, vol. 570. Springer, Berlin (2017)
6. Philipoom, P.R., Malhotra, M.K., Jensen, J.B.: An evaluation of capacity sensitive order review and release procedures in job shops. Decis. Sci. **24**(6), 1109–1133 (1993)
7. Fredendall, L.D., Ojha, D., Patterson, J.W.: Concerning the theory of workload control. Eur. J. Oper. Res. **201**(1), 99–111 (2010)

8. Thürer, M., Land, M.J., Stevenson, M., Fredendall, L.D., Godinho Filho, M.: Concerning workload control and order release: the pre-shop pool-sequencing decision. Prod. Oper. Manag. **24**(7), 1179–1192 (2015)
9. Thürer, M., Stevenson, M., Protzman, C.W.: COBACABANA (Control of Balance by Card Based Navigation): an alternative to Kanban in the pure flow shop? Int. J. Prod. Econ. **166**, 143–151 (2015)
10. Oosterman, B., Land, M.J., Gaalman, G.: The influence of shop characteristics on workload control. Int. J. Prod. Econ. **68**(1), 107–119 (2000)
11. Lödding, H., Piontek, A.: The surprising effectiveness of earliest operation due-date sequencing. Prod. Plan. Control **28**, 459–471 (2017)