Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

A NEW APPROACH TO MULTI-PASS SCHEDULING IN SHOP FLOOR CONTROL

Taejong Yoo
Daehong Kim
Hyunbo Cho

Department of Industrial Engineering
Pohang University of Science and Technology
San 31 Hyoja, Pohang 790-784, REPUBLIC OF KOREA

ABSTRACT

Real-time planning and scheduling in a shop floor are not easy to accomplish due to the concurrent flow of various parts as well as sharing of different types of resources. Multi-pass scheduling is a well known method for solving the aforementioned problem. Its success depends largely on selecting the best decision-making rule fast and effectively. Although many efforts have been made in the past, a way to minimize the computational load of rule evaluation and selection has yet to appear. The objective of the paper is to apply a nested partitioning (NP) method and an optimal computing budget allocation (OCBA) method to reduce the computational load without the loss of the performance of multi-pass scheduling. The experimental design and analysis was performed to validate that NP and OCBA can be successfully applied to multi-pass scheduling in order to enhance the performance of multi-pass scheduling.

1 INTRODUCTION

Traditionally, scheduling problems have been formulated using analytical methods like mathematical programming or network theory to provide optimal solutions under the simplified assumptions that do not reflect actual shop floor status (Chuda and Mize 1994). However, these approaches are not appropriate for real-time scheduling, due to the characteristics of dynamic shop floor, in other words, non-deterministic disturbances (e.g., machine breakdown, tool breakage, rush order) and the concurrent flows of various parts as well as sharing of different types of resources (Davis and Jones 1988).

The aforementioned difficulties led to research into rule-based approaches. These are usually dispatching rules used to prioritize the different jobs competing for the use of a given machine. In addition to the dispatching problem, real-time shop floor scheduling includes other types of decision-making problems, such as job releasing problem, job sequencing problem, etc. These problems can be resolved by event-driven job scheduling mechanism, in which each event requiring decision-making is resolved by a specific decision-making strategy. However, since the past researches indicate that the strategies' performance depends on the shop's conditions (Jeong and Kim 1998, Kutanoglu and Sabuncuoglu 2002), it would be better to change the strategies dynamically as well as at the right moment according to the conditions, instead of using a fixed scheduling strategy for every scheduling period. This adaptability gave birth to multi-pass scheduling.

Multi-pass scheduling has become a promising approach which ranks and selects the best one among the strategies by looking ahead simulation-based multiple courses of action before actual execution (Wu and Wysk 1988, Cho and Wysk 1993). Multi-pass scheduling for real time shop floor control requires a speedy response in evaluating the rule combination. The response time would depend on the number of simulation trials necessary to evaluate the rule combinations before ranking them.

The objective of the paper is to propose the methodology to speed up multi-pass scheduling used for real-time shop floor control. The efficiency depends on how fast the best rule combination is obtained. The recommended rules are then evaluated using a nested partitioning (NP) method, in which the number of simulation replications is reduced by using an optimal computing budget allocation (OCBA) method. Experimental design and analysis are performed to demonstrate the efficiency and effectiveness of the proposed methodology.

2 LITERATURE SURVEY

A multi-pass scheduling framework consists of the five components: recommendation of rules for each problem type, generation of all the rule combinations, simulation, evaluation and rank of rule combinations, and scheduling. Whenever the decision-making rules need changing, a set of promising rules for resolving each problem type are recommended. Since a few different types of scheduling problems may occur during the next scheduling period, all the rule combinations must be generated, each of which will be then evaluated by using simulation. The best rule combination is chosen and conveyed to the scheduling component. It is noted that the scheduling period during which simulation looks ahead is called a 'simulation window.'

The most computation load in a multi-pass scheduling framework occurs inevitably in simulation due to several types of scheduling problems, many rules, and replications of each simulation. To reduce the number of decision-making rules for each problem type, several methods have been utilized to quickly recommend a small number of candidate rules.

A top-down three level approach for rule selection has been developed in a prototype expert scheduling system. Another approach is to combine a rule-based expert system in order to select several scheduling rules in a real-time environment, and a multi-pass simulator in order to test which scheduling rule is the best (Wu and Wysk 1988). Neural networks have also been used to recommend candidate rules for multi-pass scheduling (Cho and Wysk 1993). Inductive learning and genetic algorithms were also used to build the relationship between shop conditions and rules (Jones *et al*. 1995). Some modified techniques of using neural networks (Kim and Kim 1994) and production rules have been used for identifying the relationships.

3 MULTI-PASS SCHEDULING USING NP AND OCBA

3.1 Introduction of NP

The nested partitioning (NP) method used for solving deterministic optimization problems with large but finite feasible region employs a global sampling strategy that is continuously adapted via a partitioning of the feasible region (Shi and Ólafsson 2000). The method can be also applied when no expression exists for the objective function, but its value is estimated through simulation. The NP method can be briefly described as follows.

- Step 0: The entire feasible region is considered the most promising region.
- Step 1: The most promising region is partitioned into disjoint subregions, unless it contains a single point.
- Step 2: Independent points are selected from each of these subregions by using a systematic random sampling procedure
- Step 3: The promising index for each subregion is estimated from the samples.
- Step 4: The most promising subregion is determined by using the estimated index. If more than one subregion is equally promising, these subregions are merged into a single subregion.
- Step 5: The selected subregion now becomes a feasible region in the next iteration and go to Step1.

This generates a sequence of region partitions, with each partition nested within the last. The final region contains only one point. When the NP method is applied to multi-pass scheduling, a rule combination is considered a point in the feasible region; while a particular subset of rule combinations are considered a disjoint subregion.

3.2 Introduction of OCBA

Suppose that the promising index for each subregion in the Step 3 of the NP method is estimated via simulations of all the samples. In other words, suppose that each rule combination in a subregion is evaluated with regard to its performance via simulation. The OCBA method can be applied to reduce the number of replications (Chen *et al*. 1997, Chen *et al*. 2003). Instead of allocating the equal number of replications to every simulation, the OCBA method employs a two stage approach. During the first stage, a small number of replications are applied to calculate the mean and variance, which are then used to determine the additional number of replications. During the second stage, the potentially promising rule combinations are simulated with more replications. This procedure can be briefly described as follows:

- Step 0: The initial number of replication is previously assigned.
- Step 1: The performance of each rule combination is calculated after initially assigned simulation replications.
- Step 2: It is checked whether or not additional replications are needed. If no additional replication is needed, stop.
- Step 3: The additional replication of each rule combination is calculated and go to step 1.

3.3 Proposed Multi-Pass Scheduling Mechanism

A new multi-pass scheduling framework is shown in Figure 1. The NP method reduces the number of rule combinations to be evaluated and ranked, and the OCBA method reduces the number of simulation replications of each rule combination.

The scheduling rules being used need changing, a number of promising rules are recommended for each problem type. All the rule combinations used to solve the problem types encountered during the next scheduling window are then generated. The NP method is applied for selecting the best rule combination. To do this, all the rule combinations are partitioned into several disjoint sets. Second, a few representative rule combinations are sampled from each set. The methods are detailed in Section 4.

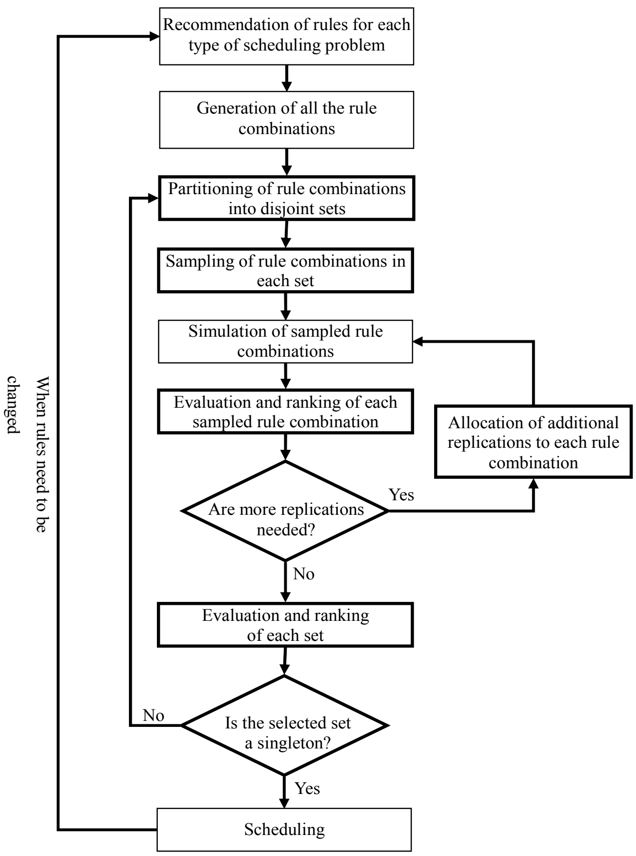

Figure 1: Detailed Flow Diagram of Proposed Multi-Pass Scheduling

All the sampled rule combinations are evaluated by using simulation respectively. Each rule combination is simulated by a small number of replications and then its performance is evaluated. If the probability of correct selection of the best sampled rule combination is less than a pre-specified threshold, the more replications must be allocated to each sampled rule combination. Once all the sampled rule combinations are evaluated, the performance criterion of each set is calculated. Unless the most promising set is a singleton, the further partitioning is performed and the above procedure is repeated.

4 STRATEGIES OF NESTED PARTITIONING

4.1 Partitioning Strategy

The basic concept of partitioning is to partition the most promising set selected at previous iteration into disjoint subset to be evaluated at the next iteration. At each iteration, the most promising set is obtained through one time performance of the procedure presented in Section 3. The partitioning strategy is crucial for the speedy convergence to the best rule combination. In multi-pass simulation, the number of rule combinations at each subset is identical. It implies that the number of partitioning iterations is equal to the number of problem types. The partitioning is repeated until only a single rule combination is selected. For the partitioning efficiency, the sequence of partitioning depends on the number of rules recommended for each problem type. The smaller the number of rules is, the earlier the rules are partitioned. This results in less computational load.

Suppose that there exist n different types of scheduling problems and m_i (i = 1,…, n) different rules recommended for each type of scheduling problem. If $m_1 < m_2 < \ldots < m_n$, then the number of iterations is n and the number of partitions in iteration i is m_i. A particular subset at the k^{th} iteration which is related to a node has all the rule combinations starting with pre-determined k preceding rules.

4.2 Sampling Strategy

Another issue is how to obtain the samples of rule combinations used to evaluate each subset at every iteration. The sampling size and method in each subset must be determined in such a way that the accuracy of finding the most promising rule combination is maximized. In multi-pass scheduling, because the all sample must be picked when the last partitioning is performed, the 50% sampling size is adopted. The method of systematic random sampling selects rule combinations at each subset throughout the sampling frame after a random start in the first subset.

An identical number of samples from each subset are sampled. When the k^{th} iteration is performed, the most promising set is partitioned into m_k subsets. If the size of each set is eight, the four samples are picked in each subset by the sampling strategy. The mean of the performance measures obtained by simulating the sampled rule combinations is used for the evaluation and ranking of each subset.

5 STRATEGIES OF OPTIMAL COMPUTING BUDGET ALLOCATION

5.1 Performance Measure of Rule Combination

Although the number of rule combinations to be evaluated is reduced at the partitioning and sampling stage in each iteration, the computation load can be still very high because each sampled rule combination must be simulated with a certain number of replications. However, the number of replications necessary to obtain the mean of the performance measure must be large due to the slow convergence rate of Monte Carlo simulation. If both the number of replications and the number of rule combinations are large, the total number of simulation replications required to estimate the mean of a particular subset can be extremely high. Therefore the number of replications needs to be minimized.

In order to calculate the mean of the performance measure of a particular set in a certain iteration, first, the OCBA method assigns the identical number of replications to the simulation of each sampled rule combination. Each rule combination is then pre-evaluated and pre-ranked according to the initial results. The OCBA allocates the more replications to the potentially critical rule combinations and the simulations are performed with additional replications until the probability of correct selection satisfies some degree of confidence interval. In conclusion, all the sampled rule combinations have their own unbiased point estimator through the enough simulation.

5.2 Stopping Criterion in Each Iteration

The OCBA method helps determine the optimal number of replications necessary to simulate the sampled rule combinations. Given a pre-specified number of replications, 'the probability of correct selection' of a rule combination is used as a criterion about whether or not more replications are needed. 'Correct selection' can be defined as the event that the selected best rule combination is actually the best. Therefore, 'the probability of correct selection' can be defined as in Equation (1), assuming that the performance measure needs to be minimized.

$$P\{CS\} = P\{\mu_{best}(N_{best}) < \mu_i(N_i)\} \text{ for all } i \neq best \quad (1)$$

where N_i is the number of simulation replications for rule combination i and μ_i is the mean of the performance measure of rule combination i.

Since Equation (2) is not easy to compute, the approximate probability of correct selection (APCS) can be estimated and used for the probability of correct selection (Chen, 1996). Note that the computation of APCS is simply a product of pair comparison probabilities.

$$P\{CS\} \geq \prod_{i \neq best}^{M} P(\mu_{best} < \mu_i) \equiv APCS \quad (2)$$

where M is the sample size of a subset. If APCS is less than a pre-specified confidence level, more replications are needed.

5.3 Determination of Additional Replications

The number of additional replications is determined according to the following two steps. First, the mean and variance of each sampled rule combination are obtained from simulation with initial replications. Second, Equation (3. 1) and (3. 2) are applied to obtain the additional number of replications. Equation (3.1) enables the more additional replications to be allocated to the rule combination with the best performance measure than that with the second best performance measure. Equation (3.2) enables the less additional replications to be allocated to the rule combination with the rest performance measure than that with the second best performance measure. This procedure guarantees that the rule combination with better performance measure is allocated with more replications.

$$\frac{N_{best}}{N_{second}} = \frac{\sigma_{best}}{\sigma_{second}}\left[\sum_{i \neq best}^{M}\left(\frac{(\mu_{best}(N_{best}) - \mu_{second}(N_{second}))^2}{(\mu_{best}(N_{best}) - \mu_i(N_i))^2}\right)\right]^{1/2} \quad (3.1)$$

$$\frac{N_i}{N_{second}} = \left(\frac{\sigma_i/(\mu_{best}(N_{best}) - \mu_i(N_i))}{\sigma_{second}/(\mu_{best}(N_{best}) - \mu_{second}(N_{second}))}\right)^2 \quad (3.2)$$

for all $i \neq best \neq second$

where, σ_i is the standard deviation of rule combination i, '*best*' is the rule combination with the best mean of performance measure, and '*second*' is the rule combination with the second best mean of performance measure.

6 EXPERIMENTAL DESIGN AND ANALYSIS

6.1 Experiment Assumptions

If all the factors in a shop floor are considered, the number of simulations may be infinite. It is necessary to create a subset of different shop floor configurations in order to avoid the combinatorial explosion of simulation runs. Some assumptions are made as follows:

1. Performance criterion is average flow time.
2. Number of different part types is 5.
3. Dispatching strategies: EDD, FIFO, STT.
4. Part releasing strategies: SPT, EDD, FIFO, STT.
5. Each part's process time is constant.
6. Each part's transportation time is constant.
7. Setup time of each machine is constant.
8. Machine-breakdowns are not considered.

6.2 Experiment Design

6.2.1 Factorial Design on the Probability of Correct Selection

The objective of the first experiment is to show that the probability of correct selection of the multi-pass scheduling using NP and OCBA is superior to that of the multi-pass scheduling not using NP and OCBA. Therefore, the methods of multi-pass scheduling and the assigned budget are controllable factors. A full factorial design for the controllable factors is used, which results in 14 (= 2×7) different experimental base design.

1. Methods of multi-pass scheduling: Use of NP and OCBA, Conventional
2. Budget: 1000, 2000, 3000, 4000, 5000, 6000, 7000

It is necessary to construct uncontrollable factors over which experiments are performed. Three different factors are considered: number of machines, variance of processing time, and variance of transportation time. A full factorial design for uncontrollable factors is used, which results in 18 (= 2×3×3) different shop floor conditions. Then, including controllable factors, the number of total experiment treatments amounts to 252 (= 18×14).

1. Number of machines: 3 and 5.
2. Variance of processing time: 0.05, 0.1, and 0.15 × processing time
3. Variance of transportation time: 0.05, 0.1, and 0.2 × transportation time

6.2.2 Experiment on the Total Number of Replications

The objective of the second experiment is to show that the number of replication simulated of the multi-pass scheduling using NP and OCBA is smaller than that of the scheduling not using NP and OCBA to reach given probability of correct selection above 0.95. Therefore, the methods of multi-pass scheduling is controllable factor.

1. Methods of multi-pass scheduling: Use of NP and OCBA, Conventional

The uncontrollable factors used for experiments are equal to those of the first experiment. The number of total experiment treatments amounts to 36 (= 18×2).

6.3 Experimental Results

6.3.1 Experimental Results on the Probability of Correct Selection

Table 1 shows the experimental results. A higher budget can obtain a higher probability of correct selection.

Table 1: Results of the First Experiment

Design	Method	Budget	Mean
1	Conventional	1000	0.600262
2	NP+OCBA	1000	0.765882
3	Conventional	2000	0.712404
4	NP+OCBA	2000	0.900655
5	Conventional	3000	0.745334
6	NP+OCBA	3000	0.936086
7	Conventional	4000	0.80584
8	NP+OCBA	4000	0.953064
9	Conventional	5000	0.832854
10	NP+OCBA	5000	0.959721
11	Conventional	6000	0.869779
12	NP+OCBA	6000	0.974814
13	Conventional	7000	0.865288
14	NP+OCBA	7000	0.982897

To claim that the accuracy of the proposed method is superior, ANOVA is performed. The paired differences between the proposed method and the conventional method are compared using the following hypothesis. The results of the ANOVA test are summarized in Table 2. The difference is significant.

$$H_0\text{: } P\{CS\}_{\text{using NP+OCBA}} = P\{CS\}_{\text{using conventional method}}$$

Table 2: ANOVA Test of the Correct Selection Probability

Design	Budget	F-value	P-value
1&2	1000	7.543099	0.009561
3&4	2000	16.46098	0.000275
5&6	3000	19.13259	0.00011
7&8	4000	13.2968	0.00088
9&10	5000	11.02996	0.00215
11&12	6000	11.2133	0.001996
13&14	7000	12.29353	0.001299

6.3.2 Experimental Results on the Total Number of Replications

As shown in Table 3, the total numbers of replications of the proposed method are five times smaller than those of the conventional method.

Table 3: Results of the Second Experiment

Design	Method	Mean
1	Conventional	5,875.200
2	NP+OCBA	1,152.889

To determine if the experiment provides enough evidence to claim that the time saving of the using NP and OCBA is superior to the conventional method, analysis of variance (ANOVA) is run. The paired difference between two methods is compared as follows:

H_0: Total replication number $_{\text{using NP+OCBA}}$ = Total replication number $_{\text{using conventional method}}$

The results of the ANOVA test showed that F-value is 9.942959 and P-value is 0.003499. It implies that the proposed method has a great effect on the time saving on replication to ensure the correct selection of the best design.

7 CONCLUSION

Multi-pass scheduling ranks and selects the best decision-making rule by looking ahead simulation-based multiple courses of action before actual execution. Even though it has been known that multi-pass scheduling performs much better than single pass scheduling, its disadvantage is to take too much time in evaluating possible rule candidates. This paper proposed a new multi-pass scheduling framework in which the number of rules to be evaluated is minimized by using a nested partitioning method and the number of replications for simulation is also minimized by using an optimal computing budget allocation method and guarantees its achievement through the application of the multi-pass simulation. The NP method reduced the number of strategies by using partitioning of the strategy regions and sampling each group. The OCBA method determines the simulation replications by allocating additional replications to potentially critical decision-making rules.

To show the efficiency of the scheduling using NP and OCBA, the experimental design and analysis has been performed and compared with that of the conventional multi-pass scheduling. The results showed that the proposed framework has a great effect on the accuracy of rule selection and the time saving for stochastic environment of a shop floor. Consequently, multi-pass scheduling using NP and OCBA is applicable for real-time shop floor scheduling in an efficient and effective manner.

ACKNOWLEDGMENTS

This research has been sponsored by the Korean Ministry of Science and Technology and the French Embassy in Seoul.

REFERENCES

Chen, C. H., 1996, A lower bound for the correct subset-selection probability and its application to discrete-event system simulations, *IEEE transactions on automatic control*, 41(8):1227-123

Chen, C. H., E. Yűcesan, H. C. Chen, and L. Dai, 1997, New development of optimal computing budget allocation for discrete event simulation, In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 334-341. Piscataway, New Jersey: Institute for Electrical and Electronics Engineers.

Chen, C. H., E. Yűcesan, J. Lin, and K. Donohue, 2003, Optimal computing budget allocation for Monte Carlo simulation with application to product design, *Simulation Modeling Practice and Theory*, 11:57-74.

Cho, H. and R. A. Wysk, 1993, A robust adaptive scheduler for an intelligent workstation controller, *International Journal of Production Research*, 31:771- 789.

Chuda, B. and J. H. Mize, 1994, Scheduling and control of flexible manufacturing systems: a critical review, *International Journal of Computer Integrated Manufacturing*, 7:340-355.

Davis, W. J. and A. T. Jones, 1988, A real-time production scheduler for a stochastic manufacturing environment, *International Journal of Computer Integrated Manufacturing*, 4(4):531- 544.

Jeong, K.-C. and Y.-D. Kim, 1998, A real-time scheduling mechanism for a flexible manufacturing system: using simulation and dispatching rules, *International Journal of Production Research*, 36:2609- 2626.

Jones, A. T., L. Rabelo, and Y. Yih, 1995, A Hybrid approach for real-time sequencing and scheduling, *International Journal of Computer Integrated Manufacturing*, 8(2):145- 154.

Kim, M.-H. and Y.-D. Kim, 1994, Simulation-based real-time scheduling mechanism in a flexible manufacturing system, *Journal of Manufacturing Systems*, 13(2):85- 93.

Kutanoglu, E. and I. Sabuncuoglu, 2002, Experimental investigation of iterative simulation-based scheduling in a dynamic and stochastic job shop, *Journal of Manufacturing Systems*, 20(4):264-279.

Shi, L. and S. Olafsson, 2000, Nested partitions method for stochastic optimization, *Methodology and Computing in Applied Probability*, 271-291.

Wu, S.-Y. D. and R. A. Wysk, 1988, Multi-pass expert control system - a control/scheduling structure for flexible manufacturing cells, *Journal of manufacturing systems*, 7(2):107-120.

AUTHOR BIOGRAPHIES

TAEJONG YOO is a Ph.D. candidate in the Industrial Engineering Department at the Pohang University of Science and Technology. His research area includes distributed simulation for supply chain execution and integration. His e-mail address is <tju13@postech.ac.kr>.

DAEHONG KIM is a M.S. candidate in the Industrial Engineering Department at the Pohang University of Science and Technology. His research area includes supply chain execution and integration. His e-mail address is <dsanddh@postech.ac.kr>.

HYUNBO CHO is an Associate Professor in the Industrial Engineering Department at the Pohang University of Science and Technology. He received his Ph.D. in Manufacturing Engineering from the Texas A&M University. His research interests include supply chain integration, systems intelligence, enterprise integration. He serves as an editorial board member of the Journal of Manufacturing Systems and the International Journal of Industrial Engineering-Applications and Practice, and the Journal of Manufacturing Systems. He is a member of IIE and SME. His e-mail address is <hcho@postech.ac.kr>.