Contents lists available at ScienceDirect

CIRP Annals
Manufacturing Technology
Journal homepage: www.elsevier.com/locate/cirp

Robust production control against propagation of disruptions

T. Tolio (1)[a,b], M. Urgo[b,*], J. Váncza (1)[c,d]

[a] ITIA-CNR, Institute of Industrial Technologies and Automation, National Research Council, Milano, Italy
[b] Manufacturing and Production Systems Division, Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy
[c] Fraunhofer Project Center for Production Management and Informatics, Computer and Automation Research Institute, Hungarian Academy of Sciences , Budapest, Hungary
[d] Dept. of Manuf. Science and Technology, Budapest University of Technology and Economics, Budapest, Hungary

In hierarchical production control systems, planning decides on capacities and operations so as to meet demand, while scheduling should guarantee the execution of production plans even in face of uncertainties. The management practice advocates rolling horizon approaches despite the danger of plan nervousness. We propose a novel framework to handle uncertainties closer to the root of their sources, when scheduling local resources. The method keeps the complexity of planning and scheduling at bay and stops the propagation of local disruptions to other resources. The paper presents the theoretical model, the solution technique, and shows their applicability on a case study taken from the tool industry.

Production planning, Scheduling, Stochastic modeling

1. Introduction and problem statement

Hierarchical production planning and scheduling (PPS) techniques aim at reducing the complexity and handling the inherent uncertainty of production control problems through aggregation along the dimensions of demand, products, resources, operations and time [1]. As for the temporal aspect, all approaches have in common that they tackle the problem of matching demand and supply on distinct time planes and horizons. The production planning phase deals with how and when to produce in the medium term, considering customer orders together with their delivery times, as well as material and resource availability [2]. Planning is aimed at optimizing a number of criteria, like service level, efficient resource utilization, inventories and overall costs [3] [4]. Production planning decisions set the target and provide input to the scheduling phase that should guarantee the timely execution of the activities on the assigned production resources in the short term [5] [6].

In production, differences between projected future and actual execution are rather the rule than the exception due to the occurrence of unexpected events. Uncertainty may stem from a number of possible sources let they be either internal or external. Activities may last more or less than originally estimated, resources may be unavailable, materials may arrive behind schedule, ready times and due dates may change, new activities like rework could be inserted in the plan. A disrupted production plan or schedule incurs high costs due to missed due dates and deadlines, resource idleness, or higher work-in-process inventory.

At the level of production planning, aggregation handles this problem in such a way that no detailed plans are generated for a future that will almost surely differ from the actual expectations [5]. In the everyday management practice, these issues are addressed by rolling horizon planning methods [6]. That is, the plan (or the schedule) is recalculated on a regular basis, or when significant unexpected events occur. Unfortunately, such approaches have the drawback of increasing system nervousness due to frequent re-planning. Academic research has recently suggested robust planning to make production plans insensitive – at least to some degree – to disruptions. Robust approaches aim at reacting to the occurrence of uncertain events (reactive approaches) or at protecting the performance of the plan by anticipating to a certain degree the occurrence of uncertain events (proactive approaches) [7] [8].

However, the vast majority of robust approaches rely on the minimization of the expected value of a performance indicator (e.g., the expected tardiness). These objective functions, even if providing a significant improvement with respect to pure deterministic approaches, do not entirely model the concept of robustness. In fact, minimizing the expected tardiness aims at assuring an average good performance in terms of due date meeting, but does not protect against the worst cases if their probability is low.

Protection against worst cases is though a basic attitude in management decisions. Plant managers facing uncertainty always try to maximize some profit measure, but, at the same time, want to avoid the rare occurrence of very unfavorable situations causing heavy losses. Hence, a balanced compromise has to be found between expected profit and acceptable risk.

Financial research had already paid specific attention to the definition of risk measures to cope with uncertainty [9]. Risk is an essential concept to take into consideration the influence of extreme unfavorable events and, hence, to evaluate the robustness of a production plan or schedule.

In the paper we suggest a novel approach that tries to handle uncertainties closer to the root of their sources, at the level of production scheduling. The key ideas are (1) to keep the complexity of the PPS process at bay and (2) to stop the propagation of local disruptions to other resources, let they be in- or outside of the boundaries of the enterprise. In particular, within the general hierarchical PPS framework, we present a single resource stochastic scheduling model that captures release and processing time uncertainties of a set of jobs to be processed and aims at minimizing a risk measure of a scheduling performance indicator. The paper is organized as follows: in Section 2 the general planning and scheduling framework is outlined while Section 3 describes the single resource risk based robust scheduling problem. In Section 4 a branch and bound approach is proposed to solve a single resource scheduling problem aimed at minimizing a risk function of the maximum tardiness. Finally, Section 5 reports on test results in a real industrial environment producing hard metal tools.

2. The general planning and scheduling framework

A production plan defines how and when to execute production activities. It usually relates to aggregate information on groups of resources and activities. A robust production planning approach guarantees the performance of the plan, like timely delivery to the customers or adequate resource utilization, even in face of events not known at planning time. Typically, to keep the performance, the plan is to be modified, for instance by shifting the execution of some activities or by shrinking lead times. Hence, the term robustness refers to 'quality robustness', i.e., the insensitivity of the plan in terms of target performance. Hence, the robustness of the production plan relies on the possibility and capability of modifying the plan with little or no penalty with regard to the value of the objective function.

At the scheduling level, all the details have to be considered: individual activities, precedence relations among them, individual resources and their capabilities. Constraints are provided by the production planning phase in terms of release times as well as due dates to complete production activities and to deliver products to the following production phases. At this level, uncertainty directly stems from machine breakdown, manually executed tasks, rework activities, late supplying of raw materials or parts. A robust scheduling approach should be able to cope with such uncertainties to guarantee the respect of the due dates imposed by the production plan or, in the worst cases, minimize their violation. Indeed, being able to block the propagation of local disruptions to the whole production system is a key factor for robust scheduling.

3. Stochastic scheduling approach

The proposed approach considers the scheduling of a single resource in a production environment. This resource can model a single machine, a group of resources, or a whole department. Although it could seem a restrictive hypothesis, a single resource model is applicable to several cases where a group of resources can only work on a single product or a single type of products at a time. This assumption fits the case of multi-model transfer lines producing batches of different products but requiring intermediate setups, or make-to-order (MTO)/engineer-to-order (ETO) systems producing complex items where a whole department works on a single job at a time (e.g., instrumental goods or navy sector). In the framework of hierarchical production control, these production facilities receive a production plan to follow. Basically this plan provides the release times of jobs (when they are available from the previous production phases) and specifies the times jobs must be completed to move forward to the following production phases (local due dates). To respect the production plan, activities should be completed on time and, hence, the maximum lateness performance is considered. From a formal point of view [8], it can be classified as a stochastic single resource scheduling problem, i.e., determining the sequence of activities on the resource given the following elements:

A: set of activities to schedule;
p_i: processing time of activity i (stochastic);
r_i: release time of activity i (stochastic);
d_i: due date of activity i (deterministic).

Stochastic release times can model the occurrence of a delay either in the case of externally supplied raw materials or components, or in the case of the connection with a different department in the same production system. Deterministic local due dates model the constraints imposed by the production plan.

However, due to the presence of stochastic variables, meeting local due dates cannot be assured. Hence, to assure the maximum effort in respecting the due dates, the objective function is a risk measure for the maximum tardiness (or, in a more general way, for the maximum weighted tardiness). This objective function is likely to minimize a stochastic function of the maximum magnitude of the deviations with respect to the due dates, thus protecting the plan from worst cases.

The vast majority of approaches in the scheduling area dealing with uncertainty usually handles expected values of performance measures without considering risk. Translating the risk concept to the maximum tardiness means, that given a certain accepted risk level α, we estimate the quantile of the distribution at α or, alternatively, the expected value of the maximum tardiness in that tail of the distribution.

These concepts are mapped to the Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) respectively, two risk measures extensively used in the portfolio management research areas [9] (see also Figure 1). However, considering more complex objective functions involves the difficulty of calculating the distribution of the objective function.

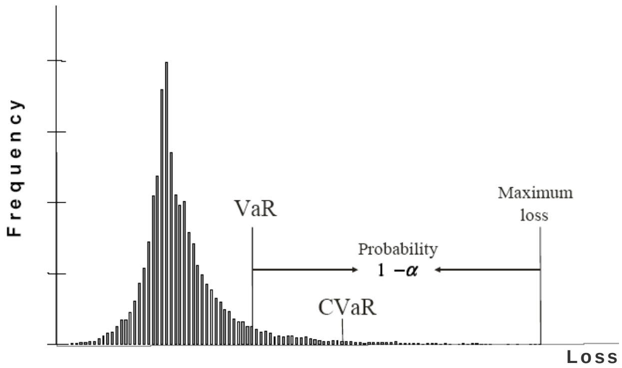

Figure 1. Value-at-Risk α (VaR(α)) and Conditional Value-at-Risk α (CVaR(α)) for a general loss distribution.

4. The proposed solution approach

As described in the previous section, the calculation of the distribution of a scheduling objective function is usually a challenging task. However, when a single resource problem is considered, this difficulty is dramatically reduced. Given the cumulative distribution function (cdf) of the processing time p_i of an activity i, $F_i(t) = \text{Prob}(p_i \leq t)$, if two activities i and j are in series, the cfd of the sum of their durations is the convolution of the individual cdf's [10]:

$$F_{i+j}(t) = F_i(t) * F_j(t) = \int_0^t F_i(t-s)dF_j(s)$$

If a stochastic release date is provided, it can be considered as an additional activity k with processing time r_j to be executed before j and activity j can be started only after either activity k or i have been completed and, hence, the cdf for the start time (S_j) and completion time (C_j) of activity j, given that j is executed after i can be defined as:

$$F_{S_j}(t) = F_i(t) \cdot F_k(t)$$

$$F_{C_j}(t) = F_{S_j}(t) * F_j(t)$$

Given d_j, the due date for activity j, the cdf of its tardiness T_j is:

$$F_{T_j}(t) = F_{C_j}(t + d_j)$$

In this way, given a single resource schedule, the cdf of the tardiness for all the scheduled activities can be computed. The cumulative distribution function of the maximum tardiness can be computed as the maximum of such cdfs:

$$F_{T_{Max}}(t) = \prod_j F_{T_j}(t)$$

Given the cdf of the maximum tardiness of the schedule, any risk measure can be easily calculated.

The solution approach grounds on a branch and bound framework. The branching scheme sequentially selects the next activity to schedule. In the leaves of the tree a complete schedule is given while the nodes represent partial schedules where only a subset of activities (A^S) has already been scheduled. For these activities, the maximum tardiness is already defined. For the not yet scheduled activities ($A \backslash A^S$), a lower bound on the maximum tardiness is provided assuming each of them starts immediately after the already scheduled activities or, if more constraining, after its release time (Figure 2, red block).

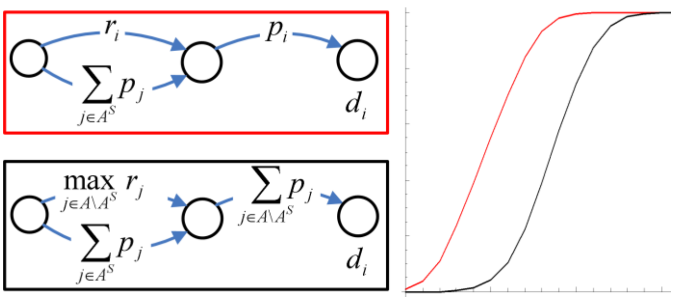

Figure 2. Activity networks for lower bound (red box) and upper bound cdf (black box) of the completion time for a not yet scheduled activity. Comparison of the two cumulative distribution functions (right).

An upper bound is obtained assuming that each of the not scheduled activities is executed as the last one in the sequence (Figure 2, black block, below). At a leaf of the tree, the upper and lower bound cdfs collapse in a single curve lying in the area between the bounding cdfs of its ancestors (Figure 2, right).

Given this property, it is easy to define a comparison criterion among the nodes. At each node, an upper and a lower cumulative distribution function is calculated. Given a stochastic performance measure, these cdfs define an upper and a lower bound on it. An example is provided in Figure 3: given a risk level α, the lower bound cdf (red) provides a lower bound on VaR(α) while the upper bound cdf (black) provides an upper bound on it. The lower bound cdf (red) stochastically dominates the cdfs of all the leaves having root in the node, as the upper bound cdf (black) is stochastically dominated by them. Hence, they provide valuable bounds for a large set of risk measures.

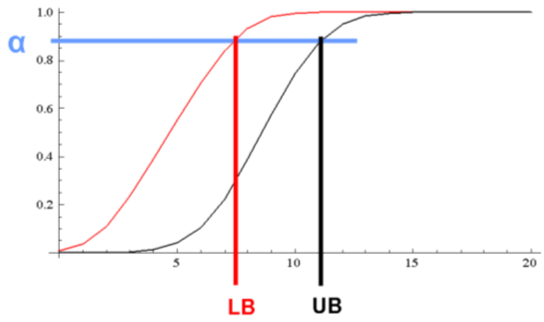

Figure 3. Lower bound (LB) and upper bound (UB) for the value at risk α in a node of the search tree.

In the exploration of the search tree, if the LB of a node A is higher than the UB of another node B, then node A can be pruned since all the deriving solutions will have a VaR greater than any solution deriving from node B. Similarly, if the LB of a node is higher than the incumbent solution, then the node can be pruned since all the deriving solution will have a VaR greater than the value of the incumbent solution.

5. Real case study

The proposed scheduling approach has been tested on a real manufacturing environment producing sintered carbide tools, from drawing tools for tubes bars and wires production to tools for cold forming and in particular for the fastener and nail making industry. A significant percentage of the production is devoted to tailor-made tools satisfying the specific need of the customers. Due to the characteristics of the products, the production system works on a make-to-order basis with really small lots ranging from a single piece to some dozens.

In particular, carbide drawing tools have been considered in the test (Figure 4).

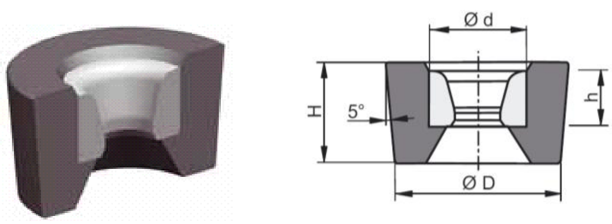

Figure 4. A carbide drawing tool.

Drawing carbide tools are obtained through a sintering process. Hard metal powders are injected into a mold or passed through a die to produce a weakly cohesive block. The block is machined to obtain a shape roughly similar to the final product and processed in a sintering furnace. After the sintering phase, the parts must be grinded to obtain the exact tool geometry and then assembled in a steel bull block. The most critical phase of the manufacturing process is grinding. Hence, we concentrate the analysis on the CNC grinding department and, in particular, on a single CNC grinding machine dedicated to the production of a specific class of tools (characterized by certain dimensions and geometric characteristics of the die). The lots to be grinded arrive from the previous production phases. Due to possible uncertainties affecting the supply of raw materials and blanks, together with possible issues affecting the other sintering phases, the release times cannot be considered deterministic. To process the lots in the CNC grinding department, a setup of the grinding machine is needed. Then the CNC grinding machine is programmed and a first piece is produced and verified. If the part does not pass, the CNC program is adjusted and the first piece is reworked until the specifications are satisfied. Then the remaining part of the lot is produced. Due to this, the time needed to machine the whole lot is a stochastic variable. As the whole lot is machined, it is transferred to the following production phase according to the internal due date defined in the production plan. Delays on these due dates likely propagate through the whole system causing a delay on the delivery date to the customer. To assure adequate satisfaction of the customer, the management considers the maximum tardiness as a proper indicator.

So as to define the test set, six months of production records have been analyzed to determine the distributions for the job arrival process, standard job processing times and due dates. The previously defined distributions have been first used to sample due date values and then the average values for release times and processing times. Then a uniform distribution and a triangular distribution have been defined to model uncertainty affecting the release dates and the processing times respectively, given the sampled average values. The parameters characterizing the distributions have been given on the base of historical data, considering the characteristics of the jobs (product types, number of pieces).

The scheduling approach has been used to schedule 10 instances consisting of 10 jobs covering a horizon of about 2 weeks aimed at minimizing the Value-at-Risk of the maximum lateness with different risk levels (1%, 5%, 10%). The solution algorithm has been coded in C++ and executed on an Intel T5600@1.83GHz CPU.

In Table 1 the results of the experiments are reported in terms of Value-at-Risk, solution time and percentage of visited nodes respect to the size of the whole tree. In addition, the proposed approach has been compared against the VaR of the maximum lateness obtained using the Earliest Due Date rule (EDD); differences in percentages between the two performance values are also reported. For each indicator, the minimum, maximum, mean and standard deviation values are provided. The results show that, as the risk level decreases, the solution time increases. This is due to the fact that, for low risk levels laying in the tail of the distributions, the cdfs of the different schedules converge asymptotically to value one, hence, the difference among them is small and the bounding and pruning criteria are less effective. The average fraction of nodes visited to find the best solution is constant (about 0.8%) for all the cases even if, as the risk level decreases, the maximum value is higher. The value of the solution remains however the same, and, in fact, using the EDD rule entails an increase for VaR of 3.47% on average.

The proposed approach demonstrated overall good time performance but, given the complexity of the problem, the computational time is likely to explode as the number of activities to schedule increases. However, it must also be noticed that, as the number of activities increases, the distribution of the sum of the processing times converges to the normal distribution. Hence, the benefits of the proposed approach are stronger when a small number of activities (less than 20) are considered.

Table 1 Results on the industrial test case

		Min	Max	Mean	StDev
Risk=1%	VaR (hours)	44.00	112.00	66.10	18.60
	Sol. Time (sec)	6.20	843.2	138.90	255.70
	% visited nodes	0.04	3.04	0.82	1.01
	% diff. vs. EDD	0.00	10.71	3.54	3.45
Risk=5%	VaR (hours)	42.00	109.00	64.90	18.17
	Sol. Time (sec)	6.00	625.00	125.30	190.50
	% visited nodes	0.04	2.59	0.86	9.51
	% diff. vs. EDD	0.00	11.01	3.45	3.81
Risk=10%	VaR (hours)	41.00	108.00	64.20	18.06
	Sol. Time (sec)	5.70	508.90	110.80	154.40
	% visited nodes	0.04	2.33	0.82	8.36
	% diff. vs. EDD	0.00	11.11	3.41	3.80

Future research will address the development of specific branching schemes and bounding criteria so as to make the approach applicable to more complex scheduling problems.

Acknowledgements

This research has been developed in the framework of the 'XVII Executive Programme of Scientific and Technological Cooperation between the Republic of Hungary and the Republic of Italy' and supported by the OTKA T-73376 and OMFB-01638/2009 grants.

References

[1] Fleischmann, B., Meyr, H., 2003, Planning Hierarchy, Modeling and Advanced Planning Systems, in: de Kok, A.G., Graves, S.C. (eds.), *Supply Chain Management: Design, Coordination and Cooperation*, Handbooks in OR and MS, 11:457–523, Elsevier.

[2] Pochet, Y., Wolsey, L.A., 2006, *Production Planning by Mixed Integer Programming*, Springer.

[3] Schuh, G., Gottschalk, S., Höhne, T., 2007, High Resolution Production Management, *Annals of the CIRP* 56/1: 439-442.

[4] Monostori L; Erdős G., Kádár, B.;Kis, T., Kovács, A., Pfeiffer, A.; Váncza, J., 2010, Digital Enterprise Solution for Integrated Production Planning and Control, *Computers in Industry* 61:112–126.

[5] Váncza, J., Kis, T., Kovács, A., 2004, Aggregation - the Key to Integrating Production Planning and Scheduling, *Annals of the CIRP* 53/1:377-380.

[6] Alfieri, A., Tolio, T., Urgo, M., 2011, A Project Scheduling Approach to Production Planning with Feeding Precedence Relations, *International Journal of Production Research* 49(4):995–1020.

[7] Tolio, T., Urgo, M., 2007, A Rolling Horizon Approach to Plan Outsourcing in Manufacturing-to-Order Environments Affected by Uncertainty, *Annals of the CIRP* 56/1:487-490.

[8] Sotskov, Y. N., Sotskova, N. Y., Lai, T.-C., Werner, F., 2010, *Scheduling under Uncertainty: Theory and Algorithms*, Belarusian Science, Minsk.

[9] Szego, G., 2002. Measures of Risk. *Journal of Banking & Finance* 26(7):1253-1272.

[10] Agrawal, M. K., Elmaghraby, S. E., 2001, On Computing the Distribution, Function of the Sum of Independent Random Variables, *Computers & Operations Research* 28(5):473-483.