Planning complex engineer-to-order products

Book Section

Recent Advances in Integrated Design and Manufacturing in Mechanical Engineering. Ed. G. Gogu, D. Coutellier, P. Chedmail, P. Ray, Kluwer, Dordrecht, March 2003, pp463-472,ISBN 1-4020-1163-6,

PLANNING COMPLEX ENGINEER-TO-ORDER PRODUCTS

Christopher Earl: Dong Ping Song: Christian Hicks
Open University, Imperial College London, Newcastle University, UK

Abstract: The design and manufacture of complex Engineer-to-Order products is characterised by uncertain operation durations, finite capacity resources and multilevel product structures. Two scheduling methods are presented to minimise expected costs for multiple products across multiple finite capacity resources. The first sub-optimises the operations sequence, using mean operation durations, then refines the schedule by perturbation. The second method generates a schedule of start times directly by random search with an embedded simulation of candidate schedules for evaluation. The methods are compared for industrial examples.

Key words: engineer-to-order, schedule, stochastic, resource constraints

1. INTRODUCTION

Planning Engineer-to-Order products [1,2,3,4] takes into account several factors: significant uncertainty in operation times; limited resources maintained by companies in response to fluctuations in demand; concurrent development of multiple products and complex product (Figure 1). The problem addressed here is to create a plan with minimum expected costs (both earliness and lateness). The majority of planning research in this area is limited to simple systems such as single machines [5], serial structures or flow lines [6,7,8,9], two stage assembly systems and two stage distribution type networks [10,11]. Research on planning multilevel products has optimised operation sequences for deterministic operation durations. However, with stochastic durations, the other factors, namely; finite capacity resource constraints, precedence constraints, assembly co-ordination and due date constraints; all cause the actual schedule to deviate significantly from plan. A further difficulty is that although the schedule starts as a useful guide, it can rapidly become a constraint as uncertainties take effect, operations are completed and new conditions for the remaining operations come into play. Two methods are developed to optimise schedules for multiple complex products over multiple finite resources. The first optimises sequences (using mean operation duration estimates) then further optimises timings for this sequence. The second optimises timings directly. The methods are tested on examples of ETO products in the power generation industry.

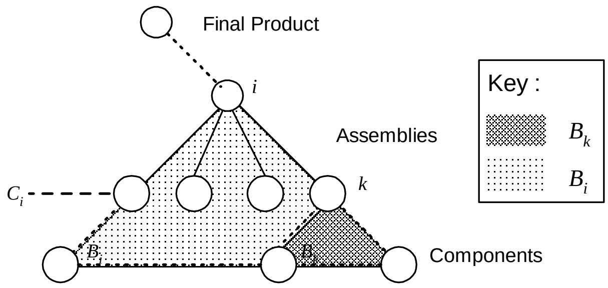

Figure 1. General product structure whose nodes represent manufacture/assembly (with component design at leaves and final assembly at the root)

2. PLANNING

Planning in stochastic systems needs to take account of how plans are implemented. For example, when a resource finishes an operation but the next planned operation has not arrived (but several other operations are queuing at this resource); which operation should the resource do next? Two choices present themselves. The first keeps the original sequence [12]. The resource ignores the queuing operations, keeps idle and waits for the next planned operation. The second choice selects one of the queuing operations by a priority rule. These choices give rise to the two methods developed here for planning. Broadly the first optimises sequences then refines the associated timings. The second optimises timings directly.

2.1 Notation

An implementation is a sample process (or real execution) of a schedule of start times $\boldsymbol{s} = \{s_i\}$ of operations i, with durations x_i sampled from their probability distributions (or given by their real values). An implementation is described by the 'actual' operation start times a_i and completion times c_i of operations on many products and resources. The following notation is used:

Γ set of all operations over all products and all resources,
L set of the final assembly operations over all products,
C_i operations which immediately precede operation i (figure 1),
$\rho(i)$ operation immediately after i in the product structure,
$r(i)$ resource used to undertake operation i
$\varphi(i)$ operation immediately preceding operation i on resource $r(i)$;
d_i due date of operation i.
h_i, h_i^- costs of earliness and lateness for operation i.

The basic relations among starting and completion times are:

$$a_i = \max\left(s_i, c_{\varphi(i)}, \{c_j : j \in C_i\}\right) \qquad c_i = a_i + x_i \tag{1}$$

The cost function which is minimised is the expected value of:

$$\sum_{i \in \Gamma\setminus L} h_i(a_{\rho(i)} - c_i) + \sum_{i \in L} h_i \max(d_i - c_i, 0) + \sum_{i \in L} h_i^- \max(c_i - d_i, 0) \tag{2}$$

2.2 Industrial ETO examples

The planning methods were tested on examples from an Engineer-to-Order company which designs and manufactures power generation equipment. Although the examples are not complete products they are significant functional subassemblies such as bearing pedestals, casings and rotor/blade assemblies. The data available covered both product structures and estimates of operation times for manufacturing processes and assembly. As mentioned above these times are not known with certainty. But the difficulties are more profound. Because the products are engineer-to-order, operations are not repeated across products and so estimations of stochastic characteristics and distributions are problematic due to small sample sizes. However, from historical data on sets of similar operations, means and variances can be approximated but estimates of distributions are more difficult to obtain. In this paper we assume that the distributions are normal.

3. TWO-PHASE OPTIMISATION METHOD

The first phase (figure 2) sets the values of operation durations at the estimated means. A sequence of operations on each resource is then determined by either: (i) a finite loading heuristic with a priority rule or, (ii) random search using Simulated Annealing (SA) or Evolution Strategy (ES). The second phase takes operation sequences from the first phase as constraints and refines operation timings using a Perturbation Analysis Stochastic Approximation (PASA). Although SA and ES can also be used to refine the timings in the second phase, PASA was found to converge faster.

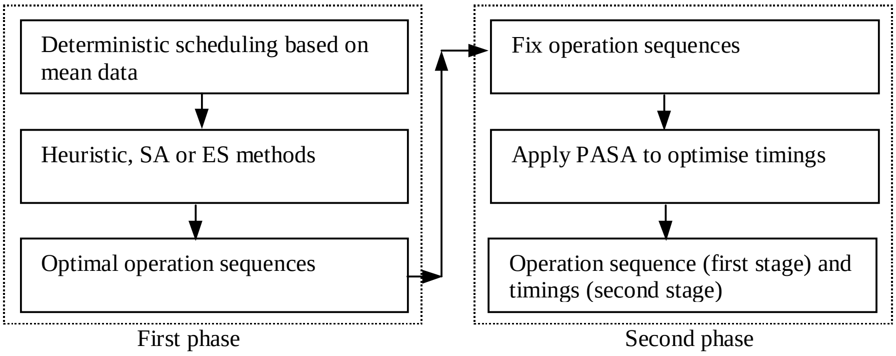

Figure 2. Two-phase scheduling

3.1 First phase heuristics

First phase heuristics are based on backwards scheduling which aims to start each item as late as possible. If more than one operation is waiting to be loaded onto a resource then a latest possible completion time (LCT) priority rule is used to decide which is loaded first. Note that if an operation is loaded first in the backwards scheduling then it will be processed last when the plan is implemented. The LCT priority rule tries to reduce waiting and thus holding costs. In cases where LCT cannot decide, the highest cost ratio h_i/x_i (HCR) selects the next operation to be loaded.

Backwards finite loading with the LCT priority rule has computational complexity less than $O(n^2)$ in the number of operations n. Several examples of ETO products, (major subassemblies), from the power generation sector were tested. One with 113 operations and 13 resources took 2 seconds and another with 239 operations on three concurrent products and 17 resources took 4 seconds. See table 1 and figure 4 for details of these products. Finite loading gives much lower costs (up to a factor of five) than infinite loading.

3.2 First phase random search

Random search methods are used to optimise start times of operations under the assumption that durations are set at their estimated means. These start times give a sequence which is then adopted in the second phase refinement of the schedule. Two different random search techniques have been applied. Simulated Annealing (SA) [13,14] numerically optimises [15] operation start times $\{s_i\}$. The performance (total cost) of a candidate schedule within a SA iteration is determined by simulating its implementation using the priority rule that the operation with the earliest planned start time (EPST) is started first if there is competition for resources. Two types of constraint are applied. Physical constraints specify that a starting event on an operation cannot occur before all immediately preceding operations C_i are completed. Planning constraints specify that an operation cannot start before its planned start time ($a_i \geq s_i$). The actual schedule $\{a_i\}$ given by the simulation is likely to deviate significantly from $\{s_i\}$ because of the finite capacity resources. Neighbouring schedules are obtained by randomly choosing new start times from a uniform distribution in the ranges $[s_i - 1/2\gamma_i, s_i + 1/2\gamma_i]$ where γ_i is the current step size. Step sizes and annealing temperatures are reduced linearly.

Evolution Strategy (ES) has iterative procedures with “selection”, “crossover” and “mutation” for generating offspring from a parent population. Offspring are selected for further generation. ES uses continuous variables and is thus suitable for numerical optimisation [16] of schedules. The chromosome of an individual schedule is represented by the vector $\mathbf{s} = \{s_i\}$ of start times. Evaluation of offspring is by simulation as in SA.

Results from running SA and ES on industrial data for a single multilevel subassembly with 113 operations and 13 separate resources indicate that SA converges faster than ES, although in the long term ES gives lower costs. See table 1 and figure 4(a) for details of this product and associated operations. Compared with the heuristic method ES and SA achieve marginally lower costs, in the region of 10%, but higher computational costs (by a factor of 100). Similar results were obtained from experiments on similar subassemblies from the power generation sector.

3.3 Second phase perturbation analysis

Perturbation analysis (PA) is used in the second phase. Consider a sample implementation (or nominal path NP) of a schedule with start times $\mathbf{s} = \{s_i\}$ and parameters (eg durations) ω, in the whole sample space Ω, determining start times $\{a_i\}$ and completion times $\{c_i\}$ by equations (1). The cost of the sample process $V(\mathbf{s}, \omega)$ is given by (2) [17,7]. Let planned start time s_j be perturbed to $s_j + \Delta$. The sample process for $\{s_j + \Delta, s_i, i \neq j, i \in \Gamma\}$ with the same ω is a perturbed path (PP). The perturbed path with start and completion times $\{a_i'\}$ and $\{c_i'\}$ can be constructed directly from the nominal path without repeating the simulation.

Perturbation generation is described by (i) if $a_j > s_j$, then $a_i' \equiv a_i$ and $c_i' \equiv c_i$, for $i \in \Gamma$ (ii) if $a_j = s_j$, then $a_j' = a_j + \Delta$ and $c_j' = c_j + \Delta$. Perturbation propagation and disappearance are described by (iii) if $a_i = s_i$ $(i \neq j)$, then $a_i' = a_i$ and $c_i' = c_i$ (iv) if $a_i = c_{\varphi(i)}$ $(i \neq j)$, then $a_i' = a_i + (c_{\varphi(i)}' - c_{\varphi(i)})$ and $c_i' = c_i + (c_{\varphi(i)}' - c_{\varphi(i)})$, (v) if $a_i = c_k$ $(i \neq j, k \in C_i)$, then $a_i' = a_i + (c_k' - c_k)$ and $c_i' = c_i + (c_k' - c_k)$. The whole perturbation gain Δ will be propagated along the perturbed path. Define $I(i) := 1\{a_i' \neq a_i\}$, where $1\{.\}$ takes 1 if $\{.\}$ is true, and 0 otherwise. Note that a_i' and c_i' have the same perturbation gain. The sequence $\{I(i), i \in \Gamma\}$ determines the difference between PP and NP. A recursive procedure implements the perturbation propagation rules and determines $\{I(i),\ i\in\Gamma\}$. Song et al [12] show that for any $\omega\in\Omega$ and $j\in\Gamma$

$\partial V(\mathbf{s},\ \omega)/\partial s_j = \sum_{i\in\Gamma\setminus L} h_i\cdot(I(\rho(i))\text{-}I(i))+\sum_{i\in L} h_i\cdot I(i)\cdot 1\{d_i>c_i\}+\sum_{i\in L} h_i^-\cdot I(i)\cdot 1\{d_i\leq c_i\}$

is an unbiased estimator of gradient, that is:

$\partial\{E\ V(\mathbf{s},\ \omega)\ \}/\partial s_j = E\{\partial V(\mathbf{s},\ \omega)/\partial s_j\}$and $E|\partial V(\mathbf{s},\ \omega)/\partial s_j| < \infty$, for any $j\in\Gamma$. Thus a PA-based Stochastic Approximation [18,19] is $\mathbf{s}_{n+1} = \mathbf{s}_n - \gamma_n \cdot\nabla J_n$, where $\mathbf{s}_n$ are start times at the beginning of iteration n, ∇J_n is the gradient estimator, and step $\gamma_n > 0$, $\gamma_n\to 0$, $\Sigma\ \{\gamma_n\}$diverges and $\Sigma\ \{\gamma_n\}^2$ converges (eg $\gamma_n = 1/n$). The gradient estimator is calculated using K sample processes.

Some results from applying PASA (in conjunction with first phase methods and using K = 100 sample processes) to complex products (figure 4) are shown in figure 5. They indicate that application of the second phase PASA yields an extra reduction in costs of between 10-20% with moderate computation.

4. ONE-PHASE SCHEDULING

Simulated Annealing (SA) and Evolution Strategy (ES) methods can be extended from their application in the two-phase process to provide one-phase optimisation methods (figure 3). Multiple sample processes are required to estimate expected cost.

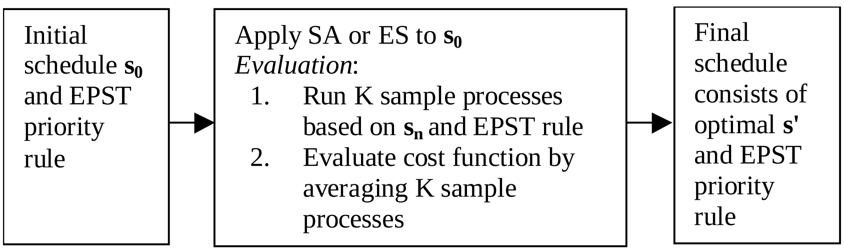

Figure 3. One phase method for scheduling

5. NUMERICAL EXAMPLES

Two examples are described here (Table 1). Example A consists of a single product and example B has three products designed and manufactured concurrently. The products are major subassemblies of power generation plant. One simplification is introduced, namely that the concurrent products are all assumed to start at the same time, with the same delivery date. In practice, for major subassemblies on a single product this is a reasonable assumption but for concurrent products, staged start and delivery will be usual practice. Example B has more pressure on resources - a common feature of ETO where resource constraints become a significant problem as the number of concurrent products increases.

Table 1. Examples of ETO used for scheduling

Example	Products	Components	Operations Manufacture/Assembly	Resources
A	1	9	100/13	13
B	3	47	210/29	17

5.1 Single product

Consider a single product (figure 4a) which is a major subassembly of power generation plant. The numbers on the nodes are references to particular operations, components and products. Normally distributed operation times have increasing variance as assembly progresses. Holding costs are set at 1% x {sum of operation times (in days) already spent on the item}x £1000. Lateness costs for final product were set at twice holding costs of the final product. The due date for this single subassembly was 180 days. Parameters for the search algorithms were established by experiment over several simulations of a range of examples, including the specific ones described here. For each method K=100 sample processes are used to assess stochastic effects. The three two-phase methods are compared in figure 5a, where the second phase PASA is compared for the various inputs provided by the different first phase methods. For single products an extra cost reduction of about 20% is achieved by applying the second phase PASA.

5.2 Multiple products

Consider three different products with the multi-level assembly structures shown in figure 4(a), (b) and (c). The aim is to plan concurrent development of these three products. The same regime of holding and lateness costs as for the single product was assumed but with a due date of 900 days. Normal distributions were assumed for all operation times with variances increasing for assembly operations closer to the finished product. In all cases K=100 samples were used. The three two-phase methods are compared in figure 5(b). The second phase PASA is compared for the various inputs provided by the different first-phase methods. For multiple products an extra 10% cost reduction is achieved by applying the second phase PASA.

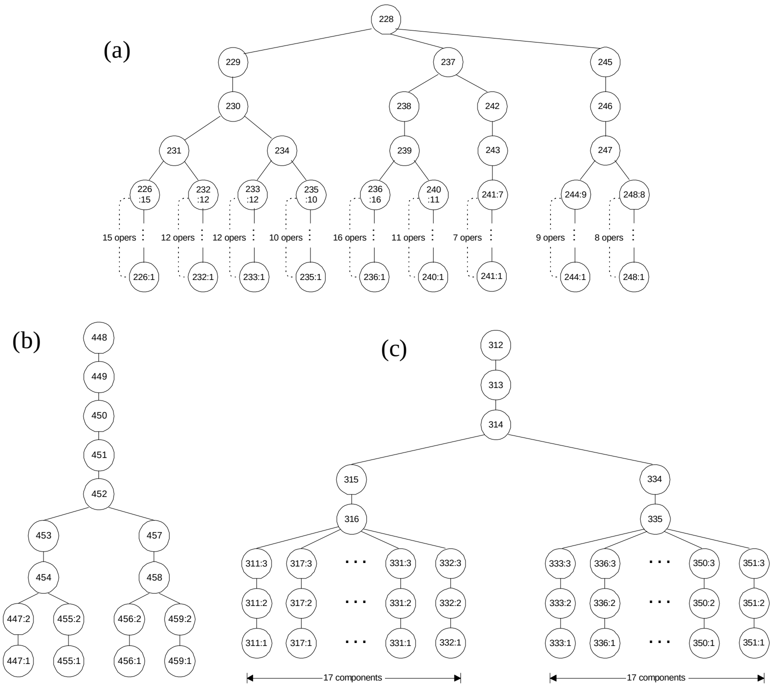

Figure 4. Examples of multilevel product structures

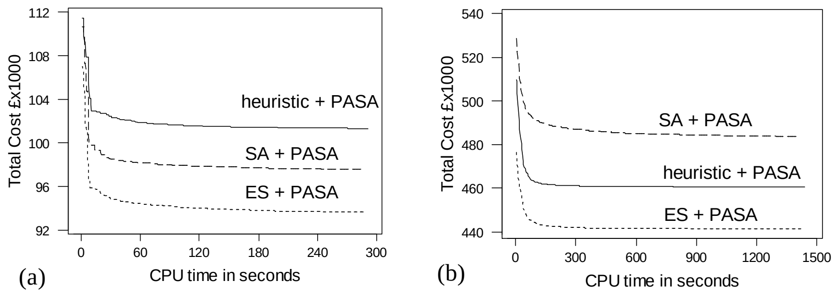

Figure 5. (a) Total costs versus CPU times at the second phase for (a) assembly of product in figure 4(a) and (b) concurrent assembly of three products in figure 4(a), (b) and (c)

For one phase methods the initial schedule $\mathbf{s}_0$ is obtained by a shifted backwards scheduling assuming infinite capacity and mean duration times. The shift, corresponding to letting the mean product completion time meet the product due date, helps to reduce SA and ES search time. The choice of parameters for SA and ES is made by repeated experiments on these and other similar complex examples. One-phase methods are applied to the single and multiple product cases (figure 4) to find optimal operation timings with the EPST priority rule (figure 6).

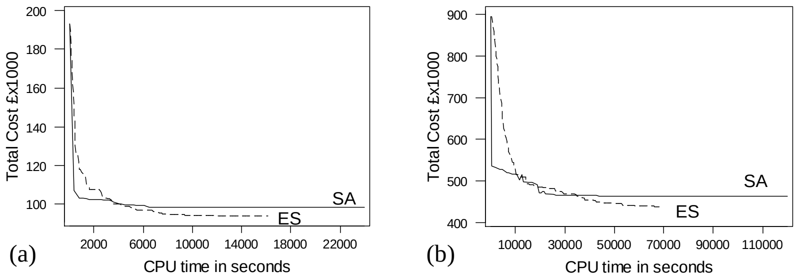

Figure 6. Total costs versus CPU time by one-phase methods for (a) assembly of product in figure 4(a) and (b) concurrent assembly of three products in figure 4(a), (b) and (c)

6. CONCLUSION

Planning large and complex products is a difficult problem especially when the operations have large uncertainties in duration. This is likely to happen for engineer-to-order products which are customised to a particular client's specification. Furthermore, companies which design and manufacture these types of product often have an unpredictable level of orders. Thus they retain core resources, but when several products are undertaken concurrently, significant competition for the resources occurs. The aim of an optimum plan is to minimise expected cost, which includes lateness and holding (i.e. work in progress or earliness) costs.

The methods developed in this paper go some way to solving this problem. One method uses a two-phase process. The first phase optimises the sequence of operations and the operation the start times (assuming means of operation times are set deterministically at estimates of their means). An heuristic is described which is computationally two orders of magnitude quicker than alternative random search simulated annealing and evolutionary strategy methods, with only a marginal reduction in cost of schedules for the industrial examples tested. The second phase uses the sequence of operations from the first phase and optimises timings for this sequence using perturbation analysis giving between 10-20% cost reduction. The other method optimises timings directly through random search but due to the stochastic operation times a priority rule is required both to implement the final plan as well as evaluate candidate solutions. Repeated simulation of sample processes for this evaluation makes the one-phase method computationally expensive. Both the one- and two-phase methods were implemented and validated on several cases with industrial data. Further research on re-planning when new orders arrive is currently being undertaken by the authors.

7. REFERENCES

[1] Hendry L C, Kingsman B G 1989 "Production planning systems and their applicability to make-to-order companies, *European Journal of Operational Research*, 40, 1-15.

[2] Bertrand J W M, Muntslag D R 1993 "Production control in engineer to order firms" *International Journal of Production Economics*, 30, 3-22.

[3] Wortmann J C, Muntslag D R, Timmermans P J M 1997 *Customer-driven manufacturing* Chapman & Hall.

[4] Giebels M 2000 *EtoPlan: a Concept for Concurrent Manufacturing Planning and Control*, Ipskamp BV

[5] Liao C 1992 "Optimal release time on a stochastic single-machine" *International Journal of Systems Science*, 23 1693-1701.

[6] Yano C A 1987 "Stochastic lead-times in two-level assembly systems" *IIE Transactions* 19(4), 371-378.

[7] Gong L G, Kok T, Ding J 1994 "Optimal leadtimes planning in a serial production system" *Management Science*, 40(5), 629-632.

[8] Homem de mello T, Shapiro A, Spearman M L 1999 "Finding optimal material release times using simulation-based optimisation" *Management Science*, 45(1), 86-102.

[9] Hasan C N, Spearman M L 1999 "Optimal material release times in stochastic production environments" *International Journal of Production Research*, 37(6), 1201-1216

[10] Yano C A 1987 "Setting planned lead-times in serial production systems with tardiness costs" *Management Science*, 33(1), 95-106.

[11] Yano C A 1987 "Stochastic lead-times in two-level distribution type networks" *Naval Research Logistics*, 34, 831-843.

[12] Song D P, Hicks C, Earl C F 2001 "Setting planned job release times in stochastic assembly systems with resource constraints" *International Journal of Production Research*, 39(6), 1289-1301.

[13] Aarts E, Korst J 1989 *Simulated Annealing and Boltzmann Machines*, Wiley, NY

[14] Azencott R (Ed.) 1992 *Simulated Annealing Parallelization Techniques*, Wiley, NY

[15] Brandimarte P, Maiocco M 1999 "Job shop scheduling with non-regular objective: a comparison of neighbourhood structures based on a sequencing/timing decomposition" *International Journal of Production Research*, 37, 1697-1715

[16] Schwefel H P 1995 *Evolution and Optimum Seeking*, Wiley, NY

[17] Baker K R, Scudder G D 1990 "Sequencing with earliness and tardiness penalties: A review" *Operations Research*, 38(1), 22-36.

[18] Rubinstein R 1986 *Monte Carlo Optimisation, Simulation and Sensitivity of Queuing Networks*, Wiley, NY.

[19] Fu M, Hu J 1997 *Conditional Monte Carlo: Gradient Estimation and Optimisation Applications* Kluwer, Dordecht