4th International Industrial simulation Conference ISC'2006 June 5-7 Palermo Italy
Eurosis-ETI. ISBN 90-77381-26-0

EFFECTIVE JOB SHOP PRIORITY SCHEDULING

Hans P.M. Veeke, Gabriël Lodewijks
Delft University of Technology, Faculty 3mE
Mekelweg 2, 2628 CD Delft
H.P.M.Veeke@tudelft.nl

KEYWORDS
MRP planning, job scheduling,
Job shop simulation

ABSTRACT

Job shop planning and scheduling have been thoroughly investigated during the last decades. The main research question was to determine the optimal load and sequence of jobs with limited resources available. Job shop environments are very stochastic (resources break down; customer orders are added at the last moment etc). Because of these stochastics, the planning is usually based on statistical data from the past. On the contrary, scheduling has to sequence individual jobs, for which statistics can not be used. There is a natural friction between statistical decision making and individual job scheduling.
In this paper the effects of some basic scheduling priority rules are illustrated which lead to the conclusion that freedom of action (i.e. there is no detailed production sequence of tasks) is a promising alternative for effective job scheduling during manufacturing. It is shown that planning decisions can be based on sets of jobs, for which the statistics hold. This combination of planning and scheduling already results in controlled throughput times and gives a clear basis for decision support during manufacturing.

INTRODUCTION

Usually Enterprise Resource Planning (ERP) systems use basic priority rules for the delivery control in job shop environments (see Russell, Taylor, 2006). Applying these rules is the last step of an MRP-based approach, starting with global and detailed load balancing. The objective of job shop scheduling is assumed to be finding a trade off between loading efficiency and delivery accuracy (Kemppainen, 2005). The resulting production schedule is a plan for what part will be made at what time utilizing what resource (Parsons, Phelps, 2001). Planning and scheduling have become even more important due to the increased pressure on throughput times. Nowadays it is necessary to deliver in short time and to react in a flexible way to changing market demands.
In this paper an alternative way is presented to schedule customer orders in a job shop production. The primary goal of the research is to support decision making during the planning phase. Rather than shortening throughput times, the starting point for research was to achieve throughput time control.
Controlling throughput time is considered the first requirement in order to achieve delivery reliability. If the required level of reliability is reached, then the next step will be to shorten throughput times.
First, a number of aspects, which influence the results of priority rules for scheduling, will be investigated. The first experiments can be considered an extended verification of the simulation model. Based on the results, an alternative approach will be proposed, which shows promising results for achieving delivery reliability with real support for decision making, and sufficient flexibility to react on disturbances. This approach preserves high occupancy values.
The simulation model used, models an existing production company with a characteristic order flow.

THE SIMULATION MODEL

The simulation model describes a company that makes highly specialized optical instruments on customer order. Every order triggers 'Material Supply' to provide the required materials from stock or suppliers. 'Parts Manufacturing' then produces the required parts and 'Assembly' finally assembles these into the ordered products. 'Parts Manufacturing' is functionally organized (job shop). There are three departments: 'Milling', 'Fitting' and a 'NC-department'. Every part to be manufactured should be processed (eventually several times) by several departments. The parts are being processed in batches. The batch size depends on the customer order size. On average a product consists of 2 components.
The model has been programmed in TOMAS (Veeke, Ottjes, 2000), an object oriented simulation extension of Delphi.

PRIORITY RULES

The simulation model contains three default scheduling alternatives:

- First Come First Served (FCFS): selection sequence depends on the arrival sequence
- Shortest Processing Time first (SPT): the batch with the smallest processing time is selected. It is well known that SPT shortens the average throughput time, but increases the standard deviation. To compensate for this, the batches are divided into two classes: batches with a task time smaller or larger than some value P. Batches with a task time smaller than P have

priority to the other batches, but are mutually scheduled according the FCFS-rule. The larger batches are scheduled in sequence of increasing task time. By varying P it is possible to apply the SPT-rule to some extent. If P = 0, the usual SPT-rule is applied, and if P = ∞ the FCFS- rule applies.

- Operation Due Date (ODD): the batch with the shortest starting time is selected.

Although these rules have been investigated intensively, the effects depend on the task time distributions, which in this research are directly derived from reality. The results of the simulation runs are shown in table 1. These results include delays in material supply.
The high standard deviation causes a low planning reliability. The table shows clearly that the average throughput time decreases significantly by using the SPT-rule. In the table SPT 98% means that 98% of the batches has a task time below P. Still, the standard deviation remains large for each SPT-alternative and the batches with a large task time show a very slow progress (see column Max). However, the table also shows that the SPT-rule, where only 5% of the batches is selected according the SPT-rule (SPT 95%) already realizes 80% of the maximum gain in throughput time, which could be realized if all batches are selected according the SPT-rule.

	T	MDD	SDD	Min	Max
FCFS	171	12.5	9.7	-3.1	39.5
SPT 98%	116	11.1	10.0	-3.0	73.0
SPT 95%	109	11.0	10.0	-3.0	73.0
SPT 0%	91	10.5	10.0	3.1	72.5
ODD	173	11.5	7.9	2.9	27.2

Table 1. Effect of FCFS-, SPT- and ODD-rules.
T = throughput time (hours)
MDD/SDD: Mean/Standard Deviation of Delivery Week (weeks)

SUPPLY DELAYS

Two factors determine the actual delivery time:

a. the performance of material supply
b. the production process itself

Both factors are a source for a planning delay. The parts belonging to one single component to be assembled, should be delivered at the same time. So a component consists of a set of parts, each requiring different materials that are delivered by different suppliers. The last supplier determines the delay with respect to final delivery. In the model, on average 2 suppliers are involved for the parts manufacturing for one component. The supply patterns for all suppliers both at the part and the component level are shown in figure 1.

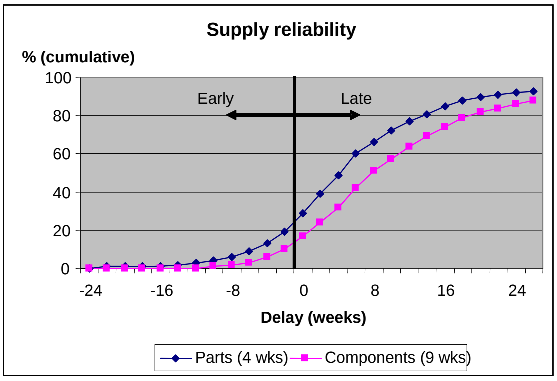

Figure 1. Delivery pattern Material Supply

On average the materials for each separate part are delivered 4 weeks late. Measuring the material delivery for each component, then the delivery is on average 9 weeks late. So the major part of the delay in component delivery is caused by part delivery. From now on this delay is removed from the model and the assumption is made that all materials are delivered in time.
The delivery to Assembly shows a similar effect. The realisation of the planned start time of Assembly depends on the last arrival of parts from Parts Manufacturing. If all parts separately are being delivered in time on average, then the set of parts for one component will be late on average. There are two ways to prevent this effect:

- each part should be planned "early"
- the standard deviation around the average should be reduced to (almost) zero.

The first solution would increase the stock of parts significantly and the second solution will be impossible for a stochastic job shop organisation. This research will focus on reduction of the deviation and investigate for this the influence of the order characteristics and the way of planning itself.

TASK TIMES

The variation in task times is large (see table 3). In reality they vary between 1 and more than 80 hours. It makes sense to investigate the effect of long task times, because these times particularly cause long waiting times (for this very reason the SPT-rule is often used).

	0-4	4-8	8-16	16-40	>40	M
Milling	22%	28%	27%	17%	6%	13.5
NC	7%	22%	33%	12%	26%	25.0
Fitting	51%	21%	14%	11%	3%	8.7

Table 3. Task times (hrs) per department
(M = Mean task time)

In the model, the longest task time is assumed to be 80 hours. In order to investigate the effect of extremely long task times, each task time larger than 40 hours is split into two tasks, one with a task time of 40 hours and one with the rest of the task time. The planning takes this into account by scheduling both tasks sequentially with intermediate waiting times of 40 hours (see table 4). Splitting the task times has no significant effect. Although the average task time decreases, the average number of tasks increases.

	T	MDD	SDD	Min	Max
Full Tasks	176	3.5	3.9	-2.1	20.9
Split Tasks	167	3.0	3.7	-2.5	23.4

Table 4. Effect of task times

PLANNING

Job shops try to achieve delivery reliability while maintaining a maximum occupancy (preferably 100%) of the available capacity. These goals are in essence conflicting. So usually one tries to achieve some optimal situation: acceptable reliability with acceptable occupancy. The occupancy of the 3 groups with normal task times is 80 to 90%, which completely consists of tasks that have been planned in detail. A number of experiments has been done to investigate the effect of adding tasks, which don't have a planned delivery time, but are merely used to occupy capacity up to 100%. The results are shown in the table below.

	T	MDD	SDD	Min	Max
P/U: 90 / 0%	176	3.5	3.9	-2.1	20.9
P/U: 70 /30%	117	1.8	2.1	-2.2	14.5
P/U: 50 /50%	106	1.5	2.2	-2.0	13.2
P/U: 50 /50% Short tasks	85	0.9	1.4	-9.2	9.2

Table 5. Effect of planned tasks vs. unplanned tasks
P/U= %planned occupancy/% unplanned Occ.

All tasks have a task time according the original task time distribution. In these experiments however, the unplanned work is only disturbing the progress of planned work. Although planned work receives priority over unplanned work, planned work may have to wait at a capacity group to complete an unplanned task.
The effect of this disturbance can be minimized by creating only short tasks for the unplanned work, as can be seen in the last row in table 5. In this experiment the tasks of unplanned work have a task time between 10 and 20 hours.
The conclusion of these experiments is that decreasing the occupancy for planned work has a positive effect on delivery reliability. However, the complete workload of Parts Manufacturing usually consists of planned work. The major question is then: how to split the workload in planned and unplanned work.

URGENCY AND PRIORITY

Until now, the starting point has been a planning, generated according MRP principles. All experiments tried to realise this planning as good as possible. Usually, MRP plans the manufacturing tasks by establishing the starting times for each task. Above that, waiting times for each task are inserted in front of the actual task times. One assumes that in this way Parts Manufacturing is able to realise the delivery time to Assembly. The original goal of this level of detail is:

- to organize the process in such a way that each task will be processed on time (or exactly according to planning)
- to provide the criteria to value the state of the manufacturing process at any moment in terms of 'early', 'on time' or 'late'.

The validity of these assumptions will now be investigated. As an example, figure 2 shows the planned and real progress of one batch of parts.

Suppose the batch should be delivered at time D to

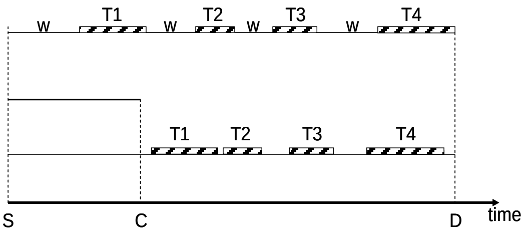

Figure 2. Planned and real progress of a batch

Assembly. The processing of the batch consists of 4 tasks: T1, T2, T3 and T4. MRP plans these tasks as shown by the top line and calculates that the required materials should be available at time S. At time C (e.g. during the next planning cycle) one observes that task T1 has not been started yet. The batch is valued as "late" on the basis of which the batch receives priority over batches that are valued early or on time. Experiments with the simulation model however show that without assigning priority values, a progress line (as the bottom line in figure 2) is just as likely as an assumed late delivery. The batch may also be delivered early without priority. The conclusion is that considering the batch as "late" at time C is premature, because one doesn't know the progress of the future trajectory yet. The final delivery time is in fact the cumulative result of temporary delays and leads.
There is another risk to planning in such detail. During execution, there is a tendency on the shop floor to realize not only the process times exactly, but also the planned waiting times. This causes a large sensitivity for disturbances, while the waiting times were meant to minimize sensitivity.
Finally, this way of planning fixes so many moments in time that a deviation of these moments is likely to appear.

In practice this happens indeed, and consequently many tasks are marked "priority tasks".
By all this, the credibility of planning is seriously damaged. In order to plan correctly it is necessary to keep the original goal of manufacturing planning in mind: "deliver the parts in time at Assembly".
Only the final delivery time should be realized by Parts Manufacturing. Planning is a control function and should only lead to interventions if this delivery time can definitely not be reached anymore. MRP apparently does not satisfy this requirement.

The planned progress of the batch of figure 2 can also be represented differently, as shown in figure 3.

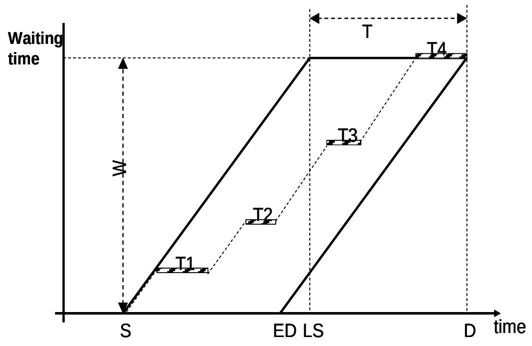

Figure 3. Planning of batch progress

Each batch is during its progress alternating in one of the following states:

- in process at a capacity group
- waiting for a capacity group

Figure 3 shows horizontally the time-axis and vertically the use of waiting time (as planned). The manufacturing process should be completed within the drawn parallelogram. At time D the batch should be delivered to Assembly, while Material Supply should have all materials available at time S. If all tasks can be processed immediately then the batch can be delivered to Assembly at time ED ("Earliest Delivery time"). If time D is to be reached then the processing should ultimately start at time LS ("Latest Starting time") and process the tasks immediately after (LS will shift to the right as soon as a task has been completed). W is the total planned waiting time and from figure 3 it is clear that

$$W = LS - S = D - ED$$

Symbolically expressed, Parts Manufacturing should reach the right hand side of the parallelogram before time D. The dotted line in the patrallelogram (the original MRP-planning) represents only one possibility to achieve this, but MRP forces this dotted line to be the single standard for manufacturing.
Suppose planning would only use S, ED, LS and L to value the progress of batches, is it possible then to effectively control the delivery times to Assembly?

Table 5 already showed the effect of the percentage of planned and unplanned work. From now on, planned work will be defined as all batches which still have slack for reaching the final planned delivery. The work without slack will be called 'urgent work'. This classification is clearly illustrated by figure 3. Periodically (in practice usually once a week), the batches with and without slack are listed. Batches without slack are the batches that will not reach the moment LS during the next week, batches that do reach this point during the next week do not have any slack left and are marked "urgent batches". These urgent batches will be planned in detail. The planning of these batches will be realistic, because all work in progress is now divided in two classes: planned and urgent batches. The urgent batches represent a far lower capacity occupancy and by assigning them priority over planned batches it should be feasible to realize the required delivery time. Two conditions should be taken into account:

- the percentage urgent batches (in working hours) should not be too large. A value around 60% occupancy by urgent batches should be preserved.
- The result of this classification is influenced by the number of tasks to be performed. The parts of the components for this company only require two tasks on average. This number is too low to control the 'urgency' effectively. In practice the number of tasks usually is around 8 on average. Therefore the number of tasks in the simulation model is changed to 8 on average (varying between 1 and 15).

During the simulation runs, the classification period is one week. The planned work is scheduled according the FCFS-rule or ODD-rule (but now based on the final delivery time D), the urgent work according increasing LS. The results are shown in table 6. This new approach is called the <u>P</u>aralleloGram <u>M</u>ethod (PGM)

	T	MDD	SDD	Min	Max
FCFS	675	9.2	11.3	-11.7	46.5
ODD	639	6.4	7.8	-11.6	27.5
PGM	617	6.1	8.2	-11.7	27.2

Table 6. MRP versus PGM

The conclusion is that PGM shows the smallest mean delivery delay, although on average still late.
In the simulation model, PGM used the sum of all waiting times as originally defined for use with MRP. The advantage of PGM however, is the applicability for decision support. PGM offers an objective way to value the progress of batches. As mentioned before, planning should preserve that the occupancy by urgent work does not (structurally) exceed approximately 60%. If it does, two possibilities are evident for use: outsourcing or working overtime. During the PGM-run of table 6 it appeared that on average 150 hours of urgent work were waiting at each capacity group. In further experiments the model was extended with outsourcing. During classification, the hours of urgent work were calculated. If

there was at least one capacity group with more than 250 (or 200 in a next experiment) hours of urgent work, it is decided that all new orders are outsourced during the next week. The results are shown below.

	T	MDD	SDD	Min	Max
PGM	617	6.1	8.2	-11.7	27.2
>250	386	0.6	4.0	-13.7	11.9
>200	374	0	4.3	-14.7	18.3

Table 7. PGM with outsourcing

Apparently, outsourcing using this criterion is successful. The outsourced work represented 5 and 10% of the working hours respectively, so the occupancy of the capacity groups dropped only slightly. The standard deviation however remains high, so further research is required to minimize this.
The comparison between MRP with ODD and PGM with outsourcing are summarized in table 8.

	MRP	PGM
Occupancy	85-90%	80-85%
Delivery	6.5 weeks late	On time
Deviation	Large (8 weeks)	Large (4 weeks)
Throughput time	16 weeks	9.5 weeks

Table 8 MRP compared to PGM

CONCLUSIONS

A number of (well known) priority rules using MRP-based planning, are examined in order to improve production control. None of these rules appeared to be satisfactory. Excluding the effect of delays in Material Supply, it is still impossible for Parts Manufacturing to realise planned delivery times. Also a change in workload composition showed no improvements. Only a significant decrease of workload that was planned in detail, resulted in an average delivery on time. The way of planning was investigated and it is shown that MRP lacks the tools to effectively intervene with production progress. An alternative approach, the parallelogram method, was primarily developed to realize the delivery time to Assembly. With this approach, better results can be achieved, if the workload can be classified in urgent and planned work. Above that, the method has the following advantages:

a. the method is simple to use. Only periodically a classification should be made of all work in progress. Two classes are defined: planned and urgent batches based on a simple unambiguous criterion.
b. The method offers more freedom of action to Scheduling. Only one priority rule applies: urgent batches have priority.
c. The method supports decision making effectively, e.g. on outsourcing, working overtime

Further research is required:

- to minimize the standard deviation of delivery times
- To improve decision support. Until now only one intervention (in this case outsourcing) has been investigated
- To preserve occupancy completely.
- To take full advantage of the shorter throughput times of PGM.

REFERENCES

Russell, R.S., Taylor III, B.W., 2006, *"Operations Management, Quality and Competitiveness in a Global Environment"*, Wiley & Sons, New York, ISBN 0-471`-69209-3

Bertrand, J.W.M., Wortmann, J.C.,1981, "*Production Control and Information Systems for Component-manufacturing shops"*, Elsevier, Amsterdam

Kemppainen, K., 2005, *"Priority rules revisited – Dominant rules, Open protocols and Integrated Order Management"*, PhD thesis, Helsinki School of Economics, ISBN 951-791-968-9

Parsons, D.J., Phelps, R.A., "Production Scheduling validity in high level supply chain models", 2001, *Proc. of the 2001 Winter Simulation Conference*, pp. 560 – 564, Arlington, USA

Veeke, H. P.M., Ottjes,J.A., 2000. "Tomas: Tool for Object-oriented Modelling And Simulation". P*roceedings of Advanced Simulation Technology Conference (ASTC2000). April 16-20, 2000*, Washington, D.C. pp. 76-81. The Society for Computer Simulation International (SCS), ISBN: 1-56555-199-0