asor

BULLETIN

ISSN 0812-860X

VOLUME 29 NUMBER 3 September 2010

Editor: Ruhul A Sarker

Published by:
THE AUSTRALIAN SOCIETY FOR OPERATIONS RESEARCH INC.
Registered by Australia Post - PP 299436/00151. Price $5.00

The Australian Society for Operations Research Incorporated

Operations Research is the application of the methods of science to complex problems arising in the direction and management of large systems of men, machines, materials and money in industry, business government and defence. The distinctive approach is to develop a scientific model of the system, incorporating measurements of factors such as chance and risk, with which to predict and compare the outcomes of alternative decisions, strategies or controls. The purpose is to help management determine its policy and actions scientifically.

ISSN 0812-860X Publisher: The Australian Society for Operations Research Inc.
Place of Publication: Canberra, Australia Year of Publication: 2010 Copyright year: 2010

Guest Editorial

The 20th National Conference of Australian Society for Operations Research incorporating the 5th International Intelligent Logistics System Conference was held on the Gold Coast, Australia, in September 2009. It is our honour, on behalf of the Australian Society for Operations Research to present these special post-conference issues, which provide a unique opportunity to maintain currency with Operations Research issues in Australia and other parts of the world. An encouraging feature of the papers is the breadth they cover in both theory and application. These special issues contain a range of papers dealing with different areas relating to the theme of the conference "Making the Future better by Operations Research". The majority of them deal with application and analysis. Some of the papers are theoretical and discuss the techniques required to analyse real life applications. As a result, the topics covered in these papers highlight the diversity of the applications of Operations Research techniques.

In this issue, Antrawan Junaputra, Katsumi Morikawa, Katsuhiko Takahashi, and Daisuke Hirotani investigate make-to-order with hierarchical workforce. The authors propose production models and rescheduling policies against order changes under such operating environment. Daniel Brodkorb and Wilhelm Dangelmaier study production planning and scheduling, and develop a new mixed integer programming model. M. A. Karim, M. R. Alam and M. A. Amin investigate the adaptation of lean production to a small make-to-order type power equipment manufacturer. The authors propose a rapid performance management technique for the effective implementation of the lean manufacturing strategy. Kari Stuart, Erhan Kozan, Michael Sinnott, and James Collier study the online scheduling relating to operating theatre department for elective and emergency patients. A robust reactive surgery assignment model is developed for minimising disruptions and cancellations, and to maximise throughput of emergency cases.

The editors of the special issues wish to express their appreciation to all authors for the contribution of their latest findings to Operations Research. We would also like to thank the reviewers for the involvement of the reviewing process in ensuring the maintenance of the highest scientific standards for these special issues. The reader is reminded that the contents prepared by the author were electronically reproduced for publication. Therefore, the views and opinions are those of the authors. Anyone with questions about a paper should contact the authors.

Guest Editors
Erhan Kozan and Andy Wong

ASOR Bulletin
Volume 29, Number 3, September 2010, pp.2-18
ISSN0812-860X | eISSN1446-6678

Production Scheduling and Rescheduling against Order Changes by Considering Hierarchical Workforce

Antrawan Junaputra[1], Katsumi Morikawa[2], Katsuhiko Takahashi[2]
and Daisuke Hirotani[2]
[1]Panasonic Corporation, Kadoma City, Osaka, Japan
[2]Grad School of Engineering, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan
(Email: antrawan@yahoo.com;{mkatsumi, takahasi, dhiro}@hiroshima-u.ac.jp)

ABSTRACT

In this research, we propose production scheduling models and rescheduling policies against order changes for make-to-order manufacturing with hierarchical workforce. The manufacturing of products is skill-dependent, and a higher qualified worker type can substitute for a lower qualified one, but not vice versa. Therefore, workers are classified into several types based on their skill, and the production capacity is defined by the number of workers in each type. In production scheduling, order assignment over the planning horizon, and order allocation among qualified workers are decided. To prepare for order changes such as changing the due dates or sizes, two mathematical models are introduced when deciding production schedules. These two models are also utilized in rescheduling with a choice of three rescheduling policies, i.e. minimum, partial and full. The performance of these models and policies is investigated in the numerical analysis. The results indicate that under higher order complexity conditions, reserving the capacity of skilled workers against order changes is preferable. In addition, this workload coordination model can absorb the effect of order changes by modifying the schedule of a smaller number of orders.

Key Words: Production scheduling, rescheduling, workforce hierarchy, order changes

INTRODUCTION

A typical multi-national manufacturing company consists of several manufacturing sites around the globe that involve in transforming raw materials into final products. On the multi-site manufacturing level, customers place orders to manufacturing sites. Upon receipt of orders, manufacturing sites distribute the orders among them, share the capacity, and order materials from suppliers or exchange materials among the sites. Each single-site manufacturing develops production order assignment and allocation, which result in production schedule. One example of multi-site manufacturing consists of three sites (e.g. Site-A, Site-B, Site-C), two suppliers (e.g. Supplier-1, Supplier-2), and two customers (e.g. Customer-I, Customer-II) as shown in Figure 1. Firstly, there are customer orders flowing from the customers to one of the sites. This site allocates each customer order to one of the three sites. Upon receiving customer orders, each site starts buying raw materials from the suppliers in order to satisfy the allocated orders. The sites are also able to exchange raw materials among themselves. Afterwards, production takes place and the finished products are either directly shipped to each customer, or they are merged in transit prior to be delivered to the customers. The production planning and scheduling problem under the multi-site manufacturing environment becomes a large schedule optimization problem and therefore it is a realistic approach to solving the problem hierarchically. Several authors have devoted their effort to modelling and solving such problems. For example, Kanyalkar and Adil (2005) develop an integrated production planning model for both aggregate and detailed level in the multi-site manufacturing environment without internal material exchanges among manufacturing sites. Meijboom and Obel (2007) utilize inventory internal exchange under multi-site manufacturing environment. Leung *et al*. (2007) develop a robust optimization model to solve multi-site production planning problem with uncertain data.

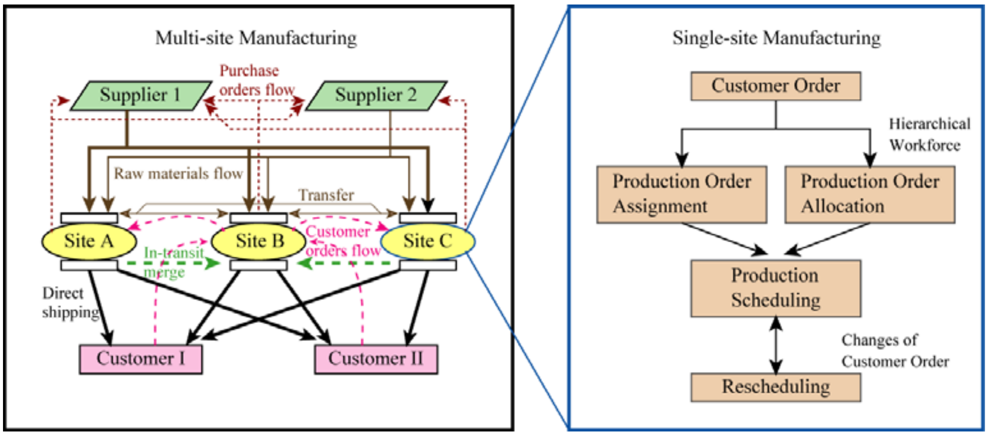

Figure 1. Example of multi-site manufacturing

As a first step towards developing a comprehensive production planning and scheduling approach to a multi-national make-to-order company that activated our study and can be modelled as Figure 1, this study focuses on the single-site production scheduling problem within the decision making hierarchy. The production scheduling to be considered involves the assignment of orders to production periods and allocation of orders to candidate workers. Complexity arises whenever the products are skill-dependant, i.e. hierarchical workforce is in place. Hierarchical workforce means that a higher qualified worker can substitute for a lower qualified one, but not vice versa. For examples there are four types of workers, i.e. worker type 1, worker type 2, worker type 3 and worker type 4. The sequence of qualification starts from worker type 1 as the highest qualified worker, followed by worker 2, 3 and 4 subsequently. In this case, worker type 1 can substitute for worker type 2, but worker type 2 cannot substitute for worker type 1. Another example, worker type 2 can substitute for worker type 3, but worker type 3 cannot substitute for worker type 2, and so on. Furthermore, complexity also arises whenever customers change their orders, either changing the due dates or sizes, which affect the current production schedule and generate the needs of rescheduling.

The purpose of this study is to present formulations of production scheduling and rescheduling, and rescheduling policies subject to hierarchical workforce and dynamic customer order changes. A production scheduling model is developed to minimize the total cost consisting of earliness cost, tardiness cost, and overtime cost. Prospective order changes are implicitly considered by expanding the above model to reserve the capacity of skilled workers. Rescheduling is conducted periodically to include order changes that have arrived within one period. Orders to be included in the rescheduling model are defined by the rescheduling policies. The minimum set of orders is given by the minimum rescheduling policy, while the maximum set is offered by the full rescheduling policy. Within these two extreme policies, another policy named partial is also introduced and these rescheduling policies are examined experimentally to investigate the performance of the final schedule in terms of the total cost and the number of schedule modifications by varying scheduling models, rescheduling policies, and the degree of order changes.

This paper is organized as follows. Related previous studies are briefly discussed in section 2. Assumptions and the formulation of the production scheduling models are presented in section 3. Order changes and rescheduling policies are explained in section 4. Numerical analysis based on make-to-order optoelectronics manufacturing cases is provided in section 5. Conclusions are made in the last section.

LITERATURE REVIEW

Mixed integer programming formulations for reactive scheduling proposed by Sawik (2007a) gives several key ideas when developing production scheduling models and rescheduling policies in this study. A hierarchical framework for deciding the long-term order assignment and the short-term machine assignment, and multi-objective long-term production scheduling are also presented by the same author (Sawik, 2006 and 2007b). Scheduling under uncertain environments and rescheduling are one of active research topics and there are many published papers. For example, Aytug *et al.* (2005) reviews over 110 papers on executing production schedules in the presence of unforeseen disruptions, and Herroelen and Leus (2005) survey the fundamental approaches for scheduling under uncertainty by referring to over 90 published papers. Vieira *et al.* (2003) present some common terms on rescheduling, and propose a framework for understanding rescheduling strategies, policies, and methods. Recent technologies such as RFID provide accurate and timely information to us, and Hozak and Hill (2009) indicate that there are important issues on replanning and rescheduling frequencies. In their research it is indicated that internal planning research has generally concluded that higher frequencies is negative.

Regarding to the issue of workforce management, Croci *et al.* (2000) indicate the importance of human resources in automated assembly systems. Bhatnagar *et al.* (2007) discuss the manpower planning under hierarchical or nested structure of worker skills. They assume the availability of contingent manpower and consider the impact of learning. Cambini and Riccardi (2009) solve the hierarchical workforce problem by an exact recursive algorithm. Multi-skilled workforce or hierarchical workforce is also often discussed in the field of personnel scheduling. Eitzen and Panton (2004) develop a dynamic rostering method for handling fluctuating staff demand. Seçkiner *et al.* (2007) develop an integer programming model under the compressed workweeks. A MILP model is proposed by Hetz *et al.* (2010) to solve a multiple-shift workforce planning problem under annualized hours. Replanning working time is discussed by Corominas and Pastor (2010). They consider two main objectives, i.e. the cost of the new plan and the stability of the schedule.

The present literature review indicates that there are many published papers dealing with production scheduling, rescheduling, or hierarchical workforce. As far as we know, however, our problem that involves these three factors has not been addressed in the literature. In our opinion, many manufacturing companies involve these factors and therefore hopefully this research can help solve the actual production scheduling problems.

MODEL DEVELOPMENT

Assumptions

Let us assume that the target production system can be modelled as a single stage process, and the production capacity of this system is determined by the working hours of workers. The number of workers is given and regular production and overtime production are available. The production scheduling horizon is divided into several periods with the same length each. A typical example is that one period corresponds to one week, and four weeks constitute the planning horizon of one month. Customers throughout the world request various kinds of products, and because of the additional differences in regulations, it is generally difficult to utilize overproduced products to other orders. Therefore, the manufacturer must produce the amount of requested quantities of the requested products after receiving orders officially. However, customers may ask the change of orders after releasing the official order.

As mentioned earlier, when considering production scheduling, orders with similar specifications, for example similar materials requirements, similar production processes, and similar processing skills are aggregated into a family. In order to consider the production capacity based on hierarchical workforce, each of these product families in this production scheduling model will be assigned to specific worker types, as shown in Table 1. Within the planning horizon, the skill level of each worker does not change, and the set of orders to be produced within the planning horizon is given from the upper decision level.

Table 1. Relationship between product family and hierarchical workforce

Product family	Hierarchical workforce
Family 1	Can be produced only by worker type 1
Family 2	Can be produced by worker type 1 and 2
Family 3	Can be produced by worker type 1, 2, and 3
Family 4	Can be produced by worker type 1, 2, 3, and 4

<u>Production Scheduling Model</u>

The notation and proposed model are as follows.

Indices:

j	production order; $j = 1,\ldots,J$
t	production scheduling period; $t = 1,\ldots,T$
w	worker type; $w = 1,\ldots,W$

Parameters:

ψ_j	set of all worker types who can produce order *j*
L_j	set of all production orders that can be produced by worker type *w*
θ_j	product family of order *j*
a_j	material arrival date of production order *j*
d_j	due date of production order *j*
s_j	size of production order *j*
z_w	number of workers in worker type *w*
p_{wt}	processing time per unit of product family *i* by worker type *w*
m_w^r	available regular production time per period per worker for type *w*
m_w^o	available overtime production per period per worker for type *w*
δ_j	cost of tardiness of production order *j* per period per unit
ε_j	cost of earliness of production order *j* per period per unit
o_w	unit overtime production cost of worker type *w*

Decision variables:

V_{jt}	1, if production order *j* is scheduled in period *t*; 0, otherwise
Y_{jwt}	fraction of production order *j* assigned to worker type *w* to be produced in period *t*
D_j	tardiness of order *j*
E_j	earliness of order *j*
O_{wt}	overtime production of worker type *w* in period *t*

Production scheduling model:

$$\text{Minimize} \quad \sum_{w=1}^{W} \sum_{t=1}^{T} o_w O_{wt} + \sum_{j=1}^{J} \left(\varepsilon_j s_j E_j + \delta_j s_j D_j \right) \tag{1}$$

Subject to

$$\sum_{t=a_j}^{T} V_{jt} = 1, \quad j = 1, \ldots, J \tag{2}$$

$$V_{jt} = \sum_{w \in \psi_j} Y_{jwt}, \quad j = 1, \ldots, J; t = a_j, \ldots, T \tag{3}$$

$$D_j \geq \sum_{t=a_j}^{T} t V_{jt} - d_j, \quad j = 1, \ldots, J \tag{4}$$

$$E_j \geq d_j - \sum_{t=a_j}^{T} t V_{jt}, \quad j = 1, \ldots, J \tag{5}$$

$$\sum p_{w\theta_j} s_j Y_{jwt} \leq m_w^r z_w + O_{wt}, \quad w = 1, \ldots, W; t = 1, \ldots, T \tag{6}$$

$$O_{wt} \leq m_w^o z_w, \quad w = 1, \ldots, W; t = 1, \ldots, T \tag{7}$$

$$V_{jt} \in \{0,1\}, \quad 0 \leq Y_{jwt} \leq 1, \quad j = 1, \ldots, J; w = 1, \ldots, W; t = 1, \ldots, T \tag{8}$$

$$D_j, E_j, O_{wt} \geq 0, \quad j = 1, \ldots, J; w = 1, \ldots, W; t = 1, \ldots, T \tag{9}$$

The objective function (1) demands the minimization of cost of overtime production, earliness, and tardiness of production orders. It is expected that all production orders can be completed within the regular production time of the workers. If the amount of time required for the production order exceeds the capacity of regular production time, as much as possible all orders are processed in overtime with the least cost. However, this production scheduling model also allows the completion of production order at an earlier or a later time, although it might occur in the earliness or tardiness cost.

Equation (2) is aimed at ensuring that production order *j* is assigned only in one period from a_j to *T*. This model assumes that required materials for order *j* are available at the beginning of period a_j, and it is not allowed to produce any order using two or more periods. Equation (3) indicates that a production order is allowed to be produced by several workers as long as the workers are qualified to produce the order. Therefore, a real value from zero to one is allocated to the worker type to indicate that a fraction of the order is produced by that particular worker type. Furthermore, in order to ensure that this allocation decision is consistent with Equation (2), the total value of order allocation for each production order must be equal to V_{jt}. As mentioned earlier, this production scheduling model allows the completion of an order ahead of or beyond its due date. It is assumed that all delivery activities are conducted at the end of every period, and thus if an order is processed in period *t*, and its due date is also *t*, then there is no earliness and no tardiness in this order. Equation (4) defines the tardiness which is calculated as the number of periods a production order surpasses its due date. On the other hand, by exchanging two terms in the right hand side of Equation (4), Equation (5) calculates the earliness of order. The capacity constraint is given by Equations (6) and (7). In every period, the processing time of all production orders is limited to the total available regular production time and over time.

This scheduling model is named *baseline* model hereafter because this model includes fundamental constraints and direct objective function of minimizing total cost over the planning horizon. Clearly this baseline model should produce an optimal solution at the beginning of period 1. If there are no order changes over the planning horizon, conducting production activities based on the solution incurs the minimum cost. This study, however, assumes that there will be some order changes over the planning horizon. Therefore, under the workforce hierarchy constraint, it may be a reasonable idea to reserve the production capacity of higher skilled workers against future order changes. Please note that higher skilled worker can substitute for lower skilled worker so that they can anticipate more varied orders. This idea is realized by adding the following Equation (10) into the baseline model. This expanded model is named *workload coordination* model.

$$\left(\frac{1}{z_w}\sum_{j\in L_w} p_{w\theta_j} s_j Y_{jwt}\right) - \left(\frac{1}{z_\sigma}\sum_{j\in L_\sigma} p_{\sigma\theta_j} s_j Y_{j\sigma t}\right) \le 0, \quad w = 1,\ldots,W-1; \sigma = w+1,\ldots,W; t = 2,\ldots,T \tag{10}$$

Please note that L_σ indicates the set of orders that can be processed by worker type σ ($\sigma \neq 1$). From the definition of workforce hierarchy, these orders can also be processed by a higher worker type $w = 1,\ldots,\sigma-1$. Equation (10) indicates that when allocating the total workload of these orders among different worker types, do not allocate higher workload to a higher skilled worker on a per worker basis. In general, each worker type has a corresponding primal product family such as product family 1 for worker type 1, family 2 for type 2, and so on. Therefore workers in type 1 must allocate their production capacity for product family 1 by priority, and this means that the worker type 1 should not accept too much workload of orders for product families 2, 3, and others. By applying this discussion for other worker types, we can assume that this constraint is relatively moderate, i.e. it does not reduce the region of feasible solutions greatly. Therefore, it is expected that the workload coordination model has a feasible optimal solution in general and this solution tends to reserve more capacity to higher skilled workers to anticipate order changes in the future.

As we will explain later, the rescheduling step is activated at the end of each period. This means that the production schedule of the first period is given by the solution of one of scheduling models, and this schedule will be implemented even if some order changes arrive within the first period. Therefore, there is no need to prepare for the order changes when scheduling the first period. Equation (10) is thus applied only from period 2 to *T*.

ORDER CHANGES AND PRODUCTION RESCHEDULING

Order Changes

In actual production environments, order changes after generating the initial schedule are not rare. Typical changes are; order cancellation, size decrease, due date push-out, due date pull-in, size increase, and additional orders. In this study the customer order changes are not processed every time the changes occur, but included into a periodic rescheduling. To be more exact, order change requests that have arrived within a period are stored, and a reschedule is generated at the end of this period by reflecting all order changes in addition to unprocessed and unchanged orders.

We assume that an order change is accepted only if it arrives earlier than its due date. For example, if the due date of an order is period 3, a change request for this order will be accepted only when the request arrives by the end of period 2. Needless to say, for a change of due date, it is automatically rejected if the requested due date is earlier than the next period. This study assumes that the production of an order must be completed within a period, and it is possible to start production of an order before its due date, provided that its corresponding materials are available and associated earliness cost is acceptable. Therefore, there is a possibility that an order is processed earlier than its due date, and then the order change request for this order arrives. Let us consider the following example; due date of order A is period 3, but it is planned to produce in period 2. Before the end of period 2, the change request for order A arrives. In this case there are two ways to handle this order change; (i) deny the request, or (ii) accept the request. This study assumes that the manufacturer must accept order changes as much as possible, and therefore, in this case we accept this change request. However, if the order change includes the order size increase or decrease, additional problems arise. If the size of order is decreased, there is no problem in demand fulfilment, but how to handle the surplus products is a problem. As stated earlier, under make-to-order environments, it is generally difficult to utilize the surplus products to other orders. Therefore the surplus products will be disposed at a cost of quantity of surplus products multiplied by the earliness of the order. This means that in this study we ignore the size decrease when calculating total cost. The total earliness cost is thus based on the initial size of the order even if the size is decreased after the production. If order size is increased, on the other hand, additional considerations are required. In general, the customer may want to receive the whole quantity at one time. Thus the order increase indicates an additional production of the order, and if this second production is behind its due date, penalty will be added to the whole quantity, i.e. changed quantity. Apparently if additional production is conducted within its due date, we can calculate its earliness cost by considering these two times of production separately.

Rescheduling Policies

Rescheduling is generally inevitable under uncertain environments. In shop floor management, however, it may be unfavourable to modify production schedules frequently because it may cause miscommunication and various troubles. Frequent schedule modifications will also result in ineffective detailed production operations. Therefore, it is a realistic approach to restricting the schedule modifications within an acceptable range. However, schedulers do not want uneconomical production schedules by severely limiting the set of orders to be rescheduled. Based on these discussions, this study considers three rescheduling policies as follows:

Minimum rescheduling: The schedule of all unchanged orders is frozen, and only changed orders are rescheduled.

Full rescheduling: All unprocessed orders can be rescheduled.

Partial rescheduling: In addition to changed orders, orders that are currently planned to be (partially) allocated to worker type 1 are also candidate for rescheduling. The schedule of other orders is frozen.

The meaning of minimum and full rescheduling policies is straightforward. The minimum rescheduling policy restricts rescheduling activities only to changed orders. This policy is, therefore, expected to produce the minimum number of schedule modifications. The anticipated negative effect of this policy is higher production cost. The full rescheduling, on the other hand, is attractive in minimizing the production cost by eliminating all restrictions in rescheduling. The number of schedule modifications is, however, expected to show a higher value. Between these two extreme policies, this study proposes an alternative policy; partial rescheduling. In addition to changed orders, this policy also adds unchanged orders that are currently (partially) allocated to worker type 1 to the set of orders to be rescheduled. Thus this policy will produce the middle performance in terms of the total cost and the number of modified schedules.

A rescheduling model is also required when applying one of these three rescheduling policies. In previous subsection, two scheduling models are developed, i.e. baseline model and workload coordination model, and these models can basically handle these rescheduling policies with minor adjustments such as fixing the value of decision variables of frozen orders, and excluding orders that have been already processed at that time. In this study, the same model is used in scheduling and rescheduling to make our approach simple. For example, if the initial schedule is generated by the baseline model, then the rescheduling is also realized by the baseline model.

NUMERICAL ANALYSIS

Experimental Setup

The purpose of numerical analysis is to investigate the performance of two production scheduling and rescheduling models under three rescheduling policies. A data set, which consists of 45 orders, 4 worker types, and 4 periods of planning horizon, was created. Three order complexity conditions of low, medium, and high were also prepared as shown in the last column in Table 2. In low complexity condition, all orders were classed as product family 3 or 4. On the other hand, in high complexity condition, 13 orders requested products in family 1, and 9 orders in family 2. The number of workers in worker type 1, 2, 3, and 4 was 1, 1, 5, and 4, respectively. The available regular time per worker in a period was varied from 40 to 44, while the maximum overtime production time was 8 per period regardless of worker types. Other important data were as follows: $(p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4}) = (0.26, 0.24, 0.22, 0.2)$, $(p_{2,2}, p_{2,3}, p_{2,4}) = (0.31, 0.29, 0.27)$, $(p_{3,3}, p_{3,4}) = (0.4, 0.38)$, $p_{4,4} = 0.8$; $\delta_j = \{2, 2.5\}$, $\varepsilon_j = \delta_j / 10$; $(o_1, o_2, o_3, o_4) = (10, 7, 5, 1)$. We assumed that a higher skilled worker can process orders faster, and incurs higher unit overtime production cost than a lower skilled worker.

Under the above conditions, three levels of uncertainty in order changes were considered; low, medium, or high uncertainty. Order changes under these conditions were generated by the following steps:

Step 1: Identify the set of candidate orders Ω. The contents of Ω is given by the order of which due date is greater than the current period.

Step 2: Select $\lfloor |\Omega| \cdot r \rfloor$ orders to be modified from the candidate set Ω, where r takes 0.1, 0.2, or 0.4 to represent low, medium, or high uncertain environment, respectively.

Step 3: Changed due date d'_j and order size s'_j of a selected order j is given by the following equations:

Low uncertainty: $d'_j = d_j,\ s'_j = [0.9s_j,\ 1.1s_j]$,

Medium uncertainty: $d'_j = [\max\{a_j,\ d_j - 1\}, \min\{d_j + 1,\ 4\}],\ s'_j = [0.8s_j, 1.2s_j]$,

High uncertainty: $d'_j = [\max\{a_j, d_j - 2\}, \min\{d_j + 2, 4\}],\ s'_j = [0.6s_j, 1.4s_j]$,

where the symbol $[x, y]$ returns an integer value within x and y uniformly.

For any order, the number of changes was limited to at most one within periods 1 to T. As the above steps indicate, cancel of orders and arrival of new orders were excluded. At the beginning of period 1, all order changes to be arrived over the planning horizon were generated randomly and stored as an order change pattern. A specified number of order change patterns were generated using a random number seed.

An initial schedule is generated over four periods, and then production operations in period 1 start. Some order changes arrive within period 1, and at the end of this period, completed products of their due dates are period 1 are delivered to the customers, and a rescheduling step over periods 2 to 4 is conducted to reflect order changes. As Figure 2 illustrates, these activities are repeated until the final production schedule in period 4 is obtained. Please note that order changes arrived in period 4 are excluded because these changes should be handled in the next scheduling.

Under the rolling horizon manner, it is also a realistic approach to fixing the planning length of the rescheduling to four periods. However, the fixed planning length causes difficulty in investigating the performance of the production (re-)scheduling models and rescheduling policies under uncertain environments because of added order information for periods 5, 6, and so on at later rescheduling steps. Therefore, in this experiment we fixed order information over four periods from the initial scheduling to the final rescheduling as shown in Figure 2.

In the numerical analysis, values of two performance measures were obtained; the total cost over four periods, and the number of schedule modifications over four periods. The total cost was obtained by evaluating the implemented schedule in four periods as shown in Figure 2. The number of schedules modified is obtained by comparing the last schedule and the obtained reschedule, and then incrementing the counter if the planned quantity is modified from the last value. The number of schedule modifications used in this study is similar to the unit reallocation cost discussed by Kopanos *et al.* (2008). Figure 3 illustrates this calculation by using an example. This example shows the scheduled quantity of a worker from period t to $t + 2$ at the beginning of period t, and the rescheduled quantity by reflecting order changes arrived in period t and generated at the end of t. By comparing two schedules, we can find that in five cells their quantities are modified. The modifications include from a positive quantity to zero, and vice versa. This value is obtained three times, i.e. at the end of period 1, 2, and 3, and the total value of them is used to evaluate the sensitivity of scheduling/rescheduling models and rescheduling policies on order changes.

Table 2. Experimental data consisting 45 orders and three levels of order complexity

Order no.	Arri-val	Due date	Order size	Order complexity		
				Low	Medi.	High
1	1	1	130	3	3	3
2	1	1	90	4	4	4
3	1	1	10	3	1	1
4	1	1	50	4	2	1
5	1	1	80	4	4	4
6	1	1	50	4	3	3
7	1	1	130	4	4	4
8	1	1	10	3	1	1
9	1	1	130	4	4	4
10	1	1	130	4	4	2
11	1	1	50	3	3	3
12	1	1	50	3	1	1
13	1	1	120	4	4	3
14	1	2	190	4	4	4
15	1	2	120	4	2	2
16	1	2	130	4	4	1
17	1	2	10	3	1	1
18	1	2	30	4	4	3
19	1	2	40	4	4	2
20	1	2	200	3	3	3
21	2	2	50	4	4	4
22	2	2	250	3	3	3
23	2	3	70	3	2	2
24	2	3	100	3	3	1
25	2	3	180	3	3	3
26	2	3	10	3	1	1
27	2	3	180	4	4	4
28	2	3	130	4	4	4
29	2	3	10	3	1	1
30	2	3	10	3	1	1
31	2	3	30	4	4	2
32	2	3	30	4	4	2
33	2	3	300	3	3	3
34	2	4	220	4	4	4
35	2	4	100	3	3	3
36	2	4	200	4	4	3
37	2	4	70	4	4	4
38	2	4	30	4	4	3
39	2	4	20	3	1	1
40	2	4	70	4	2	1
41	2	4	20	3	1	1
42	2	4	50	4	4	2
43	2	4	30	4	4	2
44	2	4	50	4	4	4
45	2	4	40	4	3	2

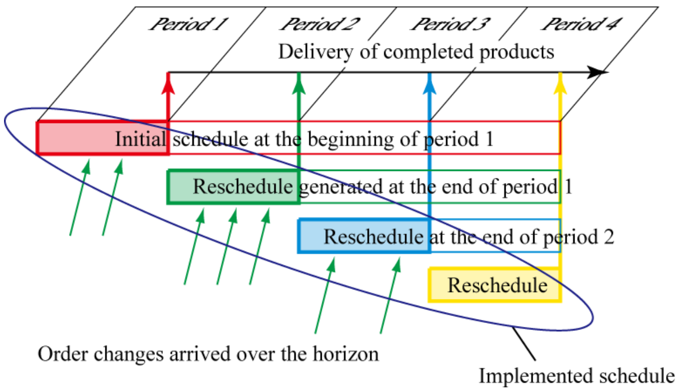

Figure 2. Scheduling, rescheduling, order changes, and implemented schedule

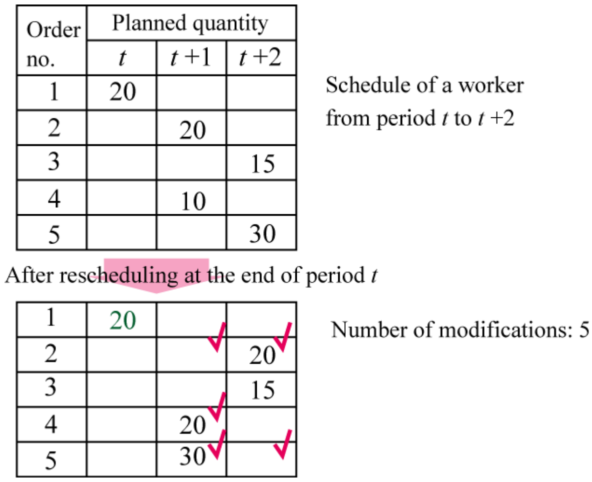

Figure 3. Example of counting the number of schedule modifications

Results and Discussions

By varying the level of order complexity (3 levels), rescheduling policies (3 policies), scheduling/rescheduling models (2 models), level of order changes (3 levels) and available regular time (5 levels), 270 combinations in total were prepared. One experimental run consisted of one scheduling and three rescheduling phases to cover the planning horizon, and in one combination, 100 runs with different order change patterns were implemented to obtain the average total cost and the average number of schedule modifications. In some cases the models indicated that there is no feasible solution. The infeasibility comes from two reasons; there is insufficient total production time, or no feasible solution that satisfies constraint (10) in the workload coordination model. When calculating average values, these infeasible problem instances were ignored and the number of infeasible problem instances was recorded. In each level of order complexity, the seed value of random number generator was fixed to reduce the variance of results.

Figures 4, 5, and 6 summarize 270 experimental results in terms of the average number of schedule modifications in upper graphs, and the average total cost in lower graphs, from low order complexity (Figure 5) to high complexity (Figure 7). If there was no feasible solution in some problem instances, the number of infeasible problem instances is displayed in the total cost graph. Solid lines and dotted lines indicate the performance of baseline model, and workload coordination model, respectively. Pertaining to the infeasible problem instances, values in bold font and normal font are related to the baseline model, and workload coordinated model, respectively.

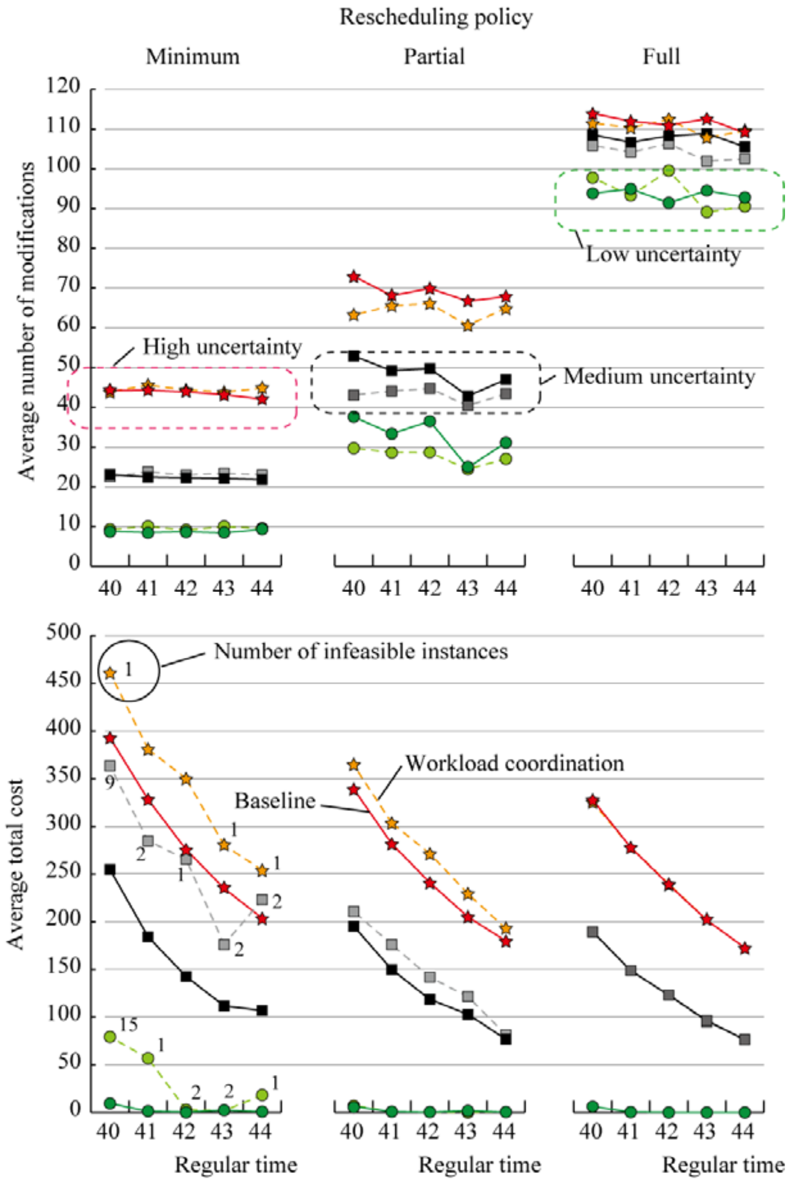

Figure 4. Average total cost and schedule modifications under low order complexity

From these graphs, we can easily find the following points.

(1) The average number of schedule modifications increased when rescheduling policy was changed from minimum to partial, and partial to full. On the other hand, the average total cost decreased by allowing more orders can be rescheduled.

(2) In minimum rescheduling policy, the average number of schedule modifications was insensitive to the available regular production time. The baseline model and the workload coordination model both indicated similar values in terms of the average number of modifications.

(3) Increasing the complexity of orders decreases the potential number of feasible schedules. This fact was supported by the graphs displaying that the average

number of schedule modifications decreased with the increase of order complexity, especially under the full rescheduling policy.

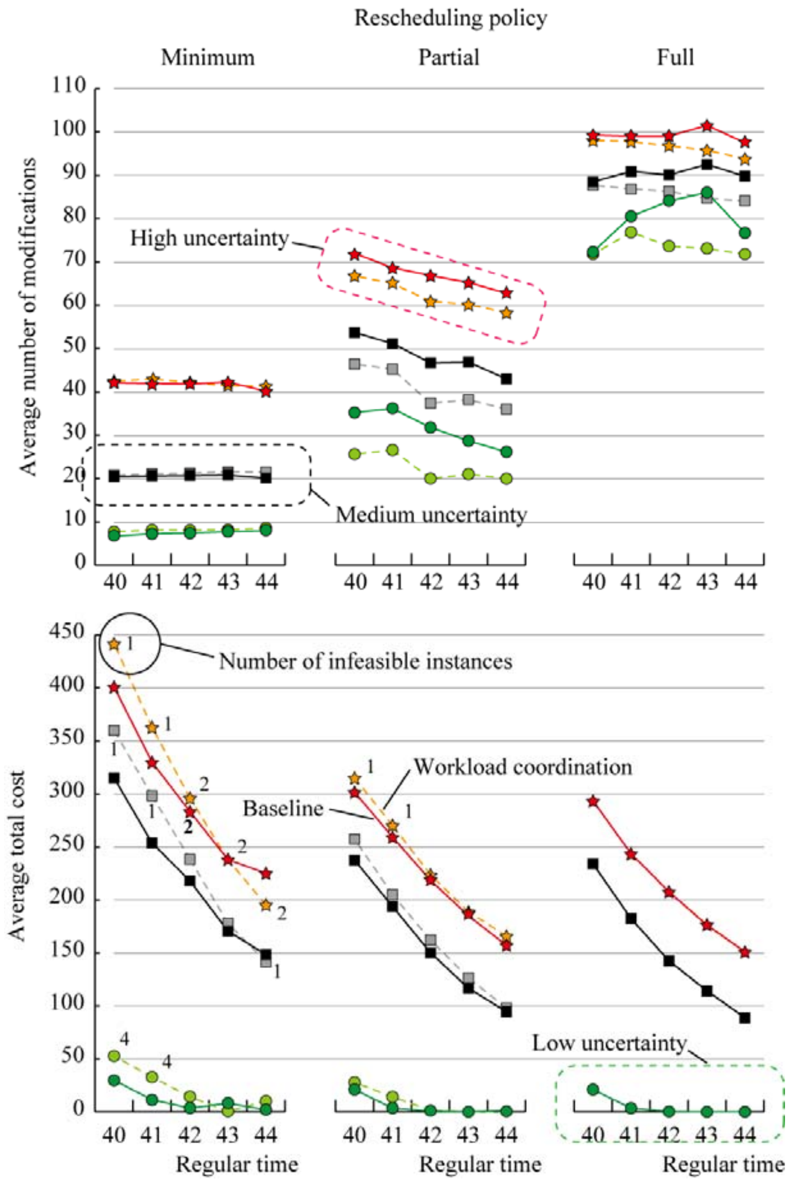

Figure 5. Average total cost and schedule modifications under medium order complexity

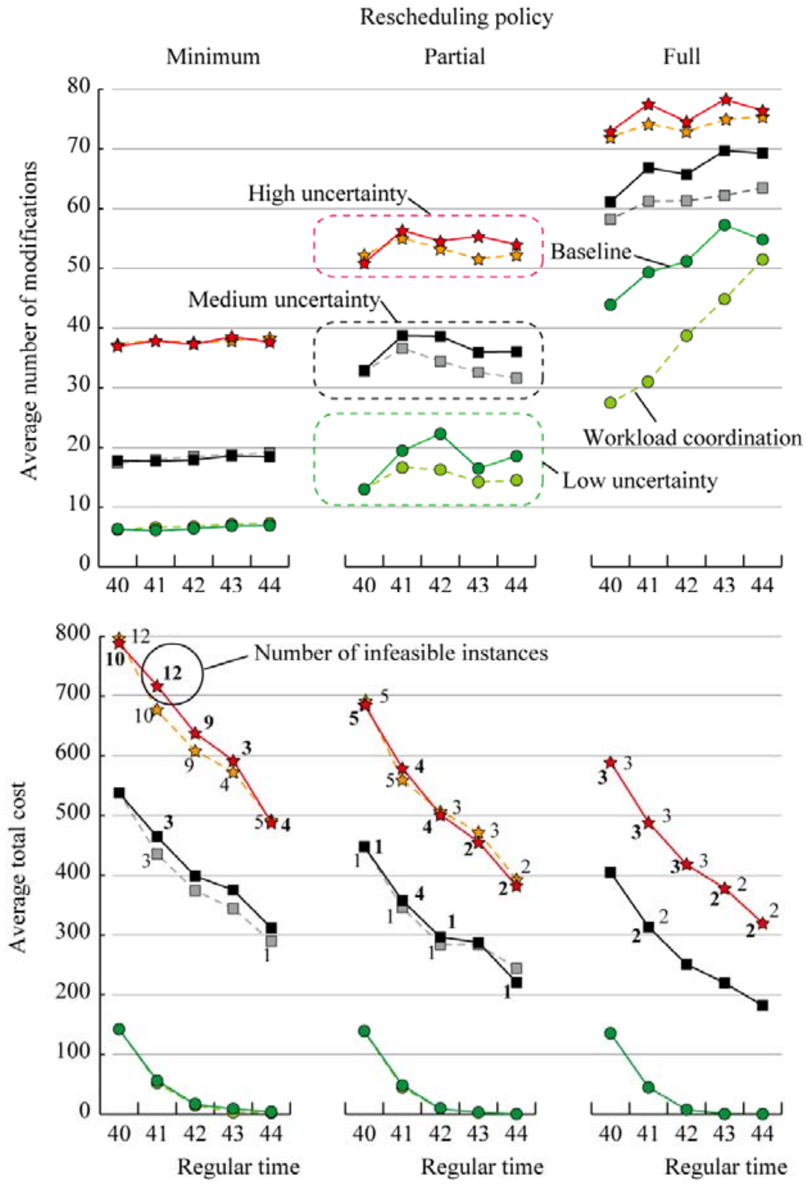

Figure 6. Average total cost and schedule modifications under high order complexity

The above points fit with our intuitive anticipation of the performance of two models under investigated conditions. Detailed examinations of these results give the following additional findings.

(1) In full rescheduling policy, the average costs of two models were almost the same, while the number of modifications often indicated significant differences between them, except that the order complexity was low.

(2) In low order complexity condition, the workload coordination model indicated higher values in average total cost under minimum and partial rescheduling policy. Although the average number of modifications was relatively favourable, the average cost indicates that the workload coordination model was unattractive except the low uncertainty level and partial rescheduling policy. The number of infeasible problem instances also indicates the inferiority of the workload

coordination model. These results are probably reasonable because under the low order complexity condition, workers in types 3 and 4 can cope with order changes, and therefore there is almost no need to reserve production capacity of higher skilled workers for order changes. This is not an intended condition of Equation (10) added in the workload coordination model. Some problem instances indicated that when order changes were fed into the workload coordination model, the model did not find a feasible solution because of the added constraint (10).

(3) The workload coordination model indicated favourable results in terms of the average number of modifications under medium order complexity level. However, by restricting the set of orders to be rescheduled, the average total cost of the workload coordination model indicated slightly higher values than the baseline model. Nevertheless, the statistical test indicated that the differences of average total cost between two models under the partial rescheduling policy and medium or high uncertainty were all insignificant under a significance level of 0.05. The number of infeasible problem instances was relatively small in this complexity.

(4) Under high order complexity condition, the average cost of workload coordination model indicated relatively favourable results in terms of the average number of modifications and the average total cost. Both models indicated that there is no feasible solution in some problem instances, but their values are similar between them.

From the above discussion we can conclude that the baseline model performs relatively favourably in terms of the average total cost, although it generally shows larger number of schedule modifications. In both the average number of modifications and the total cost, the workload coordination model is expected to perform favourably under the condition that the complexity of the allocated set of orders is roughly consistent with the available capacity of each worker type.

If schedulers want to minimize the number of schedule modifications, the minimum rescheduling policy is the best selection. The model section decision should carefully consider the level of order complexity. If there are many low complexity orders then the baseline model is the solution, while the workload coordination model becomes a superior solution under the high complexity condition as indicated in Figure 6. Except for the low uncertainty condition, increasing the weight on the total average cost leads to the partial or full rescheduling policy, and the workload coordination model is expected to indicate better performances under the medium and high order complexity conditions.

CONCLUSION

This research develops scheduling and rescheduling models, and rescheduling policies to solve problems of assigning and allocating production orders, and responding to customer order changes considering hierarchical workforce. To prepare for order changes such as changing the due dates or sizes, two scheduling models are introduced and these models are also used in rescheduling. Three rescheduling policies, i.e. minimum, partial and full policies, are also introduced to decide the set of orders to be rescheduled when responding to customer order changes.

The numerical computation reveals that, for low order complexity condition, the baseline model that involves fundamental constraints is more preferable; and for medium and high order complexity conditions, the workload coordination model that reserves the capacity of higher skilled workers is more preferable. Moreover, if full rescheduling policy is utilized, then both models indicate similar average total costs but the workload coordinate model realizes less schedule modifications. The workload coordination model is also preferable under the partial rescheduling policy in terms of the number of schedule modifications.

Apart from the potential benefits of applying the proposed scheduling and rescheduling models under appropriate rescheduling policies, there are still abundant of opportunities to improve the current proposed models. One example of direction for future research is how to link the proposed scheduling models with an upper multi-site aggregate production planning model. When deciding order allocation, expected level of order changes, hierarchical workforce of each site, and order complexity should be considered in the aggregate planning. Another example of direction for future research is to explicitly consider materials availability, especially in the rescheduling phase. Inclusion of the short-term scheduling decision is also an important research issue.

ACKNOWLEDGEMENTS

This work was partially supported by the Grant-in-aid for Scientific Research, 19510149.

REFERENCES

Aytug, H., Lawley, M. A., McKay, K. Mohan, S. and Uzsoy, R. (2005), Executing production schedules in the face of uncertainties: A review and some future directions, European Journal of Operational Research, 161, 86-110.

Bhatnagar, R., Saddikutti, V. and Rajgopalan, A. (2007), Contingent manpower planning in a high clock speed industry, *International Journal of Production Research*, **45**, 2051-2072.

Cambini, R. and Riccardi, R. (2009), Theoretical and algorithmic results for a class of hierarchical fleet mix problems, *European Journal of Operational Research*, **198**, 741-747.

Corominas, A. and Pastor, R. (2010), Replanning working time under annualised working hours, *International Journal of Production Research*, **48**, 1493-1515.

Croci, F., Perona, M. and Pozzetti, A. (2000), Work-force management in automated assembly systems, *International Journal of Production Economics*, **64**, 243-255.

Eitzen, G. and Panton, D. (2004), Multi-skilled workforce optimisation, *Annals of Operations Research*, **127**, 359-372.

Herroelen, W. and Leus, R. (2005), Project scheduling under uncertainty: Survey and research potentials, *European Journal of Operational Research*, **165**, 289-306.

Hertz, A., Lahrichi, N. and Widmer, M. (2010), A flexible MILP model for multiple-shift workforce planning under annualized hours, European Journal of Operational Research, 200, 860-873.

Hozak, K. and Hill J. A. (2009), Issues and opportunities regarding replanning and rescheduling frequencies, International Journal of Production Research, **47**, 4955-4970.

Kanyalkar, A.P. and Adil, G.K. (2005), An integrated aggregate and detailed planning in a multi-site production environment using linear programming, *International Journal of Production Research*, **43**, 4431-4454.

Kopanos, G. M., Capón-García, E., Espuña, A. and Puigjaner, L. (2008), Cost for rescheduling actions: A critical issue for reducing the gap between scheduling theory and practice, *Industrial & Engineering Chemistry Research*, **47**, 8785-8795.

Leung, S.C.H., Tsang, S.O.S., Ng, W.L. and Wu, Y. (2007), A robust optimization model for multi-site production planning problem in an uncertain environment, *European Journal of Operational Research*, **181**, 224-238.

Meijboom, B. and Obel, B. (2007), Tactical coordination in a multi-location and multi-stage operations structure: A model and a pharmaceutical company case, *Omega*, **35**, 258-273.

Sawik, T. (2007a), Integer programming approach to reactive scheduling in make-to-order manufacturing, *Mathematical and Computer Modelling*, **46**, 1373-1387.

Sawik, T. (2007b), Multi-objective master production scheduling in make-to-order manufacturing, *International Journal of Production Research*, **45**, 2629-2653.

Sawik, T. (2006), Hierarchical approach to production scheduling in make-to-order assembly, *International Journal of Production Research*, **44**, 801-830.

Seçkiner, S.U., Gökçen, H. and Kurt, M. (2007), An integer programming model for hierarchical workforce scheduling problem, *European Journal of Operational Research*, **183**, 694-699.

Vieira, G. E., Herrmann, J. W. and Lin, E. (2003), Rescheduling manufacturing systems: A framework of strategies, policies, and methods, *Journal of Scheduling*, **6**, 39-62.

ASOR Bulletin
Volume 29, Number 3, September 2010, pp.19-33
ISSN0812-860X | eISSN1446-6678

A Concept for Automated Operative Production Planning under Practical Conditions

Daniel Brodkorb and Wilhelm Dangelmaier
Heinz Nixdorf Institute,
University of Paderborn, Fürstenallee 11, 33102 Paderborn, Germany
(Email: Daniel.Brodkorb@hni.upb.de)

Abstract

Production planning and scheduling is still done manually in practice without support of mathematical methods. Optimality of created plans cannot be guaranteed. Automatic methods for production planning lack of acceptance because they do not consider the great number of constraints which exist in real life. In this paper, a concept for an automatic planning and scheduling method is presented which takes all available constraints into account. The first part of the concept is a method on the basis of the Discrete Lot-Sizing and Scheduling Problem (DLSP). The new linear mixed integer model considers the most important constraints of a practical problem. Available optimization software can be used to generate plans. The influence of less important targets is limited and computation time is reduced significantly, as they are simply output. This improves acceptance of the application of mathematical methods in practice.

Key words: *Lot Sizing, Scheduling, Mixed Integer Programming*

Introduction

As a consequence of great pricing pressure from automobile manufacturers, the strain on automobile suppliers has increased. In order to remain competitive, many automobile suppliers feel impelled to undertake extensive organizational and technological adaptations in various areas. The cutting of production costs plays a decisive role in this context. Increasing transport costs, besides leading to the relocation of production facilities to low-wage countries, also lend greater importance to the analysis and optimization of existing production sites. The technological support of formerly manually accomplished planning and control of production creates new possibilities for improvement. The complexity of the planning problem increases due to the increasing number of variables, technological solutions become necessary in order to generate production plans that are in accordance with the entrepreneurial objectives. In cooperation between the business computing group, especially CIM, a company which produces seat components, seat structures and services, and the International Graduate School in Paderborn, concepts and methods are being developed to support the planning and control of production.

To conclude, this paper describes how OR methods can contribute to better and faster planning. As the manufacturer faces with a complex planning problem good automatically generated plans can improve the competitive advantage of a company due to i.e. better capacity utilization. The required extensions to the DLSP for sequence dependent setup times, alternative machines, batch-orientated and joint production could also be transferred to other planning situations. A decomposition of the problem limits the computing time and allows a prioritization of the conditions and targets. This approach could also be transferred to other planning situations in other companies.

The following main part contains a problem description, a discussion of relevant literature, a mathematical mixed integer programming model and first computational results. The conclusion summarizes the results and gives an outlook on further research which has to be done.

Problem and Concept

The following first subsection describes the business environment and the problem which has to be solved. After that, available approaches are discussed in the literature section. Then, the concept how to solve the problem is explained and the extensions to the DLSP are given in details. Lastly, first computational results are presented.

Problem description

The production process of the analyzed automotive supplier plant can roughly be subdivided into four main parts. In front of the stamping process, which is the first main part of the production, there is a raw material stock. The raw materials which are needed for the stamping process are steel coils which differ in their material degrees, dimensions and composition. Their weights vary by a medium of about ten tons.

Several parallel stamping machines process the steel coils. Depending on the part to be produced the corresponding steel coil consisting of the required raw material is chosen and transported from raw material stock to the stamping machine. After that, the steel coil has to be put up at the stamping machine. Each stamping machine is able to use different dies which are needed to produce the parts. Some of these dies are dedicated to punch out only one part. Others produce a number of parts in combined production simultaneously. With some of the available dies the same part can be produced twice, other dies are able to produce a part and its mirror-inverted counterpart at once, for example.

The punched parts are then filled into different types of loading equipment. The loading equipment used depends on the parts and on the intended processing steps of the parts within the loading equipment. The parts are then booked into the ERP system and can then be transported to the intended processing step. Some parts are directly transported to the shipment zone where they are shipped to other plants or to external customers. But most parts are hardened, washed and polished before they are assembled or shipped to the customer. The complex material flow between hardening, polishing, washing, etc is not being analyzed in this paper and seen as a black-box. After this black-box, the parts are assembled in assembly lines. Then, the assembled parts are transported to the shipment from where they are distributed to the customers.

In cooperation with the University of Paderborn, the production process was analyzed in order to find room for improvement. Capital intensive stamping machines can be regarded as the bottleneck at the beginning of the production process and therefore have high influence of the following production steps. Therefore, it is very important to optimize production in this stage. The first step to improve the production process is to improve planning. To guarantee the feasibility of the generated planning solutions, it is of crucial importance that all conditions arising in practice are integrated in the model. In practice, there are many conditions that could vary in the course of time, for example due to work shifts or with the different days of the week.

Summing it all up, the challenge is to generate feasible and good plans for the stamping production stage of a supplier in the automotive industry. Several practical conditions have to be taken into account in order to follow the most important targets at preferably low production costs.

Literature Review

Lot sizing is always important, if there exist setup times or setup costs between the production of two different products. The properties of the production segment define the type of the lot sizing problem. In this section, different lot sizing definitions are analyzed and categorized. Single stage lot sizing problems will come to the fore in the following subsections. The first subsection is dedicated to static lot sizing problems where static demands are assumed. The second subsection lists and explains the assumptions of other available models which consider dynamic demand. The latter section is further subdivided into big bucket and small bucket problems.

Static lot sizing problems

One of the first attempts to calculate optimal lot sizes was done 80 years ago by Andler [Andler (1929)]. His square root formula is often used in practice although most of its economic assumptions restrict the applicability too much. Continuous demand and capacities without limitations often lead to infeasible production plans. Especially the consideration of only one product limits the suitability of Andler's lot sizing formula in practice.

Another static lot sizing problem is known as the Economic Lot Scheduling Problem (ELSP) [Günther et al (2005)] which is able to define lot sizes as well as lot schedules simultaneously. Although feasible solutions can be generated because of the consideration of several products, the ELSP is only applicable in situations where the demand rate for each of the product which is planned is constant over time.

Dynamic Lot Sizing Problems

A first model which considers dynamic demand was formulated in [Wagner et al (1958)]. This Single-Level Uncapacitated Lot Sizing Problem (SLULSP), which is also known as the Wagner-Whitin problem plans only one single product and it does not consider capacities. This is why generated plans are often not practicable. Nevertheless, many methods like the silver-meal heuristic [Silver (1973)], the Groff heuristic [Groff (1979)] or the part-period-procedure [DeMatteis (1968)] were developed in the past and often provide a basis for i.e. material requirement planning in enterprise resource planning systems or production planning and control systems.

There were developed several other models which consider dynamic demand. These can be subdivided into big bucket, small bucket and hybrid models. Within big bucket models macro-periods are defined. Capacities are considered for each of these big buckets. Several lots can be planned within one big bucket.

Small bucket models make use of micro-periods. Capacities of resources can then better be utilized as the planning is more accurate. Small bucket models allow simultaneous planning of lot sizes and their schedules.

Big bucket models define lot sizes and assign them to macro periods, which are long enough so that multiple lots can be produced within one of such periods. As multiple lots can be assigned to one period, the sequence of the lots within one macro period is not defined.

Examples for big bucket models are the Continuous Lot-Sizing Problem (CLSP) [Salomon (1991)] or its extensions like the Continuous Lot-Sizing Problem with Linked Lot-Sizes (CLSP-L) [Haase (1994)]. The basic CLSP extends the SLULSP with the consideration of multiple products and capacity limits. More than one resource can be considered to limit available capacity in one period. The CLSP-L is an extension of the CLSP so that the lot sizes can be linked between two periods. The problem was extended so that setup states can be transferred from one period to another. This model is known as the CLSP with setup carryover (CLSP-SC) [Haase (1996)]. The CLSP-L as well as the CLSP-SC allow to plan and to define the production at the periphery of the periods. The lot sequence within the periods is not defined.

All in all, it can be said that big bucket models are useful for specific situations. Big bucket models are suitable especially for long term planning decisions or planning situations where capacity utilization within the macro periods is negligible or flexibility within the macro periods is desired. A detailed lot sizing and scheduling is only possible if short periods are defined. Therefore small bucket models were defined.

Small bucket models define lot sizes and lot sequences and are therefore able to generate detailed plans. The accuracy depends on the length of the used micro periods. The Discrete Lot Sizing and Scheduling Problem (DLSP) [Fleischmann (1990)] acts on the assumptions that each period can only be reserved by one product. Hence, during one period it is only possible to produce either nothing or at full capacity. The so called all-or-nothing assumption guarantees on the one hand that production capacity is fully utilized but on the other hand it removes any flexibility to change lots or sequences within the micro periods. The Continuous Setup Lot Sizing Problem (CSLP) [Karmarkar et al (1985)] does not lose setup states in idle periods. Another example for a small bucket problem is the Proportional Lot Sizing and Scheduling Problem (PLSP) [Haase (1994)]. The PLSP allows doing one single product change per period which is its basic assumption. The setup state can be transferred from one period to another. Extensions to the PLSP were described in [Sürie (2005)]. Extensions for setup times, batch sizing and lot size restrictions are introduced and explained.

Another extension on the basis of the DLSP is the General Lot-Sizing and Scheduling Problem (GLSP) [Fleischmann (1997)]. With the GLSP it was intended to create a generalized model which summarizes formerly developed models respecting aspects like capacity limitations or sequence dependent setup costs. Another extension is the fusion of the big bucket and small bucket concept in one model.

To conclude, it can be said that there already exist several extensions of models for different situations. But still, there are several aspects missing especially if a practical case has to be considered. Existing model extensions can be used as a good basis. The following section describes the concept of the method which is currently developed. Model extensions are described in detail.

Concept

Lot sizing and scheduling problems are very hard to solve [Salomon et. al. (1991)]. Especially if a great number of practical aspects have to be considered, problem complexity grows so much that practical problems cannot be solved. Therefore a concept is necessary which is able to reduce the difficulty to find feasible solutions for practical problems. A decomposition approach is presented in the following subsection. After that a mixed integer programming model is described.

Decomposition Approach

As the consideration of all practical aspects at time is not possible due to complexity of the resulting problem, decomposition is essential. There already exist decompositions over time line [Sürie (2005)]. In this paper another decomposition approach of the problem is presented which limits the computing time and allows a prioritization of the conditions. The method that is currently being developed considers the most important aspects in an optimization method, while less important aspects are considered within certain tolerance regions in an ensuing improvement procedure. Besides reducing the computing time, this method prevents the discarding of good solutions because of less important constraints and targets. The prioritization of targets and constraints is as follows:

1. Fulfilment of customer demand
2. Maximization of machine utilization
3. Maximization of worker utilization
4. Maximum for lot sizes
5. Batch oriented production

The fulfillment of demand is essential for the company as it can only survive if the customers always are comfortable with the delivery.Secondly, it is important to keep machines running as long as possible, but not more than necessary. The machine utilization is therefore kept high and as production only runs if there is customer demand, stocks of inventory are also limited.Although the third target has its focus on the maximization of the worker utilization, it is also related to the machine utilization. The maximization of worker utilization which is related to minimization of workers reserved, leads also to better utilization of the machines as idle times can be reduced. If we need specialized, highly skilled cost intensive workers for setting up the machines, it is better to minimize the reservation of such personal. As follows, minimization of simultaneous setups reduces the number of costly setup teams needed.Next, it is important to consider lot size limits. The dies used for punching out the parts from the steel coils fray during the process. If lot sizes exceed the limits of the dies, the quality of the produced parts cannot be guaranteed any more.Last but not least, a batch oriented production has to be considered. As stated before, the main raw material for the stamping process is stored in the form of heavy steel coils. Workers are exposed to danger if they have to set down broached, partly depleted coils. Furthermore, stored, partly depleted coils have to be scrapped occasionally as they cannot be set up any more.

There exist many more but less important aspects in practice which should be considered. One example is limitation of capital commitment which actually only provides a basic orientation for planners to limit lot.

To conclude, the decomposition approach enables, that only the five most important production aspects are considered in computationally intensive optimization procedure, while others are regarded in an ensuing improvement procedure. The model which can be used to calculate solutions for the five most important aspects is described in the next subsection.

Notation

Sets	
P	Products
CP_p	Subset of products which are simultaneously produced with $p \in P$ in coupled production
M	Machines
T	Micro Periods

Parameters	
Cost parameters	
c_p^{inv}	Inventory holding cost for product $p \in P$ per period
$c_{m,p,q,t}^{setup}$	Setup costs for a changeover at machine $m \in M$ from product $p \in P$ to product $q \in P$ in period $t \in T$
$c_{m,p,t}^{cc}$	Costs for changing a coil at machine $m \in M$ for product $p \in P$ in period $t \in T$
$c_{m,p,t}^{prod}$	Production costs at machine $m \in M$ to produce product $p \in P$ in period $t \in T$
c_t^{team}	Costs for an additional setup team
Inventory parameters	
$I_{p,0}$	Initial inventory of $p \in P$
$I_{p,n}$	Final storage of product $p \in P, n = T_{max}$
Production parameters	
$cycle_{p,m}$	Cycle time of product $p \in P$ at machine $m \in M$
$ppc_{p,m}$ $\left(= \left\lfloor \frac{Length\ of\ Period}{cycle_{p,m}} \right\rfloor \right)$	Amount of product $p \in P$ which can be produced in one period with machine $m \in M$ (parts per cycle)
$d_{p,t}$	Demand of product $p \in P$ in $t \in T$
bs_p	Batch-Size of product $p \in P$
$minlot_p$	Minimum Lot-Size of product $p \in P$
$maxlot_p$	Maximum Lot-Size of product $p \in P$
$setupTime_{p,q}$	Setup time for a changeover from product $p \in P$ to product $q \in P$

Variables	
$r_{m,p,q,t} \in \{0,1\}$	=1, if changeover is currently done at machine $m \in M$ from product $p \in P$ to product $q \in P$ during $t \in T$: $p \neq q$ (0, otherwise)
$s_{m,p,t} \in \{0,1\}$	=1, if machine $m \in M$ is currently setup product $p \in P$ in $t \in T$ (setup state variable) (0, otherwise)
$prod_{m,p,t} \in \{0,1\}$	=1, if machine $m \in M$ is currently producing product $p \in P$ in $t \in T$ (0, otherwise)
$x_{m,p,t} \in N^+$	Production output of product $p \in P$ at machine $m \in M$ in $t \in T$ ($prod_{m,p,t} * ppc_{p,m}$)
$cc_{m,t} \in \{0,1\}$	=1, if a coil exchange has to be done at machine $m \in M$ in $t \in T$
$I_{p,t} \in N^+$	Inventory of product $p \in P$ in $t \in T$
$L_{m,p,t} \in N^+$	Lot variable which cumulates production output of product of passed production periods at machine $m \in M$ until $t \in T$
$ml_{m,p,t} \in \{0,1\}$	=1, if minimum lot size of product $p \in P$ is produced at machine $m \in M$ until $t \in T$ (0, otherwise)
$B_{m,p,t} \in N^+$	Amount of completed batches of product $p \in P$ at machine $m \in M$ until $t \in T$
$Sl_{m,p,t} \in N^+$	Slack variable, residual quantity of the last batch of $p \in P$ which is not finished until $t \in T$ at machine $m \in M$
$mst_{m,p,q,t} \in \{0,1\}$	=1, if the required setup time for a changeover from product $p \in P$ to product $q \in P$ at machine $m \in M$ was reached in $t \in T$ (0, otherwise)
$0 \le cs_{m,p,q,t} \le 1$	Slack variable, needed to store the percentage of time which is needed to complete setup from product $p \in P$ to product $q \in P$ at machine $m \in M$ in $t \in T$
$bin^{prod}_{m,t} \in \{0,1\}$	=1, if machine $m \in M$ is running and producing in $t \in T$ (0, otherwise)
$bin^{r}_{m,t} \in \{0,1\}$	=1, if machine $m \in M$ is being setup in $t \in T$ (0, otherwise)
$bin^{cc}_{m,t} \in \{0,1\}$	=1, if a coil exchanged at machine $m \in M$ in $t \in T$ (0, otherwise)
$bin^{idle}_{m,t} \in \{0,1\}$	=1, if machine $m \in M$ is idle in $t \in T$ (0, otherwise)
$teams_t \in N^+$	Amount of required teams to do setups in $t \in T$

Mathematical Model

The mathematical formulation of the practical problem begins with the objective function which is presented below:

Objective Function	
minimize	Minimize the costs for
$\sum_{t \in T} teams_t * c_t^{team}$	+additional teams which are dedicated to setup the machines
$+\sum_{m \in M}\sum_{p \in P}\sum_{q \in Q}\sum_{t \in T} r_{m,p,q,t} * c_{m,p,q,t}^{setup}$	+ setups
$+\sum_{m \in M}\sum_{p \in P}\sum_{t \in T} cc_{m,p,t} * c_{m,p}^{cc}$	+ coil exchanges
$+\sum_{m \in M}\sum_{p \in P}\sum_{t \in T} prod_{m,p,t} * c_{m,p}^{prod}$	+ production
$+\sum_{p \in P}\sum_{t \in T} I_{p,t} * c_{m,p}^{inv}$	+ inventory holding costs

The objective is to minimize the sum of the costs for additional teams needed for parallel setups at different machines, setup costs depending on the machine, involved parts and period depending costs, coil exchange costs as well as production costs which depend on the machine, the product and the period and, last but not least, inventory holding costs, which only depend on the product which has to be stored.The following equations describe the production system and guarantee that only practicable solutions are generated.

Subject to:

$$I_{p,t-1} + \sum_{m \in M} x_{m,p,t} = d_{p,t} + I_{p,t} \quad \forall p \in P, t \in T \tag{1}$$

$$ppc_{p,m} * prod_{m,p,t} = x_{m,p,t} \quad \forall m \in M, t \in T, p \in P \tag{2}$$

$$prod_{m,p,t} - s_{m,p,t} \leq 0 \quad \forall m \in M, p \in P, t \in T \tag{3}$$

The inventory balance constraint (1) ensures that the quantity stocked or sold is either produced or removed from stock without permitting backlogging. Restriction (2) is used to calculate the output and to store the output in variable $x_{m,p,t}$. The output depends on the binary variable $prod_{m,p,t}$ and the production speed $ppc_{p,m}$. Production is only possible if the machine m has the required setup state (3). In the actual practical case, there exist dies which are able to produce multiple products simultaneously. In order to represent this, the following restrictions are necessary:

$$\sum_{p \in P} s_{m,p,t} \frac{1}{\# CP_p} = 1 \quad \forall m \in, t \in T \tag{4}$$

$$s_{m,p,t} - s_{m,q,t} = 0 \quad \forall m \in M, t \in T, p, q \in P : q \in CP_p \tag{5}$$

$$prod_{m,p,t} - prod_{m,q,t} = 0 \quad \forall m \in M, t \in T, p, q \in P : q \in CP_p \tag{6}$$

Equation (4) guarantees that the state of a machine is well-defined in each period. As some dies produce multiple products in coupled production, the state for all products have to be set to the same value (5). Not only the coupled setup state but also the production has to be coupled (6).

Coupled production also has to be considered during setups. Therefore, restriction (7) was introduced. Constraint (8) provides that the setup state of a machine is only changed if the required minimum setup time has been reached. Moreover, it guarantees, that no other setup has been finished after the same period t.

$$\begin{aligned} & s_{m,p,t-1} + s_{m,q,t} - r_{m,p,q,t} \leq 1 \\ & \forall p, q \in P : q \notin CP_p \\ & \forall m \in M, t \in T \end{aligned} \tag{7}$$

$$\begin{aligned} & r_{m,p',q,t} = r_{m,p'',q,t} \\ & \forall p', p'', q \in P : p' \in CP_{p''} \\ & \forall m \in M, t \in T \end{aligned} \tag{8}$$

$$\begin{aligned} & r_{m,p,q',t} = r_{m,p,q'',t} \\ & \forall p, q', q'' \in P : q' \in CP_{q''} \\ & \forall m \in M, t \in T \end{aligned} \tag{9}$$

Conditions (8) and (9) assure that if a setup is necessary for a product all setup variables related to coupled products have to be activated.

$$\begin{aligned} & s_{m,q',t} \leq 1 - \sum_{\substack{p \in P \\ p \neq q' \\ p \neq q''}} \sum_{\substack{q'' \in P \\ q'' \neq q' \\ q' \notin C_{q''}}} \frac{1}{\#CP_p * \#CP_{q''}} * mst_{m,p,q'',t} \\ & \forall m \in M, t \in T, p, q', q'' \in P : p \notin C_{q''} \end{aligned} \tag{10}$$

The following illustrations show graphically how the setup is modeled and which cases are excluded:

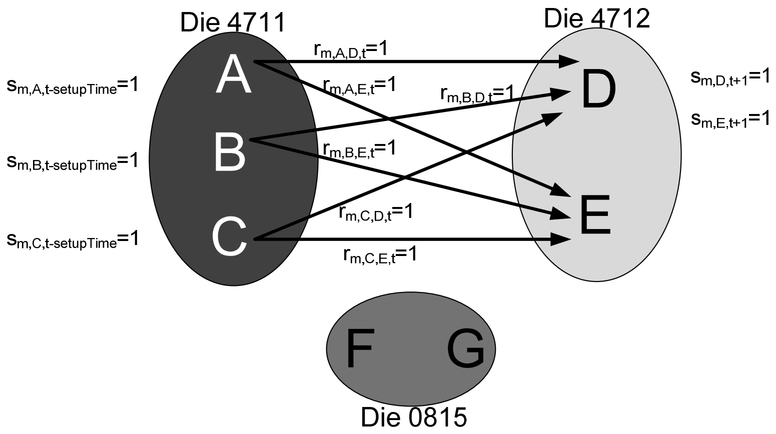

Figure 1: Restrictions ensure the generation of feasible setups

In this illustration, the standard case is shown. The machine is setup from die 4711 to die 4712 and the setup states are set correctly. Restriction (10) excludes the case that the required minimal setup time can be reached for two dies at the same time. This would represent a solution which is not feasible in practice:

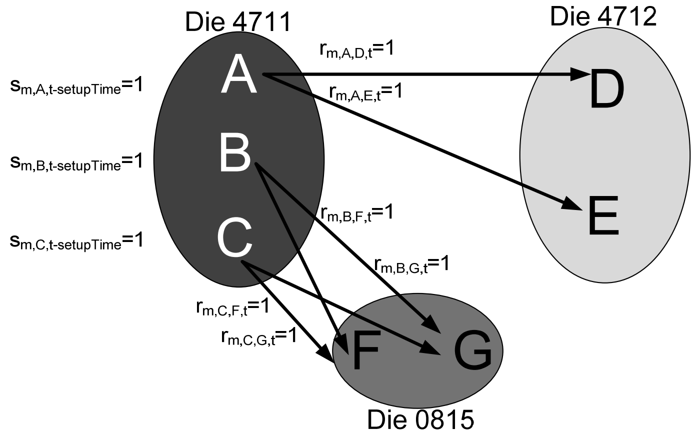

Figure 2: Avoidance of infeasible setups

Sequence dependent setup times also have to be modeled as they have great influence on optimality of the lot sequence. Restrictions (11) and (12) are used to do that. Inequality (11) assures that the variable $cs_{m,p,q,t}$ which represents the percentage of the required setup time (12) never exceeds 1 and it guarantees that $cs_{m,p,q,t}$ is only activated if a setup is being processed and is not finished, yet, which is indicated by $r_{m,p,q,t}$. In (12) the value of $cs_{m,p,q,t}$ is set to 1 if the setup is finished and $mst_{m,p,q,t}$ is activated. Compare [Sürie (2005)] for a sequence independent case.

$$cs_{m,p,q,t} \leq r_{m,p,q,t+1} \quad \forall m \in M, t \in T, p', p'', q \in P : p' \notin C_{p''} \tag{11}$$

$$cs_{m,p,q,t-1} + r_{m,p,q,t} \frac{Period\ Length}{setupTime_{p,q}} = mst_{m,p,q,t} + cs_{m,p,q,t} \quad (12)$$

$$\forall m \in M, t \in T, p', p'', q \in P : p' \notin C_{p''}$$

In order to prevent the case that variables for production are activated in the same period as setup variables, it is necessary to introduce further restrictions. It is important to consider this case as solutions would be infeasible due to the fact that it is not possible to run a machine when it is being setup. Therefore, (13) is used.

$$bin^{prod}_{m,t} + bin^{r}_{m,t} + bin^{cc}_{m,t} + bin^{idle}_{m,t} = 1 \quad \forall m \in, t \in T \quad (13)$$

$$r_{m,p,q,t} \leq bin^{r}_{m,t} \quad \forall m \in M, t \in T, p, q \in P : p \notin C_q \quad (14)$$

$$\sum_{p \in P} \sum_{\substack{q \in P \\ q \neq p \\ p \notin C_q}} r_{m,p,q,t} \geq bin^{r}_{m,t} \quad \forall m \in M, t \in T \quad (15)$$

$$cc_{m,p,t} \leq bin^{cc}_{m,t} \quad \forall m \in M, t \in T, p \in P \quad (16)$$

$$\sum_{p \in P} cc_{m,p,t} \geq bin^{cc}_{m,t} \quad \forall m \in M, t \in T \quad (17)$$

$$prod_{m,p,t} \leq bin^{prod}_{m,t} \quad \forall m \in M, t \in T, p \in P \quad (18)$$

$$\sum_{p \in P} prod_{m,p,t} \geq bin^{prod}_{m,t} \quad \forall m \in M, t \in T \quad (19)$$

Restrictions (14-19) control binary variables $bin^{prod}_{m,t}, bin^{cc}_{m,t}, bin^{r}_{m,t}$. These variables are indicator variables for the activity of a machine in a period. With inequality (20) the number of required teams for setups is defined and influences directly the objective function.

$$\sum_{m \in M} bin^{r}_{m,t} \leq teams_t \quad \forall t \in T \quad (20)$$

$$L_{m,p,t} - L_{m,p,t-1} - x_{m,p,t} \leq 0 \quad \forall m \in M, t \in T, p \in P \quad (21)$$

$$L_{m,p,t} - L_{m,p,t-1} - x_{m,p,t} \leq bin^{r}_{m,t+1} * maxlot_p \quad \forall m \in M, t \in T, p \in P \quad (22)$$

If $minlot_p > bs_p$ then (23)

$$L_{m,p,t-1} + x_{m,p,t} \geq \sum_{\substack{q \in P \\ p \neq q}} \frac{minlot_p}{\#CP_q} r_{m,p,q,t} \quad \forall m \in M, t \in T, p \in P$$

$$L_{m,p,t-1} + x_{m,p,t} \geq \sum_{\substack{q \in P \\ p \neq q}} \frac{bs_p}{\#CP_q} r_{m,p,q,t} \quad \forall m \in M, t \in T, p \in P \text{, otherwise}$$

The lot size limits are controlled by restrictions (21-23). The variable $L_{m,p,t}$ saves the amount of parts which are produced in the current lot (21). Valid inequality (22) ensures that the lot size never exceeds its maximum. The maximum is only active if no changeover is done in period t. If a changeover is necessary in period t, the lot size is reset to 0. Althogh this fact is already guaranteed by equation (21), it reduces solution space in a way that results are computed with less iterations in shorter time. In (23), it is guaranteed that the minimal lot size is produced. If the related batch size is greater than the minimum lot size, the batch size is taken as a substitute. The lot size variable $L_{m,p,t}$ can also be used to model batch size restrictions.

$$L_{m,p,t-1} + x_{m,p,t} = bs_p * B_{m,p,t} + Sl_{m,p,t} \quad \forall m \in M, t \in T, p \in P \tag{24}$$

$$Sl_{m,p,t} \leq bs_p * prod_{m,p,t} \quad \forall m \in M, t \in T, p \in P \tag{25}$$

$$cc_{m,p,t} + B_{m,p,t-1} = B_{m,p,t} \quad \forall m \in M, t \in T, p \in P \tag{26}$$

Restrictions (24) and (25) ensure that only complete batches are produced [Sürie (2005)]. Variable $B_{m,p,t}$ counts the number of complete batches produced until period t. The slack-variable $Sl_{m,p,t}$ contains the remaining quantity to complete the next batch and is reset to 0 if production ends (25). The indicator variable $cc_{m,p,t}$ for changing the coils is activated if the number of complete batches increases.

First Computational Results

Several experiments were done in order to test functionality of the optimization procedure. All tests were executed with a 1.66 GHz Dual Core and 1 GB RAM using single threaded instance of ILOG CPLEX 10. Results, which are on average 9% better than manually created plans, were computed in less than 20 minutes. One example is presented and analyzed in this section.

The part set contains 20 parts. Three machines have to be scheduled simultaneously. 70 time periods are considered in this example. There are no parts which can be manufactured in joint production. The detailed costs will not be presented in this section. It only can be said that the costs for reserving an additional setup team are the highest. About 10 times less, setup costs per period are taken into consideration, followed by costs for changing a coil, production costs and also idle costs. The inventory holding costs per period for the amount of parts which can be produced in one period are about 10 times less than setup costs per period. In this case, the batch size is for all products the same and amount to 3000 parts per steel coil. For simplification reasons, it is supposed that the production speed is equal for all products and is 500 parts per period. The demand data is as follows:

Part	Demand	Period
1	7000	20
1	3000	44
2	5000	15
3	4000	37
3	2000	50
4	1000	31
4	3500	55
5	14000	55
6	2000	20
6	5000	58

Table 1: Demands for Experiment

The resulting Gantt chart looks like this:

Timeslice	1	2	3	4	5	6	7	8	9	10	11
Machine A		Setup	Setup	Setup	Setup						
Machine B				Part 2	Part 2	Part 2	Part 2	Part 2	Part 2	Coil	Part 2
Machine C					Part 1	Part 1	Part 1	Part 1	Part 1	Part 1	Coil

Timeslice	12	13	14	15	16	17	18	19	20	21	22
Machine A				Part 6	Part 6	Part 6	Part 6	Part 6	Part 6	Coil	Setup
Machine B	Part 2	Part 2	Part 2	Part 2	Part 2	Coil	Setup	Setup	Setup	Setup	
Machine C	Part 1	Part 1	Part 1	Part 1	Part 1	Part 1	Coil	Part 1	Part 1	Part 1	Part 1

Timeslice	23	24	25	26	27	28	29	30	31	32	33
Machine A	Setup	Setup	Setup				Part 3	Part 3	Part 3	Part 3	Part 3
Machine B		Part 5	Part 5	Part 5	Part 5	Part 5	Part 5	Coil	Part 5	Part 5	Part 5
Machine C	Part 1	Part 1	Coil	Setup	Setup	Setup	Setup	Part 4	Part 4	Part 4	Part 4

Timeslice	34	35	36	37	38	39	40	41	42	43
Machine A	Part 3	Coil	Part 3	Part 3	Part 3	Part 3	Part 3	Part 3	Coil	Setup
Machine B	Part 5	Part 5	Part 5	Coil	Part 5	Part 5	Part 5	Part 5	Part 5	Part 5
Machine C	Part 4	Part 4	Coil	Setup	Setup	Setup	Setup		Part 1	Part 1

Timeslice	44	45	46	47	48	49	50	51	52	53
Machine A	Setup	Setup	Setup				Part 6	Part 6	Part 6	Part 6
Machine B	Coil	Part 5	Part 5	Part 5	Part 5	Part 5	Part 5	Coil	Part 5	Part 5
Machine C	Part 1	Part 1	Part 1	Part 1	Coil	Setup	Setup	Setup	Setup	Part 4

Timeslice	54	55	56	57	58	59	60	61	62	63
Machine A	Part 6	Part 6	Coil	Part 6	Part 6	Part 6	Part 6	Part 6	Part 6	Coil
Machine B	Part 5	Part 5	Part 5	Part 5	Coil					
Machine C	Part 4	Part 4	Part 4	Part 4	Part 4	Coil				

Figure 3: Resulting gantt chart

It can be seen that no simultaneous setups at different machines were necessary so that higher costs were prevented. The batch sizes were considered always. Therefore coils were always completely used and no rests were stored. Resulting costs from production, setups or inventory holding were minimized.

First simple experiments were done to compare costs of manual plans, which were generated in real life, with costs of generated plans in order to have a first evaluation of the automatic method. Customer demands were presented to production planners and they had to generate a feasible plan on the basis of given data and personal knowledge. Same demands and same data are used to calculate plans using the presented method. The resulting costs of the alternative plans are compared in the following table:

Exp.	Costs: Manual Plan	Costs: Generated Plan	Improvement
1	12762	11257	-11.8%
2	10537	9932	-5.7%
3	13820	10698	-22.6%
4	9502	9469	-0.3%
5	10367	9845	-5.0%
6	11051	10927	-1.1%
7	9073	8319	-8.3%
8	13869	11325	-18.3%
9	12578	10792	-14.2%
10	9452	9334	-1.2%
			Average: -8.9%

Table 2: Comparison between manual and generated plans

The generated plans are always better than the manually created plans. The average improvement is about 9%.

Further model enhancements with valid inequalities or defined cuts could improve performance. Therefore, further research is necessary in this direction. Another performance improvement could be provided by generating starting solutions with constraint programming techniques or heuristics.

Conclusion

The developed decomposition approach is useful for practice as complexity is reduced and feasible solutions can be computed in short time. The presented model combines approaches from other authors and implements several new aspects which have great importance in practice. Examples are the consideration of sequence dependent setup times and costs as well as maximum and minimum lot sizes and the consideration of a batch oriented production. But still further research has to be done. A starting heuristic could be very useful for difficult problem instances. Constraint programming techniques are currently being analyzed and tested in order to calculate an upper bound for the optimization procedure. Further research has to be done on valid inequalities or cuts for the model to improve its solution performance for practical problems. Moreover a meta-heuristic has to be developed which is dedicated to generate further solutions which consider less important aspects. In order to reduce development effort, the heuristic could also be based either on mixed integer programming or constraint programming techniques.

References

[1] Andler, K. (1929), Rationalisierung der Fabrikation und optimale Losgrösse. R. Oldenbourg, München

[2] DeMatteis, J.J. (1968): An Economic Lot-Sizing Technique—The Part Period Algorithm. IBM Systems Journal Vol. **7** No. 7, pp. 30–38.

[3] Fleischmann, B. (1990), The discrete lot-sizing and scheduling problem. European Journal of Operations Research. 44, pp. 337–348.

[4] Fleischmann, B.; Meyr H. (1997): The general lotsizing and scheduling problem. OR Spektrum, 19:11-21.

[5] Groff, G.K. (1979): A Lot-Sizing Rule for Time-Phased Component Demand. Production and Inventory Management No.1, pp. 47-53.

[6] Günther, H.-O. and H. Tempelmeier (2005), Produktion und Logistik (5. edition). Springer, Berlin.

[7] Haase, K. (1994): Lotsizing and scheduling for production planning. Lecture Notes in Economics and Mathematical Systems. 408. Springer, Berlin.

[8] Haase, K. (1996): Capacitated lot-sizing with sequence dependent setup costs, OR Spektrum 18, pp. 51–59.

[9] Karmarkar, U.S. and Schrage L. (1985): The deterministic dynamic product cycling problem. Operations Research, 33, pp 326-245.

[10] Nahmias, S. (2001). Production and Operations Analysis (4. edition). Irwin, Homewood, Illinois.

[11] Salomon, M. (1991): Deterministic Lotsizing Models for Production Planning. Springer, Berlin.

[12] Salomon, M., Kroon, L. G., Kuik, R. and Van Wassenhove, L. N. (1991): Some extensions of the discrete lotsizing and scheduling problem, Management Sci. 37, 801–812

[13] Silver, E.A. and Meal, H.C. (1973): A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment. Prod. Inventory Management 14(2), 64-74.

[14] Sürie, C. (2005): Time Continuity in Discrete Time Models - New Approaches for Production Planning in Process Industries. Springer, Berlin.

[15] Wagner, Harvey M. and Whitin, Thomson M. (1958): Dynamic Version of the Economic Lot Size Model. In: Management Science 51, 89-96.

ASOR Bulletin
Volume 29, Number 3, September 2010, pp.34-47
ISSN0812-860X | eISSN1446-6678

Lean manufacturing strategy in a make to order manufacturing environment

M. A. Karim, M. R. Alam and M. A. Amin
School of Engineering Systems, QUT, Brisbane, Australia
Noja Power Limited, Brisbane, Australia
School of Engineering Systems, QUT, Brisbane, Australia
(Email: Azharul.karim@qut.edu.au; rabiul86@yahoo.com; mdal.amin@qut.edu.au)

Abstract
Lean Manufacturing principles and techniques have been applied in a wide variety of organizations inside and outside of manufacturing industries. However, this approach has been applied more frequently in large and continuous manufacturing process than small and make to order (MTO) type manufacturing environment, mainly because of several perceived barriers in the later environment that have caused managers to be reluctant to make the required commitment. Achieving lean production status requires dramatic changes in all areas of the organization, in particular in product design, inbound logistics, manufacturing processes and outbound logistics. In this study we describe a case where lean principles were adapted in a small MTO type power equipment manufacturer. A lean manufacturing strategy was first developed for this particular company. Common lean tools such as time study, 5S, two-bin kanban and manufacturing cell and line design were applied in the proposed lean strategy. Finally a rapid performance management (RPM) technique was proposed for the effective implementation of the strategy. Implementation of the lean strategy enabled the company to achieve smooth production flow, reduced production lead-time, lower work-in-process inventory, higher productivity and higher product quality. Current progress of the research indicates that the proposed methodology would be able to make a significant contribution in similar other manufacturing organisations as well in the adoption of lean manufacturing concepts in their organizations.

Keywords: Lean manufacturing, make to order manufacturing, Time study, Kanban, Rapid performance management

Introduction

Lean manufacturing is a production strategy for organizational effectiveness focusing on waste reduction and improving productivity through application of various tools and techniques. At its core, lean manufacturing is a means by which the overall business processes are organized so as to deliver products with greater variety and superior quality using less resource and in a shorter time than can be achieved by mass production methods (Oliver et al., 1994). The goal of lean manufacturing is to reduce the waste in human effort, inventory, time to market and manufacturing space to become highly responsive to customer demand (Riezebos and Klingenberg 2009). It is now widely recognized that organizations that have mastered lean manufacturing methods have substantial cost and quality advantages (Fleischer and Liker 1997).

Today lean production has become the goal of manufacturers aiming for world-class status. There have been many reports of the successful introduction of lean principles by organizations and at the same time there have been many reports of failure (Sohal and Egglestone 1994). Samson et al., (1993) and Dawson and Palmer (1995) described the successful adoption of a variety of lean production programmes in Australian industries, while Sohal et al., (1994) provided evidence of failures where improvement initiatives "faded away" or "simply died" after a few years. Large manufacturers are more likely to implement lean practices than small manufacturers (Shah and Ward 2003).

Irrespective of how it is perceived, the concept of lean manufacturing has unarguably been discussed extensively in the past decade or so. However, there appears to be little empirical evidence in publications on the implementation of lean practices and the factors that might influence them in small and medium size enterprises (SME) (Bruun and Mefford 2004). With the notable exception, most of these publications have tended to focus on the premise of large enterprises only (Achanga et al., 2006, Bozdogan et al., 2000). Similarly, the technique is not widely practiced in MTO industries. MTO industries are still not certain of the cost of its implementation and likely tangible and intangible benefits they may achieve. Most of these companies fear that implementing lean manufacturing is costly and time consuming. Hence studies on lean manufacturing practices in MTO industries are limited.

The purpose of this research is to investigate opportunities of lean practices for MTO situations, develop a lean method and implement the method in MTO manufacturing organisation.

Lean Manufacturing and MTO Industries

The following characteristics of MTO manufacturing were depicted by Hendry and Kingsman (1993) and Karim et al. (2010):

1. There are not many standard products.
2. Resources are multi-task machinery and flexible workforce—a large number of possible end items within a product line.
3. A small volume and high degree of demand variability both in quantity and product mix.
4. Product lead time is vital for customer satisfaction and agreed with customers.
5. Price is agreed with customers before production commences.

From the characteristics of MTO described above, it can be assumed that the manufacturing process in MTO manufacturers is more complex compared to others. Often the volumes are low even at the component production stage as there is little scope for common components because products are manufactured to customer design and specification. Many of these MTO companies are small and Medium sized Enterprises (Stevenson et al., 2005). Examples of the MTOs in Australia where lean manufacturing can be implemented include automotive parts manufacturers, biomedical industries, power equipment industries, and die and mould making industries.

There has been much debate as to whether formal lean process and quality enhancement approaches can be effectively implemented and subsequently utilised by MTOs. Thomas and Webb (2003), in their work on analysing process improvement systems implementation in small and medium sized industries, highlight the lack of intellectual and financial capacity within small companies as being the primary issues that lead to poor systems implementation. They state that the uniqueness and complexity of small manufacturing operations often hinder the implementation process. Husband and Mandal (1997) identified similar problems with Australian SMEs and identified a series of issues that affect the ability of SMEs to implement formalised quality enhancement techniques. Small companies also lack resources in the form of time and personnel (Deleryd et al., 1999). In addition, they also have limited resources to provide internal training.

In many MTO companies, it is the customer that has responsibility for the design of the products. Therefore, in particular, there is less potential for the MTO companies to learn from its mistakes since they are continually in the process of making new products designed by various customers.

Background of the Case Organization

The case organisation, Power Pty Ltd* (PPL), is a manufacturer of low, medium and high voltage switchgear products. The Company is specialized in medium and high voltage auto reclosers for both pole mounted and substation applications from 10kV to 38kV. Their low voltage switchgear division is specialized in the design and manufacture of high quality low voltage motor control centers and low voltage switchgear assemblies. The design meets the demanding requirements of the mining, power utilities and process control industries. The design and development of an auto recloser depends on customers' specifications and requirements, and the regulatory legislation of a particular country.

Based on the customer requirements and specifications, product is designed and production order is being created. Variations in design and process occur due to the customer requirements and specifications. Therefore, according to the characteristics described earlier, the company is considered as MTO manufacturer. In order to stay competitive, the company is keen to embrace lean manufacturing strategy to improve productivity, on-time delivery and quality.

Current Manufacturing Practices in the Case Organization

The auto recloser product consists of OSM tank, RC control cubicle, control cable, and other electronic modules. Currently, the company has four main manufacturing lines which are electrical control and communication cubicle assembly line, OSM automatic circuit reclosers' assembly line, cable making line and switchgear assembly line. In this paper electrical control and communication cubicle manufacturing process was taken as a case study for implementing lean manufacturing philosophy.

Electrical control (RC) and communication cubicle is a microprocessor based controller that provides a directional over current, earth fault and sensitive earth fault relay, auto reclosing relay, instantaneous metering, event log, demand logger and remote terminal unit (RTU) for remote control in a single package. Figure 1 shows the steps involved in current manufacturing process of electrical control and communication cubicle.

The steps are:

1) Unpack: transport material from store to step 1 unpack area; and unpack material and transfer to step 2.

2) Decant: transfer unpacked material to workstation.

3) Fit breather and earth nuts: operator needs to fit breather and earth nuts on cubicle before putting the cubicle on assemble line.

4) Assembly RC cubicle: operator assembles cubicle at this point.

* For reasons of confidentiality, the name of the manufacturer cannot be disclosed. PPL is a pseudonym

5) Make travel card: for each work order, travel card (process layout) is prepared. Each equipment needs one travel card. One work order contains multiple travel cards

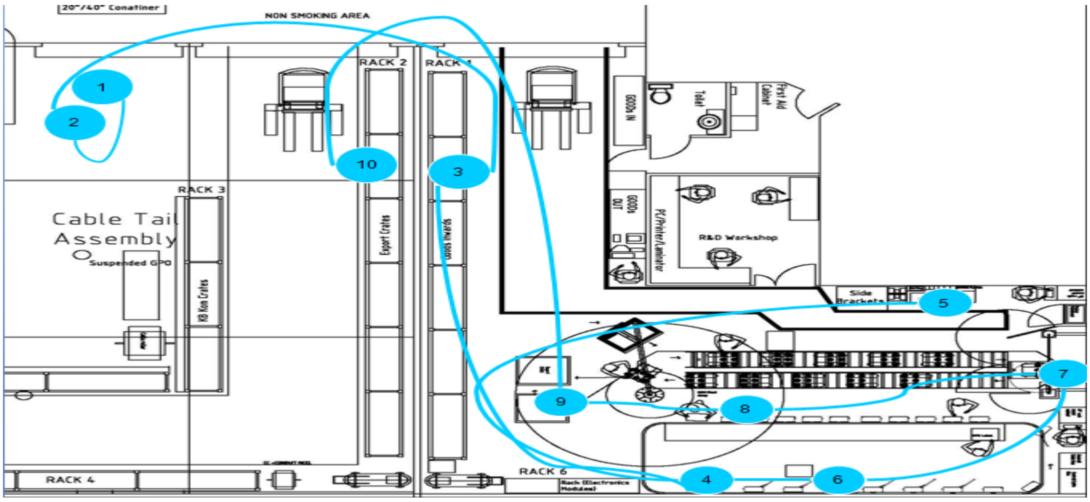

Figure 1: Current RC cubicle assembly process

6) Fit out RC cubicle: fit RC cubicle with associate modules.

7) Testing: RC control cubicle is tested at this cell together with OSM auto circuit recloser.

8) Final assembly: operator finishes the final assembly for packing.

9) Packing: packing RC control cubicle is done with OSM tank into crate; quality checks have been conducted before packing.

10) Delivery storage: The crate is moved to delivery storage area for delivery.

Based on the process flow shown in Figure 1 and steps described above, distances traveled by the operators for each activity were measured. Activities that did not add any apparent value were also identified. The process flow together with the distances traveled is shown in Figure 2. Value added and non-value added activities are also shown in Figure 2.

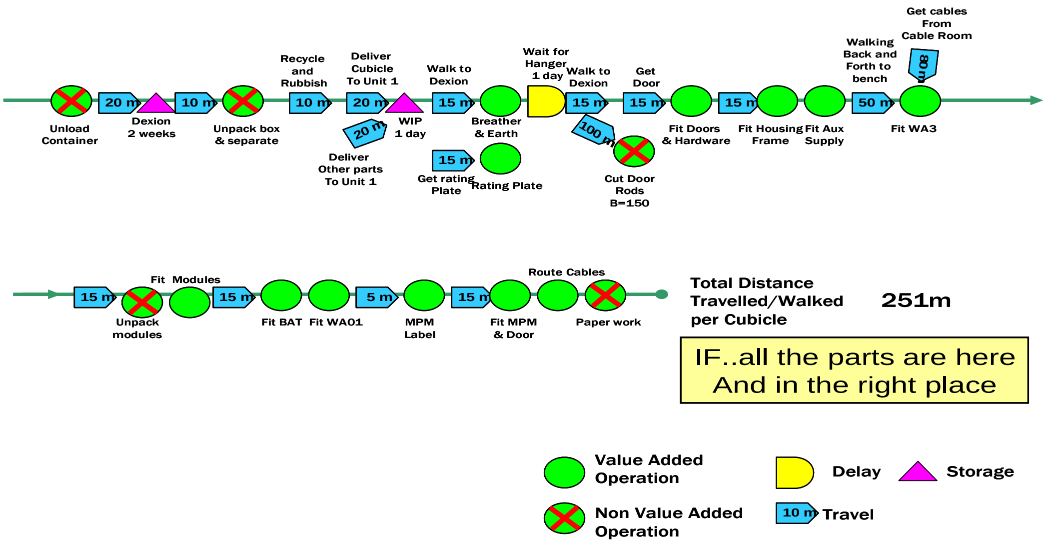

Figure 2: Current RC cubicle assembly process flow with distance travelled

Implementation of Lean Methods in the Case Organization

To implement lean strategy in the case organization, first existing work processes have been examined and possible wastes have been identified. Common lean tools such as time study, 5S, two-bin kanbans and design workstation are used to implement lean manufacturing in the case organization. RPM technique is proposed to be implemented at the end of the research.

Lean project team formation

Formation of 'lean project team' is the one of the underpinning factors within successful lean organizations. The use of integrated teams lead to improved quality, lower costs and shorter lead-time is well documented (Clark and Fujimoto 1991, Karim et al., 2008a). Therefore a multifunctional project team was formed which included the head of the engineering manager, production manager, one mechanical engineer, one lean manufacturing specialist, one skilled operator and one researcher.

Using a multidisciplinary team approach usually results in the facilitation of good communication throughout the project, ensuring that all relevant interrelations are taken care of. The faster information, decision and materials can flow, the faster a manufacturer can respond to changes, less time is spent in "fire fighting" and more time is available for performance improvement activities (Karim et al., 2008b).

In short, it is a strategy of teamwork to bring people together from different departments to work in a coordinated manner to improve efficiency.

Time study analysis

From the RC cubicle process map, the lean project team understood that efficiency of work processes are closely linked with operators and operating time. It was also found that a lot of valuable time was wasted in the process, and therefore, the work processing time should be recorded and examined in order to achieve the desired productivity as well as total improvement on processes. The main purpose of time study was to understand work processes and explore value-added time against non-value-added time. Time study is a particular method that may help the manager to understand entire work processes, identify wastes, highlight problems and imply appropriate solutions. The following steps were used for the time study:

Process recording: in this step, the operator's work process was video recorded while working on single product. Traditionally, the operator may separate the assembly work process into several stages and work on the three products at the same time. Therefore, in this project, team members asked the skilled operators to finish only one single product at a time to help the project work estimate the time for each step.

Break down of recorded time: project team reviewed the recorded video and has broken it into time segments that represented each of the details of work process.

Categorizing the process: the project team then discussed the work process with the manufacturing manager and skilled operators to determine whether the processes were of value added or non-value added category.

Sketch Non-value added and value added time spread: after estimating the time segments, an excel spread sheet was used to generate a bar chart to identify the total processing time.

The result of the time study is shown in a bar chart in Figure 3. It was observed that it took the operator around 20 minutes to finish RC cubicle outside frame assemble, about 21 minutes to finish inside module assembly, and about another 22 minutes to finish final processes.

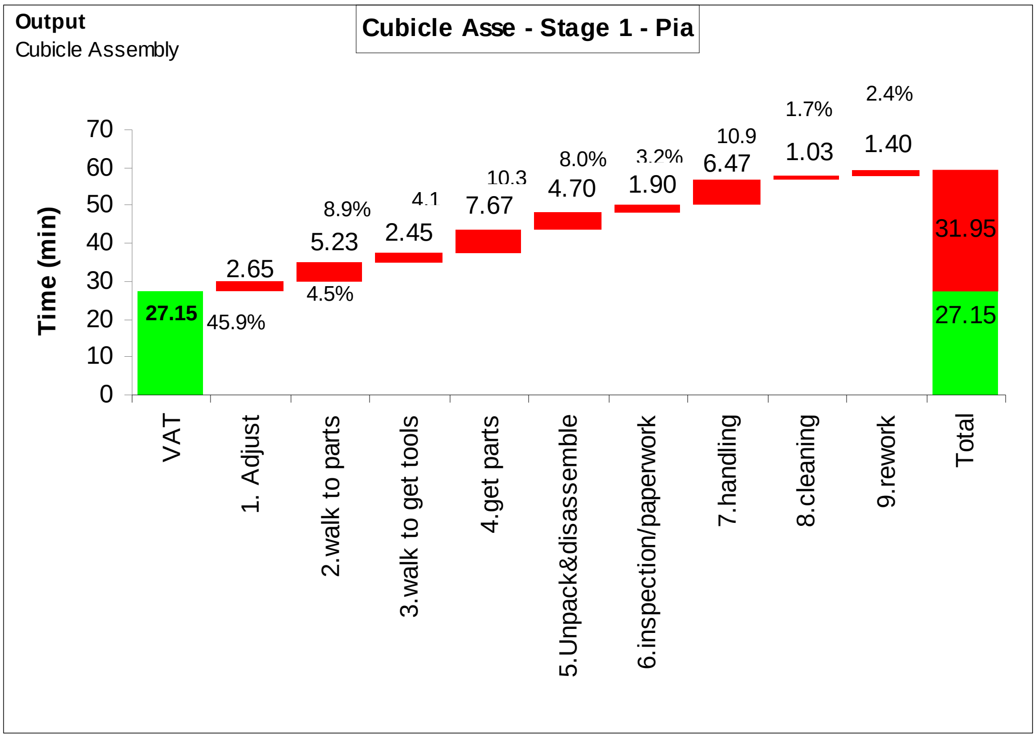

Figure 3: Time study analysis results

From the time study result, following main problems during assembly process were identified:

Handling: some double handling problems were identified, which were mainly caused by operator’s inexperience.

Part replenishment: most of the assembled parts were loaded on the work bench within single different size of bins and there was a miscommunication between operator and the person responsible for replenishment.

Walk distance: operators needed to walk to get parts and tools all the time, of which some of the walking times can be treated as non-value added and are considered waste.

Waiting and sharing tools: operators were sharing one set of tools, which may have caused increasing waiting time and can be treated as waste.

Process Improvement

Based on the work process analysis presented above, the project team sketched the work process material flow diagram and identified three main assembly problems. These are shown in Figure 4.

Figure 4: RC cubible assembly material flow

Firstly, the material supply aspect - from the material flow diagram it is obvious that assembled parts travel a long distance within the workstation (those green lines).

It causes problems not just for the operator to chase up the essential parts but also it creates a huge effort to manage the inventory. The second problem is the poor quality of the tools. Operators frequently experienced a shortage of right tools.

The third aspect is product quality, working processes experience a lot re-working process for the poor quality parts. Considering these aspects, an improved process map was designed by the lean team as shown in Figure 5.

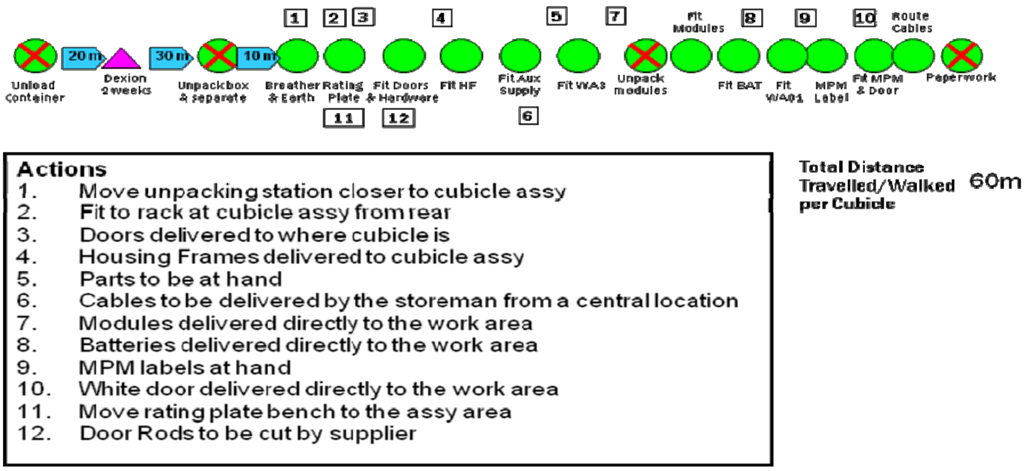

Figure 5: Ideal cubicle assembly process

The improved process map clearly shows that parts' travel distance was reduced from 251m to 60m. The improved process resulted a significant time saving for the total work process and eventually contributes to productivity improvement.

Implementation of a Kanban system

In this research, the proposed Kanban technology is a "two-bin" material handling system. A two bin kanban is one of the most popular kanban systems due to its simplicity. The basic idea is that an operator will get the material they need from one of the bins while the other bin is being refilled. This way, the two-bin system connects the production process with an invisible conveyor. When things are working properly the second bin will arrive shortly before the first bin empties. The advantage of Kanban part re-supply is that the system ensures an uninterrupted supply of parts without forecasts or complicated ordering procedures. The size of the bins is determined by the bin quantity and size of the parts. The system is described in Figure 6.

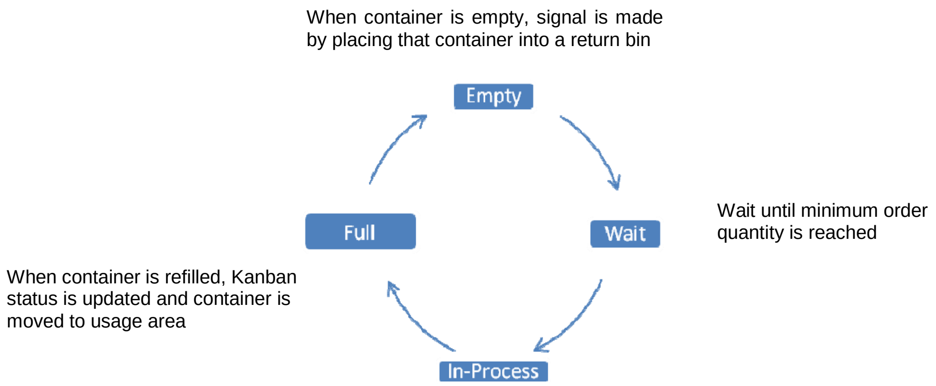

Figure 6: Replenishment Flow of Two-bins Kanban System

To design this two-bin Kanban system, several steps need to be followed and these are:

- Calculations are performed using the Bill of Material (BOM)
- Working out what modules are assembled in particular workstation
- Determining the quantity needed per unit and for the day's shift
- Determining the bin size for each part at the workstation
- Determining parts allocation for store/warehouse
- Determining parts allocation for bins
- Enter data into database or excel spreadsheet

Based on a two-bin Kanban system theory, a refill system was developed that was suitable for PPL. Every time when one of the bins is empty, the operator simply places the empty bin into the refill point which is located beside the workstation. At the end of the day, the other workman refills the empty bins and places them back to the responsible workstation. If the container quantities are correctly sized with the appropriate amount of inventory, the last part in the second container should be consumed just as the first container is returned from being replenished.

Redesign of layout and work processes

In the existing RC cubicle assembly work process, few workers assemble their own products.

As there is only one unit of tools box, it causes tool usage conflict and as a result increases the non-value added time in the work processes. The time study video record clearly showed that time was lost on changing working tools among staff frequently. This has not only produced more non-value added time, but also contributed to the poor quality of the product. Therefore, it was urgent to identify and separate the workstations to improve the working processes. This was extensively discussed in the project team discussion and proposed some changes to the existing system. The proposed changed are described below.

Three work process stages were defined based on the cubicle assembly material flow. The lean project team separated the workstation into three sections and assumed that each of them would consume equal amount of time to finish the product.

Work Stage 1: This stage is the basic process for total assembly. It contains most of the steps of RC cubicle main frame assembly. The processes in work stage 1 include following tasks:

1. Fit Breather and earth nuts.
2. Hang cubicle on the cubicle hanger.
3. Apply earth sticker.
4. Fit glands to cubicle.
5. Fit radio tray.
6. Fit gray door hinges.
7. Fit housing frame inside.
8. Assemble grey door.
9. Fit grey door and door seal then adjust door close gape.
10. Fill up paper work and put the travel card in the cubicle.

After the operator completes the work in the first workstation, the product will be passed to the next workstation and a new process will be started.

***Work stage 2*:** This stage needs the most of the special tools. The assigned tasks of this section are:

1. Fit power supply PCB board.
2. Install DIN rail into cubicle.
3. Fit GPO and MCB cable.
4. Connect earth cable.
5. Fit door stopper.
6. Assemble WA03 cable and fit in cubicle.
7. Unpack diver module, change label and fit in cubicle.
8. Fit power supply module.
9. Connect earth
10. Fit power supply module and certificate paper in cubicle.
11. Connect cable terminals and tie up cables.

***Work stage 3*:** The third stage also needs a very skilled operator to handle the RC cubicle door setup. Tools in this stage may conflict with the tools in the second stage. Therefore, the production manager needed to order second set of special tools. Tasks assigned at this stage are:

1. Unpack battery, stick data label and connect battery cable (WA02).
2. Fit battery in cubicle.
3. Connect cable terminals with module.
4. Assemble MPM and fit in white door.
5. Fit white door into cubicle.
6. Connect WA01 cable with MPM.
7. Adjust door gap.

As discussed before, the workstation has been separated into equal time sections. This may create a problem in the real life practice as the skill level of operators is not the same. Thus the lean project team recommends various training programs to increase the skill level of the operators.

Design of improved workstation layout

After the work stages were defined and work tasks were assigned, the workstation was redesigned to fit the adjustments of working process. The modified work flow of RC control and communication cubicle assembly line is shown in Figure 7. From the Figure 7, it can be seen that workstation has been redesigned and separated into three sections. Also some parts and tools relocation have been implemented. These are:

Parts and tools trolley: in order to reduce the time consumed for getting parts and tools, the lean project team decided to introduce three trolleys to store essential parts and tools right beside the operators.

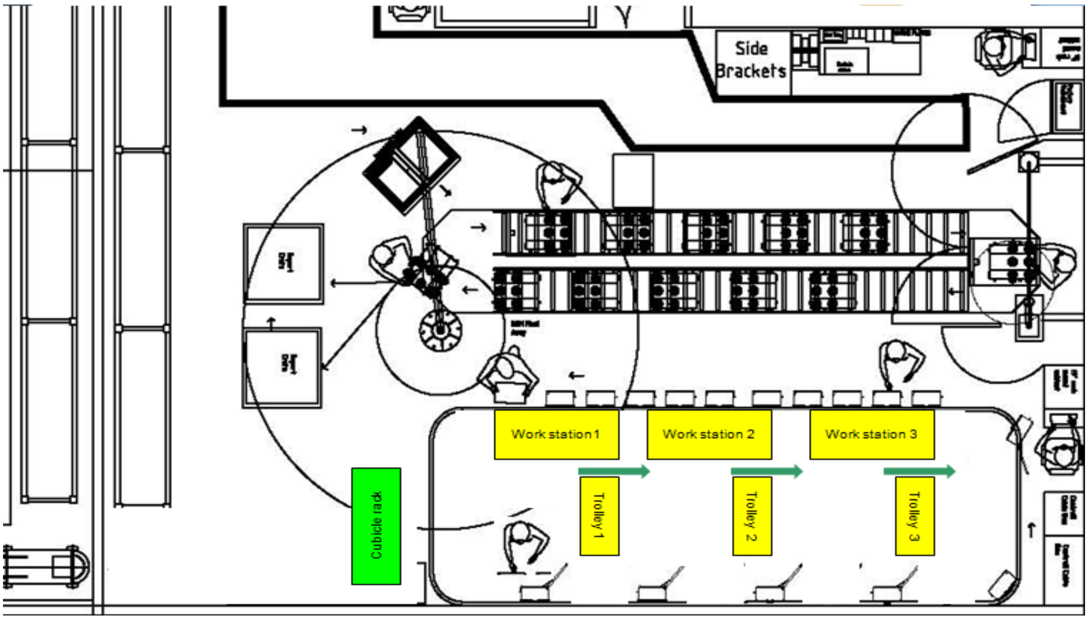

Figure 7: Modified workstation layout

Work task separation: to reduce the staff walking distance on each of the trolleys, only smaller parts have been loaded and the assembly process that needs bigger parts has to be done on the work bench.

Bigger parts relocation: under the work bench, there are some spaces for storage and these were used to store extra parts earlier in a disorderly manner. After the new layout, as the bigger parts need to be assembled in the work bench, the under bench spaces were used to store bigger parts for the each of the work stages.

Rapid Performance Management

A Rapid Performance Management (RPM) technique is proposed for the successful implementation of lean method. RPM is closely related to management and the staff. It allows the staff to operate at maximum effectiveness and to make improvement systematically. For the implementation of RPM technique, the target should be set and the performance of the employee should be recorded so that manager may explore problems within work process (Figure 8).

A review meeting (approximately 10 minutes) should be held every day between lead operators and managers to review the work records. If the work records achieve the company goal, the task is considered to be completed. However, if the work records have not reached the goal then discussions on the reasons are undertaken and a solution is proposed.

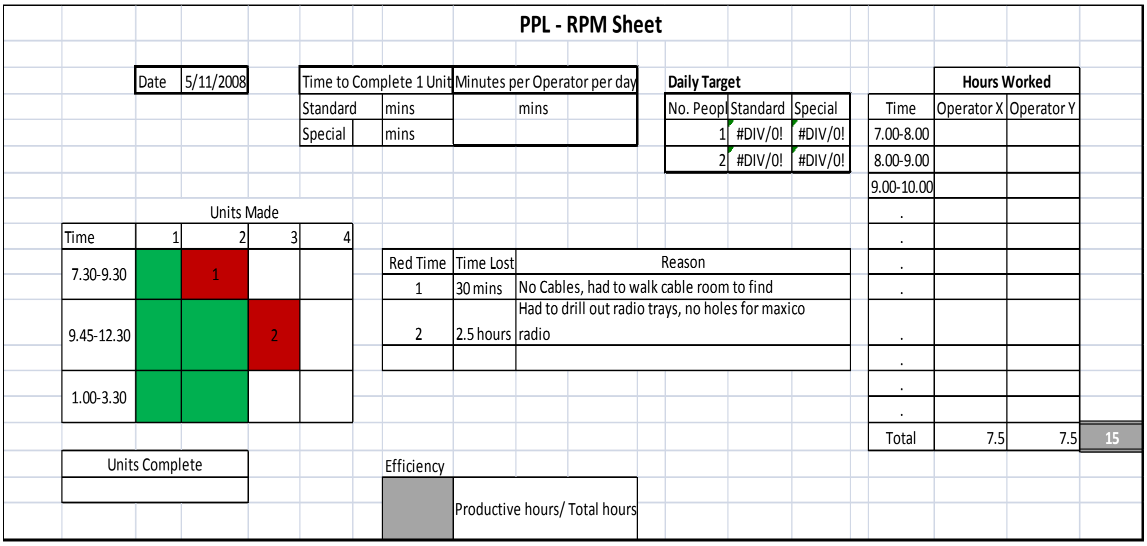

Figure 8. Proposed RPM sheet for PPL

Results of implementation of the Lean strategy

Implementation of lean strategies described above, particularly redesign of layout and work processes, implementation of two bin Kanban and rapid performance management has significantly reduced the process time in the assembly line. For RC cubicle assembly, the assembly time was reduced from 59 minutes to about 40 minutes, which is about 33% reduction. Many non value added activitites, such as walking, were eliminated or reduced. Redesign of the layout, splitting the assembly process in three work stages, implementation of Kanban, formation cross-functional lean team, implementation of rapid performance management and effective communication withing the organisation contributed to the improvement. The redesigned process significantly reduced the 'waiting times'.

Other benefits achieved are listed as follows:
1. The rate of meeting the promised delivery dates for customer orders was increased.
2. The processing time of customer enquiries was reduced.
3. Discord between different functional departments was reduced.
4. Product quality was improved as evidenced by fewer number of reworks after the assebly

Conclusion

Literature shows that applications of lean manufacturing have been less common in the MTOs. It is perceived that MTOs are less amenable to many lean techniques and there is a lack of the documented applications. This has caused the management to be reluctant to implement lean strategy. In this study manufacturing process of a MTO manufacturer has been investigated and lean methodology has been systematically implemented to reduce existing wastes and improve the process.

The main finding of this study is that lean strategy can successfully be implemented in MTOs and significant improvement can be achieved. However, for managers who are considering implementing lean manufacturing but are uncertain about the potential outcomes, we recommend that a RPM tool be used to evaluate the basic performance measures. The two-bin single card kanban system is used to manage inventory.

The Kanban system provided better parts management, ensuring continuous working process without seeking parts and helping quicker response to refill the bins.

Furthermore, the research has redesigned the workstation layout to make most of the assemble parts and tools closer to the operator in order to save process time and increase productivity. Future studies may consider many more aspects such as computer assistance during work to input values and record product performance. It is expected that lesson learned from this research would make a significant contribution to implement lean manufacturing strategy in MTO industries.

References

[1] Achanga, P., E. Shehab, R. Roy and G. Nelder, "Critical success factors for lean implementation within SMEs." *Journal of Manufacturing Technology Management* 17(4): 460-471, 2006

[2] Bozdogan, K., Milauskas, R. And Nightingale, D., "Transition to a lean enterprise", A Guide for Leaders,Vol. 1, Massachusetts Institute of Technology, Cambridge, MA.,2000.

[3] Bruun, P. and R. N. Mefford, "Lean production and the Internet." *International Journal of Production Economics* 89(3): 247-260, 2004.

[4] Clake, K.B. and Fujimoto, T., Product Development Performance, Strategy, Organization and Management in the World Auto Industry, *Harvard Business School Press,* Cambridge, MA., 1991.

[5] Dawson, P. and G. Palmer, Quality Management: The theory and practice of implementing change, Longman Cheshire, Melbourne, 1995.

[6] Deleryd, M., R. Garvare and B. Klefsjo, "Experiences of implementing statistical methods in small enterprises." *The TQM Magazine* 11(5): 341-50, 1999.

[7] Fleischer, M. and J. K. Liker, *Concurrent engineering effectiveness: integrating product development across organizations,* Hanser Gardner Cincinnati OH, 1997.

[8] Hendry, L. C. and B. G. Kingsman, "Customer enquiry management: part of a hierarchical system to control lead times in make-to-order companies." *Journal of the operational research society:* 61-70, 1993.

[9] Husband, S. and P. Mandal, "A conceptual model for quality integrated management in small and medium size enterprises." *International Journal of Quality and Reliability Management* **16**: 699-713, 1999.

[10] M. A. Karim, A. J. Smith and S. Halgamuge, A comparative study of manufacturing practices and performance variables, *International Journal of Production Economics,* 112, 841–859, 2008a

[11] M. A. Karim, A. J. Smith and S. Halgamuge, Empirical relationships between some manufacturing practices and performance, *International Journal of Production Research,* Vol. 46, No. 13, 3583–3613, 2008b.

[12] M. A. Karim, P. Samaranayake, A. J. Smith and S. Halgamuge, An on-time delivery improvement model for manufacturing organisations, *International Journal of Production Research,* Vol. 48, No. 8, 2373–2394, 2010.

[13] Oliver, N., R. Delbridge, D. Jones and J. Lowe, "World class manufacturing: further evidence in the lean production debate." *British journal of management* 5(s1): 53-63, 1994.

[14] Riezebos, J. and W. Klingenberg, "Advancing lean manufacturing, the role of IT." *Computers in Industry* 60: 235–236, 2009.

[15] Samson, D., A. Sohal and E. Ramsay, "Human resource issues in manufacturing improvement initiatives: case study experiences in Australia." *International Journal of Human Factors in Manufacturing* 3(2): 135-52, 1993.

[16] Shah, R. and P. T. Ward, "Lean manufacturing: context, practice bundles, and performance." *Journal of Operations Management* 21(2): 129-149, 2003.

[17] Sohal, A. S. and A. Egglestone, "Lean production: experience among Australian organizations." *International Journal of Operations and Production Management* 14: 35-35, 1994.

[18] Stevenson, M., L. C. Hendry and B. G. Kingsman, "A review of production planning and control: the applicability of key concepts to the make-to-order industry." *International Journal of Production Research* 43(5): 869-898, 2005.

[19] Thomas, A. J. and D. Webb, "Quality systems implementation in Welsh small-to medium-sized enterprises: a global comparison and a model for change." *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 217(4): 573-579, 2003.

ASOR Bulletin
Volume 29, Number 3, September 2010, pp.48-58
ISSN0812-860X | eISSN1446-6678

An Innovative Robust Reactive Surgery Assignment Model

Kari Stuart*, Erhan Kozan*, Michael Sinnott and James Collier****
*Discipline of Mathematical Sciences, Queensland University of Technology
GPO Box 2434, Brisbane Qld 4000, Australia
k.stuart, e.kozan@qut.edu.au
**Princess Alexandra Hospital Emergency Department
Brisbane Qld 4102 Australia
Michael_Sinnott, James_Collier@health.qld.gov.au

Abstract.

Online scheduling in the Operating Theatre Department is a dynamic process that deals with both elective and emergency patients. Each business day begins with an elective schedule determined in advance based on a mastery surgery schedule. Throughout the course of the day however, disruptions to this baseline schedule occur due to variations in treatment time, emergency arrivals, equipment failure and resource unavailability. An innovative robust reactive surgery assignment model is developed for the operating theatre department. Following the completion of each surgery, the schedule is re-solved taking into account any disruptions in order to minimise cancellations of pre-planned patients and maximise throughput of emergency cases. The single theatre case is solved and future work on the computationally more complex multiple theatre case under resource constraints is discussed.

Key Words: Robust Reactive Assignment, Online, Operating Theatre.

Introduction

The Operating Theatre (OT) department is a dynamic environment consisting of pre-scheduled elective patients and unpredictable emergency cases. Nearly every department within a hospital schedules patients for the Operating Theatre (OT) and most wards receive patients from the OT following post-operative recovery. Because of the interrelationships between units, disruptions and cancellations within the OT can have a flow on effect to the rest of the hospital. This often results in dissatisfied patients, nurses and doctors, escalating waiting lists, inefficient resource usage and undesirable waiting times. Improving the overall responsiveness to emergency patients by solving the disruption management and re-scheduling problem is used in this paper to improve the efficiency of the OT.

To date, most literature has focused on the scheduling of elective patients, which has been approached in many ways. Dexter and Traub (2000) applied decision theory to the process of sequencing patients. Pham and Klinkert (2008) treated surgical case scheduling as a generalised job shop scheduling problem. Galvin (1997) applied the cutting stock problem to the surgical scheduling problem. Persson and Persson (2006) used optimisation modelling to synchronise allocation of different resources for operating room planning. Sier et al., (1997) developed a tool for scheduling operations that considered bed availability, efficient theatre utilisation, minimising schedule deviations and emergency arrivals.

Although many good theoretical approaches exist for elective scheduling, many of these fail in the online setting due to disturbances including variability in surgical durations and the arrival of emergency patients. To address this issue, schedulers are moving away from deterministic and stochastic optimisation scheduling models towards robust scheduling models (Daniels and Kouvelis, 1995). Robust schedules are generated such that the schedule performance remains high even in the presence of online disruptions (Leon et al., 1994). This may be achieved by the inclusion of ‘buffers’ that absorb variations in treatment times that occur during project execution. Hans et al., (2008) introduced robust scheduling to the OT by assigning elective surgeries and planned slack time to the operating room days to prevent overtime. The planned slack time on each operating room-day is based on the expected variance of the surgical durations planned on that day. The effect of this is to create a schedule whose performance is relatively insensitive to the potential realisations of the task parameters (Daniels and Kouvelis, 1995).

Robust scheduling is an example of a preventive or proactive scheduling approach that serves as a baseline schedule for online production scheduling. Effective preventive schedules are important since they form the basis for resource commitment decisions. When used in conjunction with reactive scheduling models, they improve the performance of online scheduling (Li and Ierapetritou, 2008).

While robust schedules address quality robustness, they do not address solution robustness (Van de Vonder et al., 2005). Disruptions during project execution may cause deviations from a predictive schedule and even make it infeasible. Solution robustness is addressed by reactive scheduling, which is used to repair the baseline schedule following activity disruptions, by including changes whilst minimising disruptions from the original schedule. Two types of reactive scheduling include the repair of the existing schedule and full scheduling of tasks after a disruptive event i.e. re-scheduling (Li and Ierapetritou, 2008, Sabuncuoglu and Bayiz, 2000). To the authors’ knowledge, there is no literature to date dealing with reactive scheduling of elective and emergency patients for the operating theatre.

For operating theatre scheduling, two types of disruptions are defined; a theatre (machine) disruption and patient (job) disruption. Theatre disruptions occur when a theatre becomes unavailable for some period of time. Examples include equipment failure or staff shortage/unavailability or the arrival of a high priority emergency patient that requires the use of an elective theatre. This type of disruption results in the patients that were initially scheduled for that theatre (and have not yet been treated) to be delayed and the schedule must be updated to take into account such changes. Patient disruptions on the other hand occur when treatment times are less than or greater than the assigned surgery time. If surgery duration is shorter than expected, this generally means the schedule can be moved forwards without much alteration. Also, there may be time left at the end of the schedule for adding on emergency cases, or the additional time may be spent on a patient that exceeds its expected duration. If patients exceed their expected duration however, this can cause delays in surgery start time for remaining patients or may even necessitate cancellation of remaining surgeries to prevent overtime of the theatre.

Introduced is an innovative robust reactive assignment model (RRAM) for dealing with online disruptions in order to minimise cancellations of pre-planned patients and maximise throughput of emergency cases. This is achieved by keeping track of the immediately preceding schedule, before the completion of an operation. The original baseline schedule is developed using a robust scheduling approach that assumes surgical durations are lognormally distributed. This schedule provides the list of patients to be assigned to an operating theatre.

The Robust Reactive Assignment Model

The robust reactive assignment model (RRAM) is an assignment model that aims to minimise cancellations of the already assigned patients and maximise throughput of emergency cases following disruptions in the online environment. Disruptions include but are not necessarily limited to variations in a patient's estimated treatment time, the arrival of an emergency patient, equipment failure and resource unavailability (staffing or equipment). Regardless of the type of disruption, these lead to either early or late start times for the remaining patients. For example a patient that is completed early or an unexpected cancellation of an earlier patient may lead to an early start for the next patient. Conversely, a late finish or a delay in resource availability may lead to a late start for the subsequent patient/s or result in cancellations.

At the completion of each patient's surgery the RRAM is implemented. The model takes into account deviations between the actual duration of the surgery and the amount of time assigned to the procedure and any emergency arrivals during the course of the schedule. For example, if the procedure takes longer than expected, then there is less time available for the remaining patients on the schedule list. This is reflected in the amount of time remaining in the theatre, and some patients may need to be postponed if there is insufficient time to schedule all patients on the current list.

The objective of the reactive schedule is to minimise the costs of the new schedule brought on by cancellation of patients and offset by the inclusion of additional cases. The cost of cancelling (or profit earned by assigning) a patient depends on their priority level and whether or not they are an emergency or elective patient. Patients with higher priority receive a higher penalty for cancellation (or profit for inclusion).

Notations

i : Specialties considered within a surgical category, $i \in \{1,..,I\}$

j : Priority, $j \in \{1,..,J\}$

k : Type of patient (elective or emergency), $k \in \{1,..,K\}$

d_i : Expected surgical duration of specialty i

s_i^2 : Expected surgical duration volatility of specialty i

μ_i, σ_i : Lognormal distribution parameters for specialty i

M : The sum of the expected surgical durations assigned

V : The sum of the expected surgical durations assigned

T : Time remaining in the theatre

E_{ijk} : The number of available patients of specialty i, priority j and type k

X_{ijk} : The number of initially assigned patients of specialty i, priority j and type k

X'_{ijk} : The number of assigned patients of specialty i, priority j and type k following the reschedule.

C_{jk} : Cost/benefit of patients of priority j and type k

Objective function

The objective of the reactive schedule is to minimise the costs of the new schedule. These costs include the cost of cancelling a patient, the profit (or negative cost) earned for adding a patient and a penalty for deviating from the original schedule. The penalty incurred for cancelling a patient (or profit earned for adding one) depends on their priority level *j* and type *k*. This is represented by $C_{jk}\left(X_{ijk} - X'_{ijk}\right)$. If $\left(X_{ijk} - X'_{ijk}\right)$ is positive, then at least one patient has been cancelled and a penalty, C_{jk} is incurred for each cancellation. If $\left(X_{ijk} - X'_{ijk}\right)$ is negative, then at least one patient of specialty *i*, priority *j* and type *k* has been added to the schedule and a profit C_{jk} per patient is subtracted from the objective function. The second aspect of the objective function, $\left|X_{ijk} - X'_{ijk}\right|$, prevents the addition of patients at the expense of cancelling another. For example, one patient with penalty of 10 units could be cancelled and replaced with two patients, each with a profit of 5 units. In practice, this would generally be considered both impractical and destructive to the efficiency of the schedule.

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} C_{jk}\left(X_{ijk} - X'_{ijk}\right) + \left|X_{ijk} - X'_{ijk}\right| \qquad \mathbf{(1)}$$

Constraints

The number of patients assigned in the new schedule cannot exceed the number available. This allows for a cancelled elective patient to be re-assigned at a later event.

$$X'_{ijk} \le E_{ijk} \qquad (2)$$

For model simplification surgical duration mean and volatility of mean estimates are assumed independent on priority level *j* and whether the patient is an emergency or elective. Historical patient data was analysed and surgical durations are modelled with a lognormal distribution. The expected surgical duration, d_i and variance, s_i^2 of specialty *i*, are respectively given by

$$d_i = e^{\mu_i + \frac{\sigma_i^2}{2}} \qquad (3)$$

$$s_i^2 = \left(e^{\sigma_i^2} - 1\right)e^{2\mu_i + \sigma_i^2} \qquad (4)$$

where μ_i and σ_i are lognormal random variable parameters determined by analysis of historical data. The sum of the expected durations and the variance of the patients assigned to the theatre are given respectively by

$$M = \sum_{i=1}^{I}\left(\left(\sum_{j=1}^{J}\sum_{k=1}^{K} X'_{ijk}\right)e^{\mu_i}\sqrt{e^{\sigma_i^2}}\right) \quad (5)$$

$$V = \sum_{i=1}^{I}\left(\left(\sum_{j=1}^{J}\sum_{k=1}^{K} X'_{ijk}\right)\left(e^{\sigma_i^2}-1\right)e^{2\mu_i+\sigma_i^2}\right) \quad (6)$$

The amount of time that is planned for the patients assigned to the theatre depends on the level of accuracy desired by the decision maker. The level of accuracy used for the model is 15.87%, i.e. the probability that surgeries run overtime is less than 15.87%. The amount of time that is planned for the surgeries assigned to the theatre, based on this level of accuracy is given by

$$\left(\frac{M^2}{\sqrt{V+M^2}}\right)e^{\sqrt{\ln\left(\frac{V+M^2}{M^2}\right)}} \quad (7)$$

Equation 8 ensures the time available in the theatre is sufficient to complete all the assigned surgeries.

$$T \geq \left(\frac{M^2}{\sqrt{V+M^2}}\right)e^{\sqrt{\ln\left(\frac{V+M^2}{M^2}\right)}} \quad (8)$$

The number of patients assigned to each theatre is a positive integer

$$X'_{ijk} = \text{integer}, \ \forall i, j, k \quad (9)$$

Solution approach and results

The RRAM is a combinatorial optimisation problem and is a variation of a bounded knapsack problem in which the patients are the objects to be assigned to a theatre (knapsack) with limited capacity. The number of patients that can be assigned of a particular specialty, priority and type is bounded by variable E_{ijk}. Each patient takes up a portion of the limited capacity and has a benefit, C_{jk} associated with its assignment. However, in this case, the amount of time assigned for a given patient specialty is not fixed, but depends on the existing assignment of patients (see Equations 5 – 7).

The generalised bounded knapsack problem maximises the benefit associated with assigning the packed items. In our model, this is analogous to minimising $\left(-\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} C_{jk}X'_{ijk}\right)$. In order to minimise the costs of our objective, one must also solve the bounded knapsack problem, however, we have the added consideration of minimising changes from the original assignment of patients. Therefore our problem is just as computationally difficult as the bounded knapsack problem.

Knapsack problems have received much attention in the literature and are known to be NP hard in the ordinary sense and may be solved in pseudo-polynomial time. An algorithm that runs in pseudo-polynomial time has a running time that is polynomial in the numeric value of the input. Thus, as the number of patients increases, so does the computation time. However, in the case of the single operating theatre, the number of patients is relatively small and therefore can be solved without the use of metaheuristics.

Various exact solution methods exist including dynamic programming, branch and bound and exhaustive methods. Dynamic programming requires the problem to be broken up into stages that can be solved individually. Due to the robust calculations of processing time, dynamic programming cannot be applied to the RRAM. Branch and bound, commercial software or an enumerative approach could be used for the RRAM. The benefit of an enumerative algorithm over commercial software and branch and bound, is that the code can easily be adapted with changes in the problem structure. For example, if the problem is expanded to the multiple theatre case, the problem becomes analogous to a multiple knapsack problem or bin packing problem. For this reason, an enumerative algorithm is used and implemented using Visual Basic.

Implementing the model in Visual Basic has the additional benefit of a user-friendly interface, designed for use within the practical setting. The model is run after the completion of each operation. The user is prompted for information on the duration of the completed surgery, the type of surgery (elective or emergency), and the specialty and priority of the patient treated. In addition, information on emergency patient arrivals can be added at any time.

The enumerative algorithm is used to search for the best schedule based on the information supplied. The objective is to minimise changes from the original schedule whilst also searching for the schedule that produces the best objective value. This is important because it is not practical from a resource viewpoint to make large changes to the original schedule.

The RRAM was applied to a surgical care unit (SCU) (or day surgery unit). Day surgery patients generally arrive and/or are released on the day of their surgery. Using the SCU is a satisfactory representation of the whole OT department because both elective and emergency patients are treated there. The benefit to using the SCU is the reduction in problem size and hence calculations. The results obtained for the SCU can be extended and applied to the entire OT department.

The RRAM was tested by simulating the implementation of 100 offline robust schedules for a day surgery unit. Historical data was analysed to determine appropriate statistical distributions to generate surgical durations and also to determine the arrival pattern of emergency patients. Since the SCU was selected as a sample for the whole problem, specialties that are only treated in the SCU were sought after, to get a closer representation of the real life problem. If a theatre only treats a particular group (or category) of patients then the total capacity is dedicated to those patients. In other words, there is no need for adjusting total capacity available to those patients. If a number of different categories are treated in a theatre, then either all categories need to be considered, or the capacity must be adjusted for those selected. Of the categories treated in the SCU, the Ophthalmology patients are almost exclusively treated in two of the SCU theatres. Of the categories treated in these two theatres, the majority are also Ophthalmology. Most other categories (treated in the SCU) are also spread across the non-day surgery theatres. For this reason, the model is applied to the Ophthalmology patients and it is assumed that all of the available capacity is dedicated to those patients. This assumption may be changed according to a scheduler's requirements. It is important to note that the developed models may also be adjusted to incorporate any patient type and any theatre.

Historical data for the Ophthalmology patients was analysed for surgical duration estimates. The times available in the data provided were 'time in suite', 'in anaesthesia', 'in OR' and 'Out OR'. These times indicate the time the patient enters the operating theatre suite (which is composed of the operating rooms and their anaesthesia workrooms, day surgery unit, ICU, PACU etc), the time the patients enters anaesthesia and the times of entering and exiting the operating room respectively. The time of anaesthesia may be calculated as the difference between 'in OR' and 'in anaesthesia'. Likewise, the time in the OR is the difference between 'Out OR' and 'in OR'. These varied according to the type of surgery being performed. The total processing time of a patient was assumed to include anaesthesia preparation time, however this assumption can easily be changed if desired.

The Ophthalmology category was broken down into specialties for the calculation of surgical duration estimates. Six surgical specialties were determined based on the type of surgery performed. A histogram of actual data suggested a possible lognormal distribution for each of the surgical specialties. Goodness of fit tests supported the hypothesis that surgical durations may be described with the lognormal distribution. An example plot of the fitted distribution against actual data for the first specialty is provided in Figure 1.

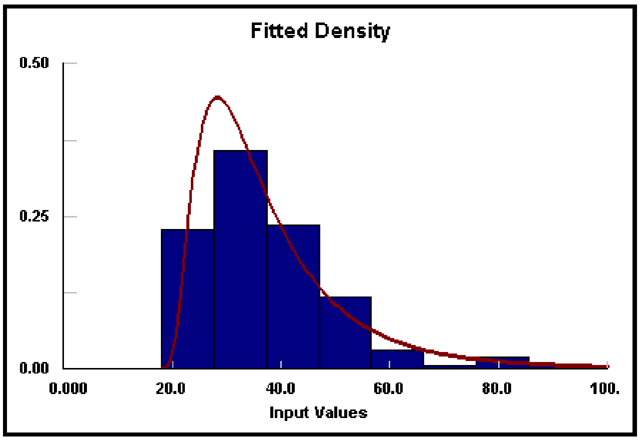

Figure 9. Fitted distribution versus actual data for specialty 1

The lognormal distribution is a continuous distribution, based on the normal distribution. It is used to describe many applications including physicians' consultation time, lifetime distributions, the long-term return rate on a stock investment and weight and blood pressure of humans. It has also been used in the literature for describing surgical durations (Jebali et al., 2006, Strum et al., 2000).

Parameters for the lognormal approximations were determined for each of the surgical specialties based on the data analysis and are presented in Table 1. The Lognormal distribution is described with 3 parameters, i.e. the mean of the included normal μ, the standard deviation of the included normal σ and the minimum value or location parameter γ. For data analysis, surgical durations were given in minutes.

Table 1. Data for surgical specialties

Specialty	Lognormal		
	γ	μ	σ^2
1	18	2.78	0.674
2	20	3.38	0.561
3	16	3.42	0.779
4	12	3.89	0.777
5	55	4.0	0.79
6	29	3.82	0.452

The elective patients for these test schedules were generated randomly, and assigned to the theatres using a multiple knapsack approach. Procedure durations, drawn from the appropriate lognormal distributions that were fitted from historical data, were generated for each patient. Emergency patients were randomly generated using an exponential distribution with an average inter-arrival time of 225 minutes (based on historical data). The results of the test cases are given in Table 2. The performance measures presented are the number of elective patients originally scheduled that were cancelled, the number of emergency cases performed, the number of electives cancelled but later re-scheduled and the number of emergencies added to the schedule that had to later be cancelled.

Table 2. Results of reactive schedules

Results	TOTAL
Number Electives Initially Assigned	741
Total Patients Completed	793
Number Emergencies Completed	178
Number Electives Cancelled	126
Re-scheduled Electives	71
Cancelled Emergencies	34

Results indicate that across the 100 examples there were a total of 126 elective cancellations and the RRAM was able to re-schedule 71 electives. The ability to re-schedule patients, when an already completed patient uses less time than it was allocated, illustrates the benefit of the robustness built into the model. In addition, by allowing a calculated amount of extra time for each surgery based on a percentage determined by the decision maker, the number of cancellations is kept low and allows for additional emergency patients to be seen. In this case, the model saw 178 additional emergencies performed across the 100 test cases. This ability to schedule additional patients ensures theatre capacity is used efficiently rather than being left unused. In 34 instances, emergency cases that had been tentatively assigned to a schedule had to be cancelled due to lack of time. Allowing for their cancellation also helps to maintain the efficiency of the theatre utilisation by preventing overruns. For the 100 test cases, only 16% resulted in over-run theatres.

In addition to measuring performance indicators, changes in the schedule may be presented in Gantt charts. Figure 2 illustrates the Gantt chart for one of the schedules. In the initial schedule, 9 elective patients are assigned to the theatre, each of which is illustrated with a different colour. After each surgery is completed, the RRAM is run and an updated schedule is generated. Following the late completion of the first patient (in yellow), the RRAM shifts the original schedule to the right and post-pones one elective patient (patient 8, in green) and fills the remaining capacity with an emergency patient (patient 10, in orange). The next seven procedures (patients 2 – 7 and 9) finish either early or on time. No more emergency patients are added to the schedule and patient 8 is not re-scheduled. After each of these procedures, the RRAM is implemented and there is either a 'left' shift in the schedule (for an early completion) or the schedule is unchanged (for on-time completions). The final patient (patient 10, in orange) requires an additional 5 minutes of surgery time. It is evident from the Gantt chart that this particular schedule completes 'on-time' meaning that it does not exceed the capacity of 480 minutes.

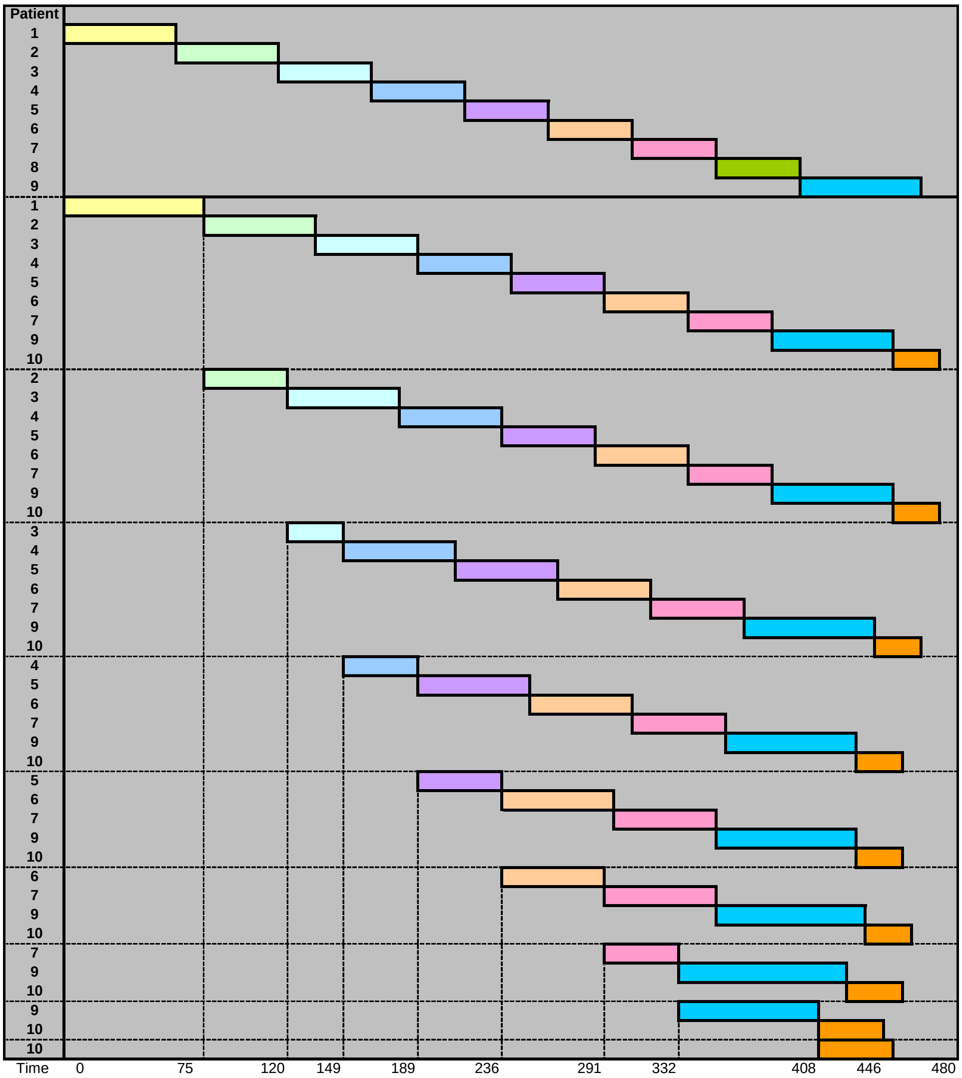

Figure 10. Gantt Chart illustrating results for one schedule

Conclusions

An innovative online assignment model for a single Operating Theatre is developed and solved. The model is run in real-time following the completion of each operation and minimises cancellations whilst also allowing for additional scheduling of emergency cases (time permitting), which may arise during the schedule's implementation. The problem is NP hard in the ordinary sense and hence an exact solution approach was used. The model was developed and implemented using Visual Basic.

Results for the RRAM showed it was capable of adapting appropriately to disruptions in the online environment by delaying, rescheduling or adding additional surgeries according to the available operating time capacity. The ability to re-schedule patients, when an already completed surgery used less time than it was allocated, illustrated the benefit of the robustness built into the model. In addition, by allowing a calculated amount of extra time for each surgery based on a percentage determined by the decision maker, the number of cancellations was kept low and additional emergency patients could be treated. This ability to schedule additional patients ensured theatre capacity is used efficiently rather than being left unused. Allowing for the cancellation of emergencies also helped to maintain the efficiency of the theatre utilisation by preventing overruns.

One limitation of this research is that it only considers the assignment of patients to a single operating theatre and does not consider patient sequence. Future work of the authors includes the adaptation of the model to include patient sequence and also to extend the problem to include multiple theatres.

As mentioned earlier in the paper, the objective of the model is to minimise the costs of the new schedule by minimising the number of elective cancellations and maximising the number of electives that may be added to the schedule. Because the model is re-run after each patient's completion it does not keep track of the preceding objective function values. The resulting schedule may be different if the objective function values were carried through for each schedule. For example, the number of elective cancellations may possibly be reduced with an accompanied decrease in emergencies added to the schedule. This could also form a future topic of research. Other changes to the schedule results could be induced by changes in the scheduler's objectives and should be investigated. For example, if an elective is cancelled relatively towards the beginning of a schedule, then the user may opt not to fill in the remaining capacity with available emergencies at that point in time and wait until later in the schedule.

References

[1] R. L. Daniels and P. Kouvelis, "Robust scheduling to hedge against processing time uncertainty in single-stage production," *Management Science*, vol 41, pp.363-376, 1995.

[2] F. Dexter and R. D. Traub, "Sequencing cases in the operating room: predicting whether one surgical case will last longer than another," *Anesthesia & Analgesia*, vol 90, pp.975-979, 2000.

[3] W. Galvin, "Hospital theatre scheduling and cutting-stock problem," *ASOR Bulletin*, vol 16, pp.3-12, 1997.

[4] E. Hans, G. Wullink, M. Van Houdenhoven and G. Kazemier, "Robust surgery loading," *European Journal of Operational Research*, vol 185, pp.1038-1050, 2008.

[5] A. Jebali, A. B. Hadj-Alouane and P. Ladet, "Operating rooms scheduling," *International Journal of Production Economics*, vol 99, pp.52-62, 2006.

[6] V. J. Leon, S. D. Wu and R. H. Storer, "Robustness measures and robust scheduling for job shops," *IIE Transactions*, vol 26, pp.32-43, 1994.

[7] Z. Li and M. G. Ierapetritou, "Process scheduling under uncertainty: Review and challenges," *Computers and Chemical Engineering*, vol 32, pp.715-727, 2008.

[8] M. Persson and J. A. Persson, "Optimization modeling of hospital operating room planning using a logistic perspective," *Proc. of 2006 Annual Conference of OR Applied to Health Services*, 2006.

[9] D.-N. Pham and A. Klinkert, "Surgical case scheduling as a generalized job shop scheduling problem," *European Journal of Operations Research*, vol 185, pp.1011-1025, 2008.

[10] I. Sabuncuoglu and M. Bayiz, "Analysis of reactive scheduling problems in a job shop environment," *European Journal of Operational Research*, vol 126, pp.567-586, 2000.

[11] D. Sier, P. Tobin and C. McGurk, "Scheduling surgical procedures," *Journal of the Operational Research Society*, vol 48, pp.884-891, 1997.

[12] D. P. Strum, J. H. May and L. G. Vargas, "Modeling the uncertainty of surgical procedure times: comparison of lognormal and normal models," *Anesthesiology*, vol 92, pp.1160-1167, 2000.

[13] S. Van de Vonder, E. Demeulemeester, W. Herroelen and R. Leus, "The use of buffers in project management: the trade-off between stability and makespan," *International Journal of production economics*, vol 97, pp.227-240, 2005.

A Recognition for ASOR Bulletin

ASOR Bulletin has recently been recognised as a refereed journal and upgraded it from unranked to 'C' journal in the latest 2010 ERA listing prepared by the Australian Research Council (ARC). See the details below with the field of research codes. Note that APJOR and ITOR are also ranked as 'C' in 2010 ERA. I like to thank all the editorial board members, our reviewers, and web editor for their advice /suggestion and effort in achieving such a wonderful recognition.

Ruhul Sarker
Editor-in-Chief
ASOR Bulletin

Journals on 2010 final list with

ERA Journal ID	Ranking	Title	ISSN
		Field of Research: 0806 Information Systems	
41356	C	ASOR Bulletin	0812-860X 1446-6678
		Field of Research: 1503 Business and Management	
41356	C	ASOR Bulletin	0812-860X 1446-6678

For 2010 journal and conference ranking visit: http://lamp.infosys.deakin.edu.au/era/index.php

ASOR BULLETIN

Editorial Policy

The ASOR Bulletin is published in March, June, September and December by the Australian Society of Operations Research Incorporated.

It aims to provide news, world-wide abstracts, Australian problem descriptions and solution approaches, and a forum on topics of interests to Operations Research practitioners, researchers, academics and students.

Contributions and suggestions are welcomed, however it should be noted that technical articles should be brief and relate to specific applications. Detailed mathematical developments should be omitted from the main body of articles but can be included as an Appendix to the article. Both refereed and non-refereed papers are published. The refereed papers are *peer reviewed* by at least two independent experts in the field and published under the section 'Refereed Paper'.

Articles must contain an abstract of not more than 100 words. The author's correct title, name, position, department, and preferred address must be supplied. References should be specified and numbered in alphabetical order as illustrated in the following examples:

[1] Higgins, J.C. and Finn, R. Managerial Attitudes Towards Computer Models for Planning and Control. Long Range Planning, Vol. 4, pp 107-112. (Dec. 1976).

[2] Simon, H.A. The New Science of Management Decision. Rev. Ed. Prentice-Hall, N.J. (1977).

Contributions should be prepared in MSWord (doc or rtf file), suitable for IBM Compatible PC, and a soft copy should be submitted as an email attachment. The detailed instructions for preparing /formatting your manuscript can be found in the ASOR web site.

Reviews: Books for review should be sent to the book review subeditor A/Prof. G.K.Whymark, c/- the editors. Note that the subeditor is also interested in hearing from companies wishing to arrange reviews of software.

Advertising: The current rate is $300 per page, with layout supplied. Pro-rata rates apply to half and quarter pages and discounts are available for advance bookings over four issues.

Subscriptions: ASOR Bulletin electronic version is free for all members and non-members which is accessible through ASOR web site.

Deadlines: The deadline for each issue (for all items except refereed articles) is the first day of the month preceding the month of publication.

Editor: Address all correspondence and contributions to:

A/Prof. Ruhul A Sarker,
School of Engineering & IT, UNSW@ADFA
Northcott Drive, Canberra ACT 2600
Tel: (02) 6268 8051 Fax: (02) 6268 8581
Email: r.sarker@adfa.edu.au