A RE-ENGINEERING PROJECT BASED ON THE TOC FUNDAMENTAL MEASUREMENTS

Jos Schaefer and Riad Aggoune
Laboratoire des Technologies Industrielles
Centre de Recherche Public Henri Tudor
L-4221 Esch-sur-Alzette, G.D. Luxembourg.
Riad.Aggoune@tudor.lu

Abstract

This paper presents a case study investigating a re-engineering project based on the Theory of Constraints (TOC) fundamental measurements. Bad bottom line results triggered the project. The main constraint is the market itself. Moreover, a policy constraint encouraging local optima in the planning and scheduling process limits the company from reaching its goal and prevents it from exploiting the market constraint. The paper also discusses briefly a risk analysis of such projects.

Keywords: planning, scheduling, theory of constraints, real case study.

1. INTRODUCTION

E.M. Goldratt developed the theory of constraints in the early 80s. Rather than a theory, TOC is a philosophy and set of techniques to manage and optimize the activity of the business (see Goldratt, 1992). It can be seen as a triplet, namely the DBR (Drum-Buffer-Rope) process for synchronous manufacturing, the Thinking Process for problem solving and the measurement system, which keeps the system running towards its goal. These 3 tools are needed to actually succeed in improving bottom line results and leveraging the limits of any organization. The basic assumption is that every system is limited by a constraint. At the same time TOC teaches that there are only a few actual constraints in any system, depending on the manufacturing typologies, response to the market, production technology, cost of raw material and added value. Eventually, elevating those constraints allows the system to improve. Therefore TOC provides a 5-step process for continuous improvement: 1. Identify the constraint. 2. Exploit the constraint. 3. Subordinate to the constraint. 4. Elevate the constraint. 5. Return to step 1. Each time an organization needs to change, TOC asks 3 basic questions: What to change? What to change to? How to change?

The planning process has a direct influence on a company’s productivity, independent of what it produces. A significant amount of literature on best planning practices exists and APICS (American Production and Inventory Society) recommend the best well known.

In the following sections, we first describe the problem that was subject of this study then we present the developed solution before concluding on the results obtained after the implementation of the planning model.

2. PROBLEM DEFINITION

2.1 Context of BMA

BMA is a small and medium enterprise (SME), while embedded in a multinational group. It is one link in a commodity product supply chain for special metallic sheets that buys big coils of metallic sheets, the mother rolls, from the group. Raw material may be differentiated by thickness and metal quality. BMA’s business consists on cutting mother rolls into smaller coils with reduced lengths and widths. These coils are then transformed into foils split and wound to customer widths and lengths. As final items, rectangular plates are cut out of the foils, with additional width and form (plates or coils) parameters. The products of the different steps in the process are given in Fig. 1.

BMA’s manufacturing environment has two specificities. Firstly, it only adds little value to the product. Secondly, the cutting process is always constrained by the optimization of raw material utilization.

In the wood and metal industries, two types of cutting problems are encountered, the surface and the linear cutting. The first one is a problem of optimization of the cut surface, which consists in finding the best raw material utilization when cutting n customer orders out of m raw material plates, with the objective of minimizing the quantity of used material. As mentioned by Hartmann (2000), it is very similar to the resource-constrained project scheduling problem (RCPSP) with the makespan minimization. This kind of highly combinatorial problems being strongly NP-
hard (Garey, et al., 1976), heuristics and meta-heuristics are the most used optimization processes. For further details, the authors refer to the recent survey of Lodi, et al. (2002), reviewing solution methods and complexity results. Second is the linear cutting, which is exactly the problem of BMA. It is different from the surface cutting since the whole surface to be cut is not available at the same time, the roll being unwound (see Fig. 1).

2.2 Previous Planning System

Before the new model was implemented, the BMA planning was based on local performance criteria. It aimed at optimizing the manufacturing shop comfort rather than customer service. The system was almost completely decoupled from customer demand, but designed to maximize raw material and resources utilization. Therefore, the scheduler was used to identify customer orders to be cut from the same mother roll and to release the manufacturing orders independently of customer due dates. The result was that some customer orders were finished in advance and sat in the ready-to-ship inventory for weeks and months incurring inventory costs and space problems. Some of these orders became late because someone forgot to ship them. But there were huge customer orders waiting to be produced since customer demand for the specific raw material was too low. Furthermore, this planning system created at least the same amount of work in process (WIP) at the level of coils that was not linked to any customer order. The re-utilization at those coils became difficult for two reasons. First, the dimensions did not meet the majority of incoming customer orders. Second, the lack of identification and follow up of this WIP caused reproduction of the same coils in the event of an actual need. The WIP thus generated additional WIP (reproduction in this case means cutting a new mother roll) inducing a spiral effect.

Bottom-line results were a disaster. This can easily be shown with the three basic performance indicators used in the theory of constraints that are the Throughput (T), the Inventory (I) and the Operating Expense (OE). Measuring form those indicators it could be stated that there was no money generated. Current throughput was negative because late orders mean late payments. The company's activity had low added value compared to raw material costs. Future throughput was in danger because a bad customer service level in a commodity market looses customers (Christopher, 1998). There was high money absorption. Inventory, including all investments, was growing continuously with WIP generation. Raw material consumption was continuously above sales. Operating expense increased. Order expediting became the most frequent activity. This is the typical end of month syndrome when customer orders are shipped despite additional costs. This means overtime, but also high transportation costs. Eventually operating expense was spent to absorb money instead of transforming absorbed money into throughput. To conclude, all three indicators were pointing in the wrong direction. The company was loosing money and might have been constrained to stop its business soon or later.

3. SOLUTION METHOD

3.1 Dominant Flow: Identifying the System Constraint

A first analysis, focused on improving bottom line results, showed that most of the problems could be solved with an efficient planning model respecting customer due date and trying for maximizing raw material utilization. The only remaining problem would then be the improvement in the vendor relationship. In terms of supply chain management, this means accelerating the dominant flow and mastering the control flow. Efficiency in a commodity supply chain means having a chain with a global low cost (Christopher, 1998). This again means that participants must be synchronized with the global flow speed.

An adapted planning system is a necessary condition for synchronization. Accelerating the dominant flow needs a deep knowledge of this flow. The project therefore started with mapping the flow. The flow analysis, for which we refer to the paper of Schaefer, et al. (1997), first includes an analysis of the communication and decision structures, horizontal and vertical, then the mapping realization, and finally the validation of the study. Considering processes instead of sector of activities, and subordinating theses activities to the flows, allows simplifying the original mapping. The results are illustrated in Fig. 2.

A flow map allows an understanding of the company's organization and focuses on its dysfunctions. It is furthermore an excellent tool for measuring the maturity for company change, and is thus a very powerful tool for evaluating project risk. Finally, the flow map enables us to draw an ideal future flow and define an action plan. The major differences between the original and final map are firstly, the introduction of a basic flow coordination function (an actual logistic function) and secondly, a change in subordination. In the original flow map, the flow is subordinated to the functions. In the final flow map, functions are subordinated to the flow. In terms of TOC, the dominant flow was identified as the main constraint (directly linked to the market constraint). The second step was exploiting the constraint. Exploiting required a fundamental flow simplification, im-

![Fig. 1 Production stages at BMA.](image-url)
implementing an efficient planning process and linking it to a simple scheduling system integrating secondary constraints at all levels. The third step was subordinating the whole company to the constraint. An initial requisite condition was straightforward customer order flow management including planning and scheduling realized by a flow management and synchronization function. The subordination process became the project part with a high change risk. It has been the author’s experiences that change processes have three requisites. First, give information and ensure transparency, second, provide education and third, create win-win situations.

3.2 Planning Model

The planning model was developed from a rather chaotic planning system: BMA is in a make-to-order situation with the exception for a few large customers that have business contracts and replenish periodically. The planning model is in two parts: the master production schedule and the finite capacity schedule. The basic principles for the entire system are built on the following objectives:

- Aim 100 % of customer service level.
- Optimize global material and capacity resource utilization.
- Keep raw material at minimum entropy.
- Minimize inventories.
- Minimize operating expenses, such as overtime and energy consumption.
- Minimize transportation cost.

As all of these elements are interdependent, it is important not to try for local optimization. All elements play a part in the solution choice and there must be compromises regarding them. However, customer service remains the dominating criteria. In fact, it is the main qualifying and winning order in the market in which BMA competes. In order to successfully develop a planning system, one needs two basic parameters: customer lead-time, which is the time to deliver an order and manufacturing lead-time, representing the time needed to produce the goods. Both allow us to define the planning horizon and the main zones delimited by the demand and the planning time fences (Vollmann, 1992). A short analysis of the sales and the manufacturing department revealed that customer lead-time was around 6 weeks and manufacturing lead-time including the shipment but excluding raw material replenishment could become easily stable at 1 week. This led us to consider a planning horizon of a minimum of 3 weeks with a frozen zone of 1 week. Fig. 3 shows the basic zones.

In Fig. 3, the segments corresponding to orders are obtained by first marking the orders due dates on the planning horizon. Then, starting times for production are obtained for each of the orders by subtracting manufacturing lead-time and a two days security from these due dates. The planning is initiated in week W. The plan is manufactured in week W+1 and shipped at the beginning of week W+2. All firm orders existing in the backlog for the next weeks W+3, W+4, etc. are considered. As a consequence, all orders that have due dates in W+2, must be produced during week W+1.

This model has to integrate a certain number of constraints or criteria for exploiting the flow constraint to a maximum. The fact that the cutting business has an exceptional product cost structure: 80 % of raw material, 20 % of added value, is particularly important, and raw material utilization therefore is a major concern. Traditionally, the (linear) cutting industry exploits total backlog orders and tries to maximize material utilization by optimizing the cutting process with respect to the mother roll. This means that lanes to be cut must have an optimal distribution with respect to the roll’s width. It could be estimated that for a representative business period, up to 60 % are produced for inventory. This method of optimizing raw material utilization is therefore decaying productivity. It is just a local optimum on manufacturing resource utilization and does not respect the constraint. This model subordinates cutting process backlog, thus minimising set up and operation time. However, the main objective must be to ensure customer service in order to keep and win customers; use overcapacity for flexibility and rush orders, thus creating a niche for

![Fig. 2 Original (top) and final (bottom) flow maps](image)

![Fig. 3 Planning zones](image)
higher current throughput via higher prices and for increased future throughput via customer satisfaction.

3.3 Planning Constraints

The new planning model is designed with a simple master production schedule respecting initially the customer due dates and then integrating all mentioned constraints. The basic idea is to produce exactly what the customer asks for as the moment he asks for it. The different criteria are not aggregated in a unique economic function, and the customer service is the optimized one. The other criteria are then used to select a solution, with the importance order given by the following subsections.

3.3.1 Raw material utilization

Developing the idea of optimal material utilization in the sense of the TOC measurement system changes completely the cutting philosophy and the material optimization system. Rather than optimizing with respect to the width of the mother roll and cutting down the whole roll, the new method optimizes the length to be cut and places the customer order in the correspondent virtual rectangle. This means that coils, foils and plates due in week W+2 are the only material cut. To validate this method for cutting the metal sheets, a simulation was run with the backlog of a representative business period. This simulation showed that there were 0 tonnes of WIP produced. The material losses were less than 5 percent, i.e. the company’s objective. The cutting lines that appear during unwinding must remain lined up and cannot stop until the end of the unwinding process. In the literature, this last constraint is known as the guillotine constraint (Lodi, et al., 2002). Moreover, successive lanes that are lined up must have widths that are multiples. These placement constraints ultimately create material loss and the sum of surfaces for different customer orders will always be smaller than the global surface out of which the orders are cut. This is a first-decision parameter in the planning process. The backlog may in fact contain more orders than those due for W+2 as shown on Fig. 3. Ideally, orders that are due for W+2 will be processed exclusively. Nevertheless, for decreasing material losses, it might be necessary to include some more orders that are due for later. Material utilization thus becomes an important planning parameter. If orders due for W+3, W+4 etc. must be cut together with current orders, the smaller material losses will be partially neutralized by inventory costs. Therefore, unless customers accept earlier shipments and unless supplementary shipments do not disturb (render more expensive) the shipment schedule, the sum of all costs generated by order pre-production must be considered. The condition for pre-production can be calculated by the following rate, namely it is accepted if ratio \( P \) is less than 5 \%, with:

\[
p = \frac{S_{\text{tot}} - \sum_{k=1}^{N} S_{njk}}{S_{\text{tot}}} < 5\%
\]

\( S_{njk} \) : surface of order \( n \) being produced in period \( j \) from mother roll \( k \).

\( S_{\text{tot}} \) : total surface of raw material being cut \( (S_{njk} \leq \sum_{k=1}^{M} S_{njk}), N \) being the total number of orders to produce from \( k \).

3.3.2 Capacity utilization

Line capacity is measured in m/min rather than in tons/min, thus eliminating the thickness factor. In fact, the speed of the production varies by levels depending on the material thickness. To calculate the capacity condition let us define the following quantities:

\( C_j \) : total available capacity during period \( j \),
\( L_j \) : total load for period \( j \),
\( m_{kj} \) : total number of mother rolls \( k \) in period \( j \),
\( l_{kj} \) : total load for material \( k \) in period \( j \),

Latter is calculated since \( l_{kj} = S_{nk}/W \), \( W \) being the standard mother roll width. The load for the different material characteristics depends on the number of speed levels. It can be calculated from the formula:

\[
L_{kj} = \sum_{i=1}^{N} l_{ki} \quad \text{and} \quad L_j = \sum_{k=1}^{N} l_{kj} + \sum_{k=N+1}^{M} l_{kj+1}
\]

The last term in this expression represents the load due to forwarded orders. The capacity condition can be calculated from the following:

\[
x \times C_j > L_j + M \times T_{su}
\]

where \( T_{su} \) represents the capacity loss due to machine set up and where \( x \) is a safety capacity coefficient varying from 0 to 1. This coefficient has a strategic value adjusted to actual business analysis and flow performances. It avoids planning for total available capacity thus preventing queues. Furthermore it saves capacity for rush orders. Safety capacity is an interesting lever to respond to the VIP customer as well as to the new potential customer in trouble with his regular vendor. During the machine loading process special attention must be paid to capacity constrained resources.

3.3.3 WIP inventory

The model aims for zero WIP inventory. WIP in this context means material cut to coils that have no link to customer orders. There is another kind of temporarily produced WIP that is the result of optimal material utilization. Each time, as described for the raw material condition, the cutting losses generated by the production for week W+2 is too high, a heuristic forwards orders for improving the loss coefficient. This process creates temporary WIP, the so-called ready-to-ship-inventory if the corresponding customers do not accept early deliveries. In the latter case, this WIP creates costs for different reasons: inventory costs, management costs (it needs specific identification for avoiding wrong shipments), handling costs, (the shipping area is partially blocked), and finally operating expense. For these reasons, material loss and inventory generation must be compared by using the basic indicators of throughput, inventory and operating expense.
The condition is:

\[ c_j \cdot d \cdot l_{cc} \leq \Delta \Lambda \]  

(4)

where \( c_j \) is the item cost entering WIP, \( \Delta \Lambda \) is the variation of raw material losses due to additional order cutting, \( d \) is the period of WIP, and \( l_{cc} \) the inventory carrying costs.

3.3.4 Transportation criterion

Transportation costs are one important issue especially in the case of shipping heavy products. The organization of a shipping schedule is therefore highly critical. As we will see in the next paragraph, the shipping schedule will be the drum in the sense of synchronized production technology (Goldratt, 1992) for the final manufacturing schedule. Sending trucks with a bad filling rate is to be avoided. If the shipping supervisor has problems filling all trucks with the orders due in W+2, he might suggest extending the production of orders allowing a better fill rate. This is possible if the corresponding customers will accept early shipments. It is only recommended if early shipments will not generate additional costs. Those costs can be a request from the customer willing to accept the material in the form of a discount. Again all additional costs must at least be balanced by the transport cost reduction. It should be mentioned, finally, that efficient shipment scheduling and strong relationship with the transportation companies generally avoids such constraints.

3.3.5 Final production and shipment schedule

The constrained planning model generates a list of planned orders. The next step is to calculate the final production schedule thus defining the work order sequence in the manufacturing shop. In the particular case of BMA, manufacturing and shipment schedules are strongly linked. The market (and therefore the dominant flow) is the main constraint at BMA. A shipping buffer is needed to protect the constraint. This buffer is currently estimated at 2 days. Upstream, the flow must be synchronized to the customer delivery schedule. Thus shipping and manufacturing schedules must be synchronized. Of both processes, the manufacturing process is the most flexible. The shipment process in fact has two constraints, the time and the filling rate. For these reasons, the shipping process was chosen as the drum for synchronization. Ultimately it makes the global scheduling very simple but very efficient. It starts from the definition of transportation zones as in Fig. 4. There are 4 main shipment zones with a partial coverage allowing 8 shipment possibilities, zones 1 to 4 and their combinations as well.

The planned orders are sorted since the customer due dates. Then, by starting with the most urgent order, we select all orders with a destination in the same shipping zone. The due dates of the selected orders are then updated, what provides us a new planning list, which is finally sorted according to urgency. The planning list being established, the drum-buffer-rope scheduling system is applied to obtain the final schedule.

First step of the DBR algorithm, namely developing the drum, consists in establishing the schedule on the constraint for all the orders. Then, we subtract the time buffer, equals to the manufacturing lead-time of 1 week plus a 2 days security, to obtain the starting times of the gating operations (developing the rope). The rope implements for the gating operations the schedule of the constraint, offset by the time buffer length. The final schedule consists in following these dates, keeping the FIFO order (First-In-First-Out) as a priority rule. For more details, we refer to the book of Goldratt and Fox (1992).

This schedule, establishing the manufacturing schedule and a shipping list as well, has to be defined within the weekly production meeting. This scheduling method has multiple advantages. Firstly, the dominant flow from the customer to the customer will ultimately become synchronized. Secondly, the shipping list exists 10 days in advance of the shipping period, thus allowing having low transport costs and simultaneously giving the opportunity for grouping BMA’s orders with other shipments, thus avoiding bad fill rates.

4. RESULTS

The development and implementation of the planning and scheduling system took 3 months. The lead-time was mainly due to the flow mapping, flow analysis and simplification. The results of the project are listed in Table 1.

5. CONCLUSION AND PERSPECTIVES

It is certainly neither the first nor the last time drastic bottom-line improvements were possible by implementing an improved and adapted planning and scheduling system. In this case, results could be obtained in a very short time. This is particularly due to the fact that the approach had been based on the principles of the theory of constraints thus aiming global optimum since the beginning of the project. The first three steps of the basic process helped in making the right decisions and accelerating the implementation phase. BMA was losing money before the planning and scheduling model was implemented. After the implementation BMA was making money. Improvement has been made however regarding the WIP reuse level. Furthermore, the replenish-
Table 1 Results

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A logistics function</td>
<td>Non existent</td>
<td>(synchronization head)</td>
</tr>
<tr>
<td>Definition of performance</td>
<td>No indicators implemented</td>
<td>indicators</td>
</tr>
<tr>
<td>Development of a planning</td>
<td>No tool available</td>
<td>and scheduling tool</td>
</tr>
<tr>
<td>A cutting heuristic for</td>
<td>No tool available</td>
<td>optimal material utilization</td>
</tr>
<tr>
<td>Customer service:</td>
<td>Service level: 34%</td>
<td>87% (steady during 6 months)</td>
</tr>
<tr>
<td>Reduction of transportation</td>
<td>Confidential (but very high)</td>
<td>costs by 25%</td>
</tr>
<tr>
<td>Lateness gravity to nearly 0</td>
<td>From 21 to 182 days</td>
<td>euros / days</td>
</tr>
<tr>
<td>Manufacturing lead-time</td>
<td>50% of each mother roll</td>
<td>stable at 10 days</td>
</tr>
<tr>
<td>Potential to reduce WIP to</td>
<td>No planning methodology</td>
<td>0</td>
</tr>
<tr>
<td>TOC based generic planning</td>
<td>Just-in-case system</td>
<td>model</td>
</tr>
<tr>
<td>DBR based synchronized</td>
<td></td>
<td>production system</td>
</tr>
</tbody>
</table>

ment system might be improved through a better supplier synchronization. In the particular case of dealing with inflexible suppliers working with long manufacturing campaigns, the vendor lead-time is important and not very reliable. Therefore it was suggested, as a first step, to qualify a few new suppliers and to try to exploit or to use their phasing in campaigns. In this case, the aggregated lead-time can be reduced by differentiation. Ultimately the best solution will still be the full synchronization of the entire supply chain as viewed from BMA.

Postulate: Unfortunately the excellent results and performances decayed, after having remained stable for 10 months. This fallback is, according to one BMA manager, due to a partially unmanaged change process. This failure concerns the entire project team since a consequent resistance came from the production supervision. The presumed win-win situation, transforming a fire fighter in a production team coach was not perceived as such by the production supervision. This is also one of the reasons why, even since the first implementation actions, the material optimization procedure could not be implemented despite excellent simulation results. The question “what went wrong” needed to be answered. One element is that the risk evaluation was probably incomplete. The process of detecting negative branches was not done sufficiently (Scheinkopf 1999). Another element is without doubt, that the main project driver within BMA left the company. This means that there was no actual appropriation and capitalization process, taking place in addition to the responsibility of sustainability incurring in the project team, internally at BMA. Change projects should include this part as well. The model is being implemented in a floor-manufacturing environment with partially continuous and partially discrete processes. First results are fortunately very promising. Risk evaluation has been particularly improved.

References

Scheinkopf, L. J. (1999). Thinking for a change, St. Lucie Press.