ANNALS of the ORADEA UNIVERSITY.
Fascicle of Management and Technological Engineering, Volume VII (XVII), 2008

APPLYING THEORY OF CONSTRAINTS IN THE PRODUCTION MANAGEMENT

Adriana FOTA
Transylvania University of Brasov
e-mail: fota.a @unitbv.ro

Key words: theory of constraints, management, production control.

Abstract: Theory of constraint (TOC) is a management philosophy that advocates that every system has at least one constraint which limits the system from achieving high level performance. General principles of TOC can be applied successfully to improve performance of manufacturing and service organizations. Aim of the study is to demonstrate how a method of TOC is used to improve performance of the production management. According to this theory, on industrial level, the company has to focus itself on constraints. Constraints may have very different natures, but we are interested here, especially in capacity constraints.

1. INTRODUCTION

The theory of constraints is the most productive methodology of process synchronization. The achievements of companies applying TOC (the originator Dr. Eliyahu Goldratt) results prove that TOC is one of the most effective business management methodologies in the world. It is proved by practical achievements that TOC is more efficient than other management methodologies, e.g. JIT (Just-In-Time), TQM (Total Quality Management), LEAN (Toyota Production System), Six Sigma etc. By using TOC many companies in the world increased their profitability, reduced inventories, found new profitable markets and got out of bankruptcy. The practical study held by APICS (The Association for Operations Management) showed, that using Six Sigma and LEAN together with TOC provides the results which is 24 times more effective than using only Six Sigma and 10 time more effective than using LEAN alone. TOC is a global theory, which claims, that there must be a constraint in every system, and the goal of business organization is to earn money.

Success stories: 1. Boeing Inc. -There was achieved a 75 % reduction of lead time; 60% reduction of WIP inventories; the increase in throughput - 150%, delivery rate approaches 100%; 2. Motorola Inc. - Throughput increased by 150%; the cycle were reduced by 20%; that helped to free excess resources. The reduction of cycle time impacted the much faster adoption of new technologies; 3. Ford Motor Company - The lead time was reduced by 60%; there were 50% less of quality defects; the efficiency of investments increased by 20% and all above helped to save about $100 dollars.

2. COMPONENTS OF THE THEORY OF CONSTRAINTS

Theory of constraints consists of separate, but related processes and interrelated concepts, including the following: the performance measures and five focusing steps, logical thinking processes, and logistics.

2.1 Performance measures

According to [3], there are three key performance measurements to evaluate: throughput, inventory and operating expense. TOC emphasizes the use of these three global operational measures rather than local measures (e.g., efficiency and utilization). Goldratt places the greatest importance on increasing throughput. Throughput is defined

as the rate at which the system generates money through sales, not through production. Goods are not considered an asset until sold. This contradicts the common accounting practice of listing inventory as an asset even if it may never be sold. Goldratt has advocated a new accounting model as an alternative to traditional cost accounting procedures and measures. Inventory is defined as the money invested in goods that the firm intends to sell or material that the firm intends to convert into salable items. The concept of value-added and overhead are not considered. Operating expense includes all the money the firm spends converting inventory into throughput. The objective of the firm, therefore, is to increase throughput and/or decrease inventory and operating expense in such a way as to increase profit, return on investment, and cash flow (more global measures). In *The Goal* [3], Alex explains to Jonah that his plant's use of a robot has resulted in a thirty six percent improvement in one area. Jonah then asks if Alex is now able to ship more products, and if he has fired any employees or reduced inventory as a result (in other words, whether increased throughput, reduced operating expense, or reduced inventory resulted). When the reply was no, Jonah questions how there can be any real improvement; and of course, there can't.
Increasing throughput and/or decreasing inventory or operating expense should lead to the accomplishment of the firm's goal: to make money now as well as in the future. Anything that prevents a firm from reaching this goal is labeled as a constraint. Constraints may appear in the form of capacity, material, logistics, the market (demand), behavior, or even management policy. TOC thinking regards all progress toward the goal of making money as relating directly to management attention toward the constraint(s). The marginal value of time at a constraint resource is said to be equal to the throughput rate of the product processed at the constraint, while the marginal value of time at a no constraint resource is said to be negligible.

2.2 Five focusing steps

The five focusing steps are a tool Goldratt developed to help systems deal with constraints. These steps ensure improvement efforts remains on track towards system-level improvements. Dettmer [1], believes that these are collectively the most important aspect of TOC.

TOC's five focusing steps are:

Step 1: Identify the system's constraint(s).
Step 2: Decide how to exploit the system's constraint(s).
Step 3: Subordinate everything else to the decisions made in Step 2.
Step 4: Elevate the system's constraint(s)
Step 5: If a constraint is broken in Step 4, go back to Step 1, but do not allow inertia to cause a new constraint.

The orientation of TOC is toward the output of the entire system, rather than a look at a discrete unit or component. The five focusing steps assist with identifying the largest constraint that overshadows all of the others. These steps constitute an iterative process. As soon as one constraint is strengthened, the next weakest link becomes the priority constraint and should be addressed. Thus, a process of ongoing system improvement is applied to the business practice of the firm.

2.3 Logical thinking process

Goldratt introduced [3], a staged logical thinking process to be used in conjunction with the five focusing steps. The thinking process assists with working through the change process by identifying the following: what to change, what to change to, and how to effect the change. The thinking processes consist of logic tools used to identify problems, then develop and implement solutions. These tools include effect-cause-effect (ECE) diagramming and its components: negative branch reservations, the current reality tree, the future reality tree, the prerequisite tree, the transition tree, the evaporating cloud, the negative branch reservation, and the ECE audit process. These tools allow an organization to analyze and to verbalize cause and effect.

The following is a brief description of the thinking process. A current reality tree, a cause-effect diagram, is drawn in order to discover the problems. These problems are known as undesirable effects. The cause of an undesirable effect is known as a root cause. The first goal is to find the causes of these undesirable effects. Each statement in a current reality tree that is not a derivative of another must be a root cause. If you build a tree that is comprehensive enough, at least one root cause will lead to most of the undesirable effects. This particular root cause is labeled a core problem, the major improvement target. The fewer root causes responsible for the undesirable effects, the better. The solution to this core problem is apparently not readily available. If it were, then the problem would have already been solved. Some conflict, therefore, must exist that prevents an immediate solution. This conflict becomes evident upon the construction of an evaporating cloud.

An evaporating cloud is a conflict-resolution tool. The process begins with a statement of the desired objective, one that is the opposite of the core problem. Then, the prerequisites necessary to achieve the requirements are listed. Any conflicts and assumptions that exist between the prerequisites are verbalized. For example, if one objective is to increase profit, then the requirements may be to improve the product and to decrease expenses. Prerequisites for each, respectively, might be to increase expenditures on capital equipment and to decrease expenditures, two obviously conflicting elements. The best solution is to remove the conflict; a compromise is not desirable. The next move involves finding an injection, a breakthrough idea that will evaporate the cloud. The "evaporating" refers to the tool's ability to dissipate conflict and to create a win-win solution. Usually, the original injection is not sufficient to fully solve the problem, but additional needed injections become clear when building the future reality tree.

A future reality tree is another cause-effect diagram. The tree starts with the proposed solution to the core problem and delineates the injection(s) and the ensuing desirable effects. The future reality tree is a "what if." It provides the opportunity to evaluate and to improve a solution before it is implemented. It is noted that one should be careful not to allow the solution to cause new undesirable effects.

A prerequisite tree describes the implementation of the injection(s) and is composed of an obstacle and an intermediate objective. This diagram breaks the implementation tasks into smaller increments, noting expected obstacles and intermediate objectives whose accomplishments will overcome the obstacles. The intermediate objectives are sequenced, displaying the necessary order of accomplishment and determining which ones can be achieved in parallel. This tool is powerful in that it does not ignore the obstacles. It uses them, rather, as the main vehicle for this phase.

Finally, a transition tree or implementation plan is constructed. This element presents a detailed description of the gradually evolving change envisioned. This task forces one to carefully examine which actions are really needed and if they are sufficient to guarantee the required change.

The thinking-process tools are powerful resources when used effectively. They have found successful use in the logistics and medicine areas of the United States Air Force, in primary education, and in the service sector. James Cox and Michael Spencer, both college professors and "Jonahs," state in *The Constraints Management Handbook* that the thinking processes may be the most important management tools developed this century.

2.4 Logistics

Logistics in TOC include drum-buffer-rope scheduling, buffer management, and VAT analysis. The Drum-Buffer-Rope model (DBR) is a method to identify and exploit the constraints in a production system. It uses Process Mapping, a popular process improvement tool, as a main tool to identify the bottleneck and then apply solutions.

The DBR demonstrates that a manufacturing facility will only produce as much as it's most limited resource or constraint allows. These constraints or bottlenecks can be displayed as:

A. Drum – Buffer – Rope. Drum-buffer-rope is a TOC production application and the name given to the method used to schedule the flow of materials in a TOC facility. Each component will be define as follows, [5]:

- Drum. The drum is the constraint and therefore sets the pace for the entire system. The drum must reconcile the customer requirements with the system's constraints. In simpler terms, the drum is the rate or pace of production set by the system's constraint.
- Buffer. A buffer includes time or materials that support throughput and/or due date performance. A buffer establishes some protection against uncertainty so that the system can maximize throughput. A time buffer is the additional planned lead time allowed, beyond the required setup and run times, for materials to reach a specified point in the product flow. Strategically placed, time buffers are designed to protect the system throughput from the internal disruptions that are inherent in any process. A stock buffer is defined as inventories of specific products that are held in finished, partially finished, or raw material form, in order to fill customer orders in less than the normal lead-time. Stock buffers are designed to improve the responsiveness of the system to specific market conditions.
- Rope. The rope is a schedule for releasing raw materials to the floor. The rope is devised according to the drum and the buffer. The rope ensures that non-capacity constraint resources are subordinate to the constraint. Restated, the rope is a communication process from the constraint to the gating operation that checks or limits material released into the system to support the constraint.

B. Buffer management. Buffer management provides the means by which the schedule is managed on the shop floor. Buffer management is a process in which all expediting in a shop is motivated by what is scheduled to be in the buffers (constraint, shipping, and assembly buffers). Buffers can be maintained at the constraint, convergent points, divergent points, and shipping points. By expediting this material into the buffers, the system helps to avoid idleness at the constraint and missed customer due dates. Also, the causes of items missing from the buffer are identified, and the frequency of occurrences is used to prioritize improvement activities.

C. VANT analysis. VAT analysis determines the general flow of parts and products from raw materials to finished products. It conceptualizes an organization in terms of the interaction of its individual component parts, both products and processes. Three general categories of production structures result from this standpoint, each necessitating a unique approach to management planning and control.

In Figure 1 is presented a Drum-Buffer-Rope model in the production management system of a woodworking company. The Drum is Cutting process with less throughput of

40 parts per hour. A buffer level is determined, a detailed work schedules (Rope) is released to keep the Drum running smoothly and produce more with less.

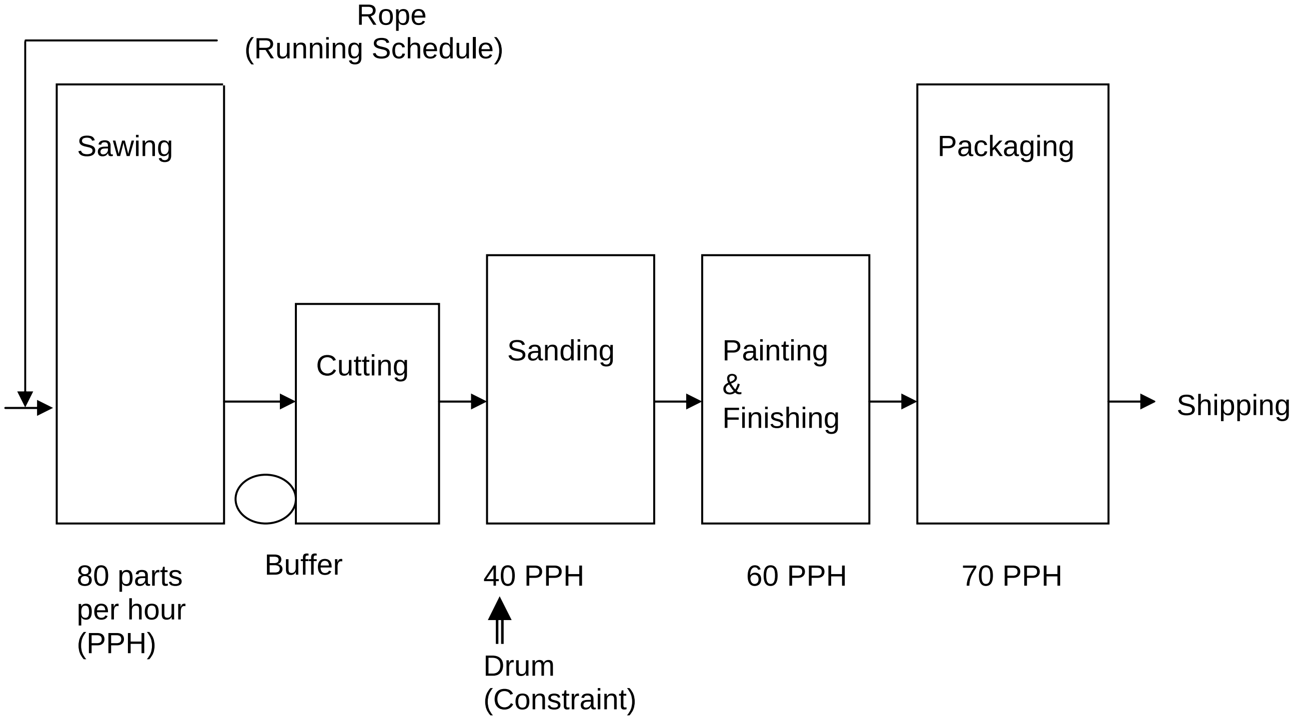

Fig.1. A Drum-Buffer-Rope model in the production management

The logical structure [2] is the sequence of operations through which each product must pass in order to manufacture and assemble a product or product family (figure 2). A V logical structure starts with one or a few raw materials, and the product expands into a number of different products as it flows through its routings. The shape of an A logical structure is dominated by converging points. Many raw materials are fabricated and assembled into a few finished projects. A T logical structure consists of numerous similar finished products assembled from common assemblies and subassemblies. The graph shows the general appearance of each structure. Once the general parts flow is determined, the system control points (gating operations, convergent points, divergent points, constraints, and shipping points) can be identified and managed. This determination focuses management's attention on a few control points where buffers can be used to protect and to maximize throughput. Five control points are used to manage the process: (1) the constraint, (2) the points of divergence (where a part or material is diverted to different routes in order to make different products), (3) the points of convergence (where two or more parts are combined in subassembly), (4) the gating operation (releases work into the shop), and (5) the shipping operation.

The shape of the structure determines which control points are utilized to manage production. A *T* structure focuses attention on the constraint and the gating operation. The five-step focusing process is used to manage the constraint with a buffer placed before the constraint to absorb variations in the process. The output from the gating operation is tied to the constraint; that is, since the constraint controls the amount of throughput; the gating operation cannot process more than the constraint.

A *V* structure also uses a buffer to protect the constraint and the gating operation releases orders at the same rate as the constraint as seen in the *T* structure. However, an additional control point exists in the *V* structure, the divergent point. The divergent point is

controlled by a schedule derived from the shipping schedule. This derivation prevents misallocation of material to a product not currently in demand.
The *A* structure also manages the constraint and gating operation in a fashion similar to the *T* structure. Any diverging points are scheduled in accordance with the shipping schedule. In addition, an assembly buffer is used to maintain the flow into the convergent points. An additional schedule based on the shipping schedule (similar to that used in the *V* structure) is used to keep capacity from being misallocated to the wrong order. By using VAT analysis, significant improvements in the production process can result.

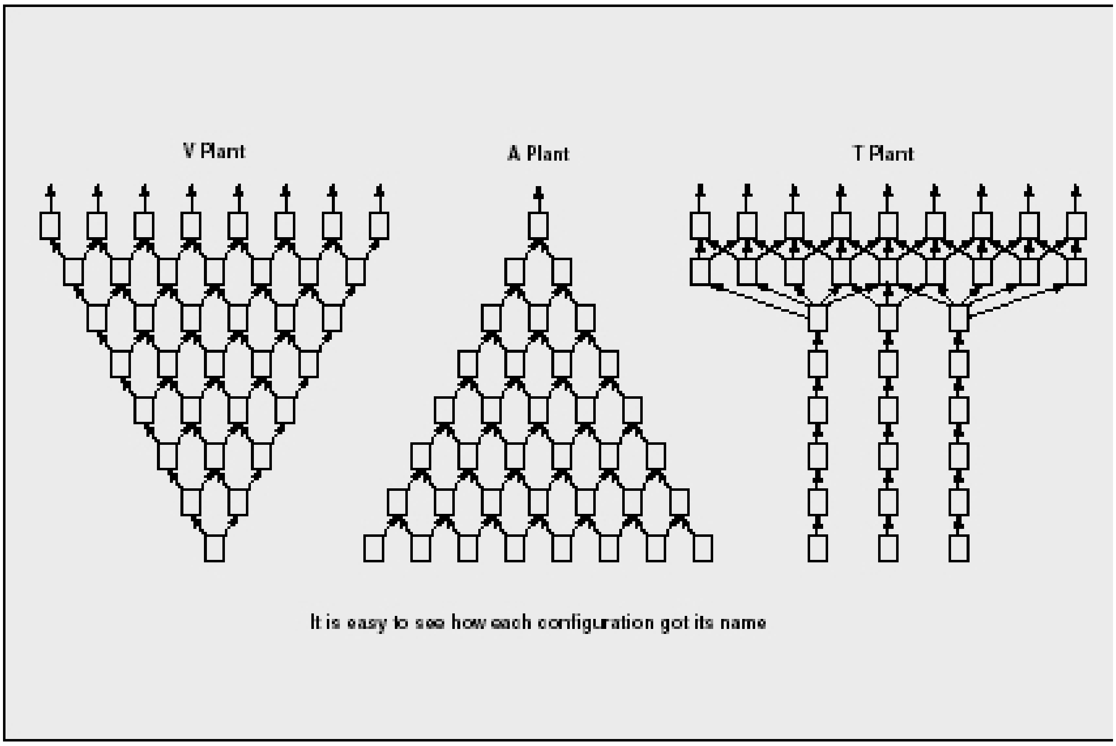

Fig. 2. The logical structure

3. BASIC CONCEPTS OF THE CONSTRAINTS THEORY

According to this theory, on industrial level, the company has to focus itself on constraints. A constraint means a limiting factor for the performance of a system. Constraints may have very different natures, but we are interested here, especially in capacity constraints (limitation). Thus, a capacity constraint may be described as a mean of production, which has a less or equal capacity to the load resulting from a request.
Plants' unbalance imposes generally the appropriation of a dual view, distinguishing two types of resources and treating them in a different manner: constraints (which have average capacities approximately equal or inferior to the needs) and non-constraints (which have average capacities surpassing the needs).

On economic level the constraints theory is guided by the principle of profit as main aim of the whole industry. It is possible to express this aim from a financial point of view in two manners:

- Raising net benefit by improving simultaneous the efficiency of investment and treasury. Generally, this aim is expressed according to more tangible in and close to the data of industrial culture.

- Raising the added value issued by sells by reducing global load to be covered. This means thus an increase of product selling, by simultaneous reducing stocks and operating costs.

This method is based on three operational indicators:

- Sold production output (*Throughput* -T): corresponding to the money produced by the production system through its sells
- Stock (*Inventory* - I): meaning invested money in buying materials proposed by the production system to be sold
- Operating costs (*Operating Expenses* - OE): representing money consumed by the system for transforming I on T.

Relating the above three indicators with the classical financial indications, featuring the company's operation (figure 3): net result, revenue on investment and treasury, it results:

- Net result: T-OE
- Revenue on investment: (T-OE) / I.
- Treasury.

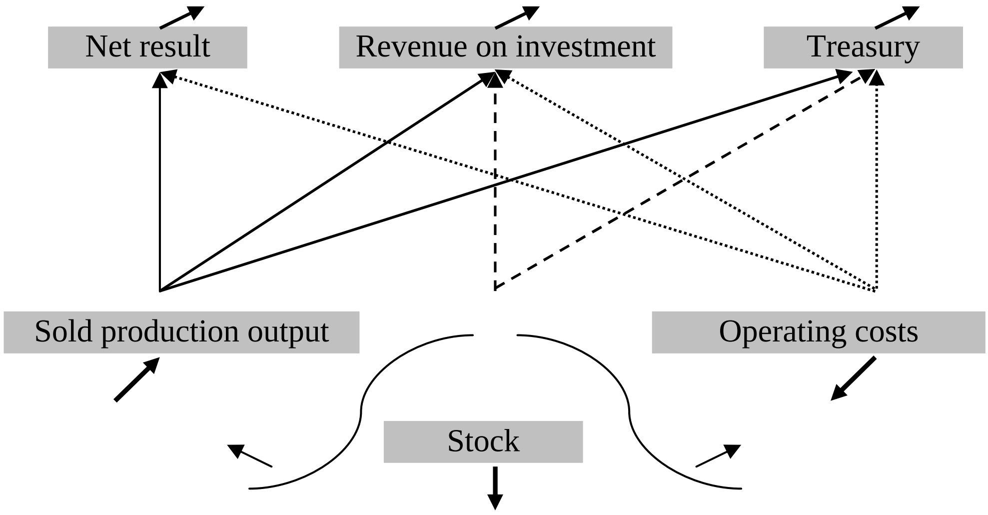

Fig.3. Value shaft

This shaft gives the impression that the stock indicator represents an important performance criterion. By this analysis, the presence of a constraint shall have as consequence the limitation of the sold production output and a raise of the upstream stock. A detailed presentation is made by [4].

The philosophy of the TOC method consists in nine main rules:

Rule 1: Balancing streams, not capacities. The test to balance capacities ends by prolongations, stock increasing and stream reducing. If a process is perfectly balanced, its resources have neither capacity margin, nor excess for compensating a delay. The system may react to the most insignificant random phenomenon, disturbing the stream.

Rule 2: The using level of a non-restrictive resource is not determined by its own potential, but by other constraints of the system. It has not to be studied the maximal use of non-restrictive resources. These, have to be regulated and constraint by a limitation. If not activated beyond this limitation, waste stocks are created. Thus it may be noted, that the old notion of local productivity is resumed.

Rule 3: Maximal usage and use of a resource are not synonymous.

Rule 4: A lost hour with a restrictive resource is a lost hour for the whole system. The entire process being rhythmic by the capacity of the restrictive resource, any loss in its time, is a loss for the whole system and any casting in time in a large space has no impact. This confirms the importance to be granted to the restrictive resource. All means for improving the performances resulted from the constraint have to be searched, avoiding losses in production time, by using substitute resources, remedying or reducing spoilage after constraining.

Rule 5: A gained hour with a restrictive resource is an illusion.

Rule 6: Constraints determine the output flow on the stocks level. The theory searches to reduce stocks but doesn't has in view to eliminate them, because as we saw it above, the time lost with the restrictive resource is lost for the whole system: the stock may be perceived as a damper of random aspects.

Rule 7: Often, the transfer batch doesn't have to be equal to the manufacturing batch. The duality specific to the management by constraints leads to managing rules such as the original batches: two batch-types are identified: manufacturing batch (the assembly of items treated by the same resource between two serial changes0 and the transferring batch (the quantity of pieces transferred from an operation to the other). The production batch size from a restrictive resource has to be determined dynamically to synchronize the activity.

Rule 8: Manufacturing batches have to be variables and not fixed. For accelerating the above mentioned product flow by keeping an identical manufacturing time, manufacturing batches have to be fractioned in more transfer batches. The production batch size within a large space, as well as that of all transfer batches have to be minimized. The production batch sizes, according to this method, have to be adapted to the constraints and circumstances to allow a better adaptation to commands and to reduce waiting times. For an ordered work, the most usual method is creating buffer stocks. But it is not applied systematically, because it hides the problems.

Rule 9: Fixing schedules simultaneously having in view all constraints. Manufacturing deadlines are a result of a schedule and thus, they may not be determined. Often, production management limits the plant's performances, supposing that manufacturing cycles have fixed and pre-determined lengths. Constraints management deduces the lengths of their ordering cycles; it is really enough to describe the operational sequences on these few resources, for previewing the plant's performances. Constraints attribution controls the flow, keeping the schedule and, partly, the stocks level.

4. THE CAPACITY CONSTRAINTS RESOURCES

The above mentioned capacity constraints resources are restrictive resources, i.e. a resource with average capacities equal or inferior to the capacity. Generally, there are more capacity constraints types. Thus it may be described the CCR (Capacity Constraint Resource) which designates all resources, which, administrated incorrectly, may alter the production schedule. Nonrestrictive resources administrated incorrectly may become CCR and thus, apparently restrictive resources.

For a clearer explanation, we take an example.

4.1 The case of a company

We shall consider a plant disposing of four machines M1, M2, M3 and M4, for manufacturing two products A and B. The typical request is a combination of 2A and 5B.
The plant operates in three shifts and (the time it is open amounts 24 hours). Table 1 indicates the time allotted to each component of the combination.

Tab.1. The time allotted to each component of the combination

	M1	M2	M3	M4
A	3	2	9	1
B	5	4	1	1

For determining what resources are CCR, we shall check the capacity, the test profile and the ratio: test/capacity per time unit.
For a typical request consisting of 2A+5B:
- Test's M1: (2 x 3) + (5 x 5) = 31 hours
- Test's M2 = 24 hours
- Test's M3 = 23 hours
- Test's M4 = 7 hour.

Knowing that the daily available time is here 24 hours, M1 and M2 are restrictive, M3 and M4 not. According to the definition of a constraint, M1 determines the global flow (stream), meaning that any issue on M1 reflects itself on the whole flow of the stream. M1 is a CCR, when M2, as restraint space by definition, shall not be an issue, since it shall be supplied by M1.

Representing as scheme the different machine occupying times and observing the order of passing from a machine to an other, we find that the production schedule is completed within 37 hours if the command treating order is A, followed by B (figure 4).

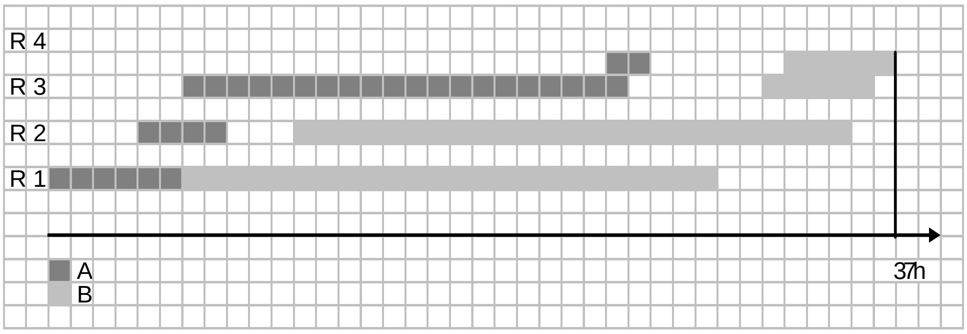

Fig.4. Representation of different occupation times of the machine A, than B.

In return, if the order is first B and than A, the production schedule is completed within 51 hours (figure 5).

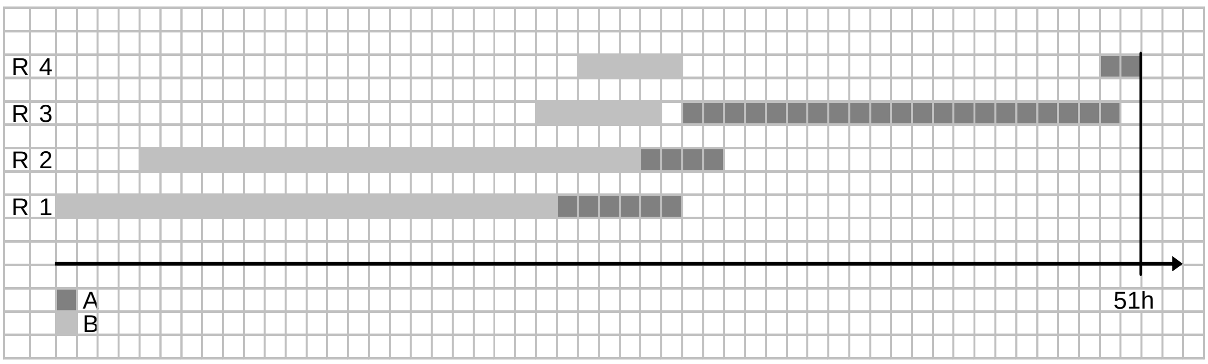

Fig.5. Representation of different occupation times of the machine B, than A.

With these two information, if the commercial department promises the delivery of the registered combination in the production schedule within a 38 hours term, this promise will observed if the order is A followed by B, but if the order is B followed by A, this order will not be ready on time.

In this last case, ignoring the influence of M3 will delay with 14 hours the production completion. Thus, M3 is a CCR. Though M3 is not a restrictive resource, its influence may become essential if no attention is paid to the priority choice between A and B.

5. CONCLUSION

Applying theory of constraints offer main benefits:

- *Increase revenues.* Assuming that the market demand exists, by eliminating the most significant constraint, the company will be able to generate higher revenues by producing a higher volume.
- *Reduce costs per unit produced.* If many processes or departments within a company have capacity which isn't fully being utilized, it means that fixed costs associated with those processes or departments are not being used efficiently. When the company's total throughput increases as a result of eliminating the most significant constraint, the fixed costs can be spread over a larger number of units and therefore the fixed costs per unit of throughput are lower. There are also other hidden costs associated with significant constraints – such as management time spent dealing with the negative consequences of the constraint, or increased wastage due to overproduction at the processes upstream of the constraint.
- *On-time delivery.* Often the part of the company which is the most significant constraint is not as controlled or predicable as the other processes within the company. Consequently, the constraint itself is often the source of shipment delays. As such, removal of the most significant constraint can lead to more predicable lead times and better on-time delivery performance.
- *Allocate management's time in a way that has the highest return.* By focusing on the most significant constraints, management's time will be allocated towards solving whatever problems are likely to have the most significant positive impact, at least in the short term.

REFERENCES

[1] Dettmer, H. William. *Goldratt's Theory of Constraints: A Systems Approach to Continuous Improvement.* Milwaukee, WI: ASQC Quality Press, 1997.
[2] Fota, A., Sarbu, F, A., *Production management through enterprise resource planning (ERP) instruments,* Revista RECENT – Industrial Engineering Journal, vol. 8 (2007), no. 3b (21 b), ISSN 1582-0246, pp. 464-473.
[3] Goldratt, Eliyahu M., and Jeff Cox. *The Goal: A Process of Ongoing Improvement.* 3rd ed. Great Barrington, MA: The North River Press, 2004.
[4] Marris, P., *La Gestion par le contraintes,* editions d'Organisation, 1994.
[5] McMullen, Thomas, B., Jr. *Introduction to the Theory of Constraints (TOC) Management System.* Boca Raton, FL: St. Lucie Press, 1998.