HOW DOES COLLABORATIVE INVENTORY MANAGEMENT MAKE PROGRESS?

Togar M. Simatupang
Department of Industrial Engineering
Institut Teknologi Bandung

:: Abstract ::
Collaborative inventory management (CIM) has revolutionized electronics, textile, apparel, and grocery industries. An intriguing question is how far does the movement of CIM challenge the traditional practice of inventory management? This paper illustrates how to expose and challenge flawed assumptions of traditional inventory management. It also proposes a collaborative replenishment process that consists of a cyclic process of tactical planning, execution, and control. The proposed scheme makes it possible for the chain members to apply collaborative inventory management. The paper also outlines directions for future research.

Keywords : inventory management, supply chain collaboration, logistics, theory of constraints.

1. INTRODUCTION

Collaborative inventory management (CIM) has revolutionized electronics, textile, apparel, and grocery industries in many countries. Different names have been given to CIM such as quick response (QR), efficient-consumer response (ECR), vendor-managed inventory (VMI), and collaborative planning, forecasting, and replenishment (CPFR). CIM serves as an excellent scheme for creating a just-in-time inventory system because it encourages collaborative efforts amongst the chain members to effectively match supply and demand at the lowest cost and avoids the pursuit of antagonistic bargaining over inventory.

Collaborative inventory management creates mush potential for improvement opportunities to dramatically reduce the supply chain pipeline inventory. This pipeline inventory and its associated costs of carrying, capital, and obsolescence is quite large. Fuller *et al.* (1993), for instance, estimated the grocery pipeline alone carried an inventory value of $75-100 billion, which comprised one-quarter to one third of annual sales (estimated at $300 billion). Companies have recognized this opportunity and collaborated with their immediate upstream and downstream links to realize mutual benefits of reducing inventory levels (Simatupang & Sridharan, 2002). For example, General Electric (GE) collaborated with its retailers to respond to actual demand instead of filling demand from inventory (Treacy & Wiersema, 1993). Collaborative inventory management enabled both parties to eliminate inventory holding costs and assemble full truckload orders. GE saved about 12% of distribution and marketing costs and obtained half portion of the retailers' sales. The retailers reduced out-of-stocks and gained increased profit margins on GE products.

An intriguing question is how far does the movement of CIM challenge the traditional practice of inventory management? This paper aims to show how CIM makes significant progress and proposes a generic scheme for applying CIM. The evaporating cloud diagram (ECD) is used to scrutinize the movement of CIM by exposing and challenging flawed assumptions of traditional inventory management. This paper contributes to the literature of inventory management in two ways. First, the paper demonstrates how to invalidate flawed assumptions of traditional inventory management. Second, a collaborative replenishment process is proposed as a guideline for the application of collaborative inventory management.

The paper is organized as follows. The next section describes the scope of collaborative inventory management. Then the dilemma of inventory management is described. Next, methods of solving the dilemma are described. In the subsequent section, a collaborative replenishment process is proposed as a means to apply CIM. The paper ends with conclusions and recommendations for future research.

2. COLLABORATIVE INVENTORY MANAGEMENT

Collaborative inventory management can be defined as two or more chain members agreeing to improve the velocity of inventory in order to maximize revenues at minimum cost. The distribution system shown in Figure 1 is a well-known example of collaborative inventory management between a supplier and a retailer. This distribution system consists of a supplier plant, several distribution centers and many stores as points of sales, with material flowing from supplier, to distribution center, to store in order to satisfy demand at the store. The supplier plant produces a range of products. Distribution centers are required to supply the store very quickly. Stores serve end customers. The retailer is assumed to own distribution centers and stores.

The distribution system can be described as a series of individual links. Each link consists of a process, buffer (base stock), and replenishment procedure. A process is normally composed of one or more tasks. One link carries out all tasks before work and responsibility are passed on to the next link. The task might include production activity that alters the physical form of the product, shipping activity that moves the product to the next link, and packaging. Buffers or base stocks are located behind each process and hold the material that is waiting for the process. Figure 1 contains three strategic buffers: store buffer, distribution buffer, and plant buffer. Replenishment procedure specifies the quantities of material required and the timing of supply from the upstream member.

The replenishment process is described as follows. The demand during each period (day, week, or month) at each store has the highly random uncertainty. The end customer enters the store to buy a product. Demand is met as long as there is stock on hand at that store. When a stockout occurs, unfilled demand is backlogged until sufficient replenishment arrives. Every review period (r), the store places a replenishment order with the distribution center so that its inventory position (on-hand plus on-order minus backorders) restores to a fixed base stock level. The distribution center satisfies the store order from on-hand inventory. Demand in excess of available inventory is backlogged. Replenishments are assumed to be delivered to the store at leadtime L_a after being ordered by the store.

Every r review period, the distribution center places a replenishment order with the supplier so that its inventory position restores to a fixed base stock level. The supplier satisfies the distribution center from on-hand inventory. Demand in excess of available inventory is backlogged. Replenishments are assumed to be delivered to the distribution center at leadtime L_s after being ordered by the distribution center.

The main issue of collaborative inventory management is to design and manage the supply chain in a way that requires the lowest level of inventory investment necessary to provide a pre-specified level of customer service (Schwarz & Weng, 1999). The goal is to find values for each site's base stock inventory (buffer stock), batches or lot sizing, and the average total leadtime that minimize the total expected costs for inventory-holding, lost sales, backlogs, and transportation costs. However, this goal is elusive and so the dilemma of inventory management persists.

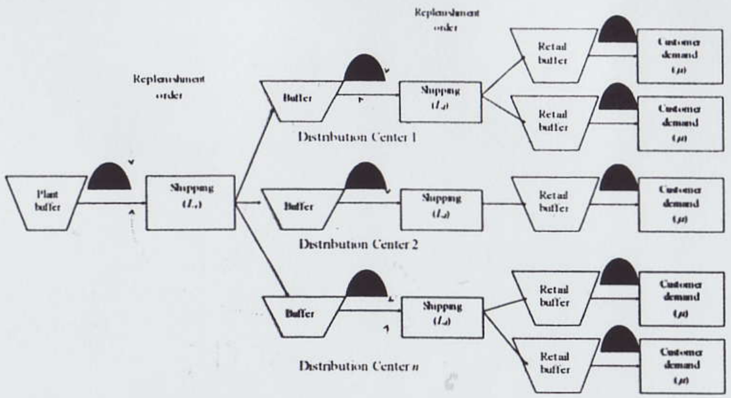


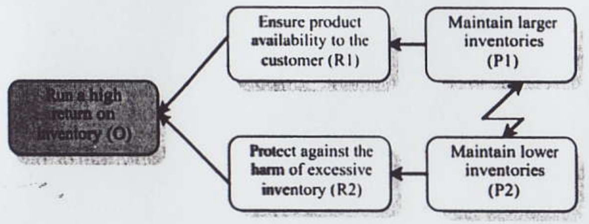
Figure 1. Collaborative inventory management between the supplier and the retailer

3. THE DILEMMA OF INVENTORY MANAGEMENT

The evaporating cloud diagram (ECD) is used to scrutinize the movement of CIM by exposing and challenging flawed assumptions of traditional inventory management (Goldratt, 1990). Figure 2 captures the dilemma of inventory management faced by a retailer in the supply chain. Other upstream members also experience this dilemma. Inventory management is a means of matching demand and supply by protecting sales from demand and supply fluctuations so the system can attain a high return on every dollar of inventory investment. In order to run a high return on inventory investment, the retailer depends upon two requirements: "ensuring product availability to the customer" (R1) and "protecting against the damage of excessive inventory" (R2). Product availability means a customer can obtain the desired products off the shelf when shopping at the store. The disadvantages of excessive inventory include high carrying costs, high capital costs, obsolescence, large-lot quality costs, and reduced-capacity costs.

The requirement to "ensure product availability" needs specific action - "maintaining larger inventories" (P1). The retailer holds larger inventories to reduce certain costs such as ordering costs, stockout costs, and acquisition costs. The retailer also often takes advantage of forward buys and vendor deals such as discounts and promotions. On the other side of the diagram, the requirement to "protect against the harm of excessive inventory" requires specific action - "maintaining lower inventories" (P1). The problem of reducing inventory becomes obvious as one side of the conflict recommends the retailer to "maintain larger inventories", while the other side encourages the retailer to "maintain lower inventories" (see the conflict arrow between P1 and P2 in Figure 2). The important issue becomes how much inventory to hold.

There are two fundamental decisions in inventory planning: how much to order (order quantities) and when to place the orders (order points). Order quantities and order points determine the amount of inventory at any given time. The traditional solution to determine order quantity is to compromise the conflict arrow. Order quantity is placed so that the cost of ordering too much is balanced against the cost of ordering too little on each order. For example, economic order quantity (EOQ) is determined by balancing carrying costs against ordering costs so that the total stocking costs are at minimum.



When setting order points (OP), the retailer faces an uncertain demand during lead time. Safety stock is carried to cover demands for a product during lead time so that a stockout can be minimized. If too much safety stock is carried, the carrying costs become excessive. When too little is carried, the stockout cost becomes excessive. The compromise is sought to balance these two costs in setting order points. In determining the relationship among variables in setting order points and safety stock, it is known that order point equals expected demand during lead time (EDLT) plus safety stock (SS). Since the difficulty in estimating the costs of stockouts, service level (*SL*) is often used to measure the probability that a stockout will not occur during lead time. In other words, the probability of a stockout () equals to 1-*SL*. For example, a 95 percent service level means a 95 percent probability that all orders can be immediately filled out of inventory or there is 5 percent of probability that all orders cannot be filled from inventory during leadtime. To set order points, service level is determined by management policy. The order point becomes the demand during leadtime that has stockouts about percent of the time. In the formula, $OP = EDLT + Z(\sigma_{LT})$. *Z* is the number of standard deviation that the order point is away from the expected demand during leadtime. The amount of safety stock is then the order point minus the expected demand during leadtime.

However, the pressure to maintain larger inventories is more prevalent due to influences from upstream members in terms of price and volume discounts, and from end customers in terms of higher service levels, and thereby the retailer tends to rely on the logic of P1 to R1 rather than following the compromise logic. The main assumptions behind the retailer holding large levels of inventory are inaccurate demand forecasting and long replenishment times. Tolerating inaccurate forecasting means that the members along the supply chain produce or buy goods far in advance based on demand forecasts during lead time. Demand during leadtime is the amount of a product that will be demanded while waiting for a new order to arrive and replenish inventory. There are two sources of the variation in demand during leadtime. First, demand for a particular product is subject to great daily variation. Second, the leadtime to deliver an order is subject to variation. For instance, the supplier can have difficulty in procuring raw materials or may suffer machine breakdowns, and the truck company can have unexpected disruptions that delay deliveries.

The variation in demand during leadtime often makes each chain member set minimum and maximum levels of inventory. The inventory on-hand (actual stock plus any outstanding replenishment orders) is often reviewed periodically. The order is placed only when at review the inventory on-hand reaches or falls below a minimum level. The order size is equal to the maximum level less inventory on-hand. If the inventory on-hand is above the minimum level, no replenishment order is placed. This ordering policy is not based on when an item is actually sold. Pushing the inventory toward end customers has little to do with actual demand fulfillment and thereby wastes resources through the retention of more inventory at the store. Additionally, the combination of large inventories and inaccurate forecasts leads to mismatch between the retail inventory and the customer demand.

Although the retailer has a large inventory, stockouts often occur for certain items.

Another prevalent assumption underlying P1 to R1 is that the order quantity must be immediately delivered in one shipment. Shipping in large quantities means cheaper transportation costs. Many suppliers optimize the transportation costs by offering the retailer a free delivery if placing orders within the order maximum and minimum. As a result, the retailer should receive and maintain large inventories which are expected to be sold in the future eventually. However, maintaining large inventories ties up cash and exposes the retailer to prohibitive costs of obsolescence, especially for seasonal and innovative products (Fisher, 1997).

4. BREAKING THE DILEMMA OF INVENTORY MANAGEMENT

The presence of an arrow of relationship in Figure 2 indicates the existence of hidden assumptions between entities described within the diagram. The key to unlocking the conflict is to expose all hidden assumptions about the prerequisite-requirement-objective relationship and examine those which are not valid. The chain members will be stuck in compromise unless they challenge hidden assumptions. Injections (actions or conditions) need to be created to break or invalidate the underlying assumptions in a way that is adequate to resolve the problem. Applying this concept to breaking the dilemma of inventory management, this section attempts to identify and challenge invalid assumptions.

Table 1 presents a list of underlying assumptions behind each arrow. There are many other hidden assumptions that can be discovered. However, the assumptions presented are adequate to break the dilemma. The dilemma can be broken by attacking several underlying assumptions behind the logical links. Table 1 also provides injections used to invalidate underlying assumptions.

The link ensuring product availability (R1) and high return-on-inventory (O) assumes that the only way to attain higher return on inventory is by ensuring availability to generate sales. However, sales only occur when end customers buy products from the store. The concept of availability needs to be revisited depending on the position of the chain member. Only availability at the store can generate more sales into the supply chain and thereby contribute to return-on-inventory. Collaborative inventory management recognizes that only availability of products at the store maximizes revenues.

Underlying assumptions between P1 and R1 might be: the leadtime is too long, the order quantity must be immediately delivered in one shipment, the distribution center requires large inventories to distribute to many different stores, demand forecasts are inaccurate, and products out of stocks mean lost sales. All these assumptions can be countered.

CIM invalidates the assumption that leadtime is too long by strategically locating inventory buffers to protect strategic processes such as production capacity, distribution, and shelf life. For example, a distribution center is useful where the production leadtimes are large compared to the delivery times. Inventory buffers serve as de-coupling points that divide long leadtime into smaller leadtime segments. Smaller leadtime segments enable the previous link in the supply chain to achieve speedy delivery to replenish products when stock has been drawn from the buffer by the next link downstream. The length between two stages is measured by the maximum time required to reliably replenish, and the frequency of replenishment. The size of each strategic buffer thus depends on the level of consumption it services and the time taken to replenish it.

CIM breaks the assumption that the replenishment of stock is based upon the forecast of future consumption. This assumption about the replenishment process takes place between the supplier and the distribution center, and between the distribution center and the store. The injection is that the replenishment of stock is based on the actual consumption of the products. The order quantity placed with the supplier does not need to be delivered immediately in one shipment, but it can be shipped in smaller batches as needed. The supplier provides rapid and reliable replenishment to the retailer distribution centers. Shipments can be consolidated from the supplier's plant to many distribution centers to create full truckloads. As the supplier is able to ensure rapid and reliable replenishment, the distribution center does not need to hold such large inventory in order to fulfill the demands from the store. This explains how the velocity of product flow can be transferred into lowered inventory levels.

Similarly, the assumption that the distribution center requires large inventories to distribute to many different stores can be countered. If the supplier frequently replenishes whatever the distribution center delivers to the stores, the distribution center only needs to hold enough inventory to replenish what is sold at the stores. The reaction time from the distribution center to the store is not subject to the procurement leadtime but to the transportation time.

Traditional inventory management often tolerates inaccurate demand forecasts and thereby obliges the retailer to hold higher levels of inventory (Eppen & Martin, 1988). It is technically difficult to predict the daily demand for hundreds of individual items at a store. However, the accuracy of the forecast varies depending upon position in the supply chain. Therefore, the injection asserts that the best place to forecast is the point where the largest pooling effect occurs. For example, the distribution center has better forecast accuracy for items at the store level. Likewise, forecasting for the large groups at the plant provides better accuracy compared to the aggregation of products at the distribution center. Demand forecasts of large groups are often disaggregated into items based on historical data. Furthermore, shortening the forecast period increases demand forecast accuracy. However, CIM encourages the chain member to be less dependent on forecasts. Forecasting is only used to predict the replenishment time in reacting to demand changes, or to predict aggregate items for tactical planning.

Another assumption underlying the P1 to R1 link is that stockouts mean lost sales. However, only stockouts at the store lead to lost sales. The upstream members can choose whether to backorder or not when a stockout occurs. Therefore, collaborative inventory management employs low levels of inventories at the store to protect availability from demand uncertainty and variability. A low inventory system is very responsive to demand changes because rapid and reliable replenishment enables the store to quickly respond.

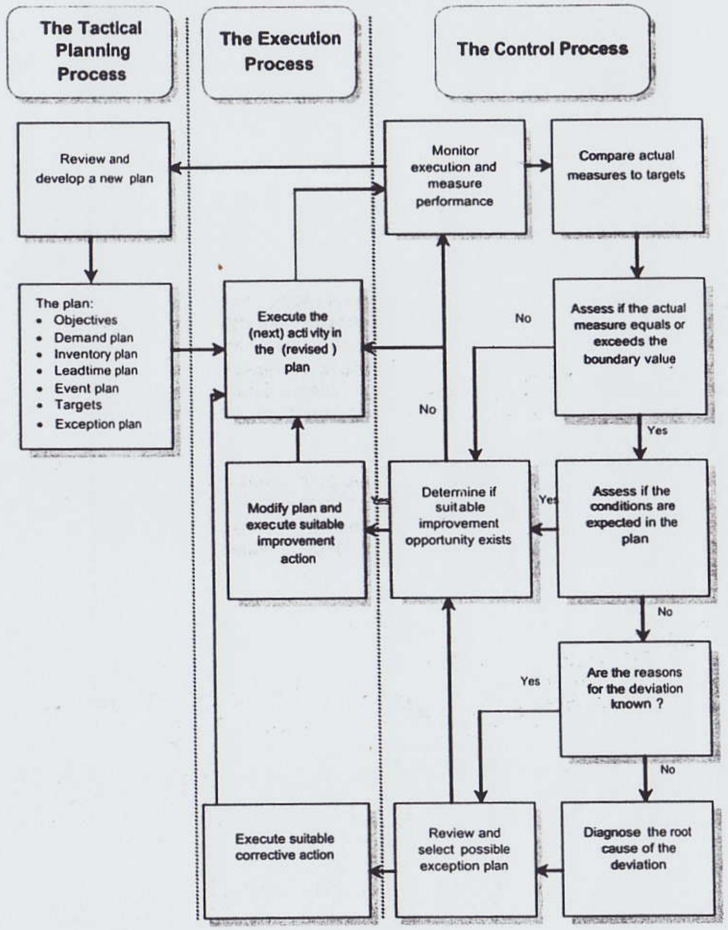
The more contemporary explanation of CIM comes from the proponents of the theory of constraints (TOC) (Goldratt *et al.*, 2000; Simatupang *et al.*, 2004). The TOC approach disputes the conflict arrow (P1 P2). The invalid assumption is that the commitment to locate inventory at a certain place often limits flexibility for further choices. It is not surprising that traditional inventory management often actually has enough inventory only it is located in the wrong place. The injection is that the better place to hold inventory is the point in the supply chain where there are the most choices about where it can be sent or processed. These strategic points along the supply chain include the plant and the distribution center. The distribution center ensures rapid replenishment time to the stores and the plant ensures replenishment time to the distribution center. Holding inventory closer to the plant not only provides supply flexibility to match actual demand with the least amount of investment, but also decreases the replenishment time to the distribution center. What the plant produces should be based closely upon what the market is consuming. Therefore, the market consumption should be triggering the pull to the plant. This pull behavior eliminates larger inventory at the stores, reduces the entire cycle time, and trims down the total capital investment.

After determining where to hold the inventory buffer, the next task is to quantifythe right buffer at each strategic point in the supply chain. The amount of inventory buffer required by every link is the total leadtime demand for *L* time periods plus safety stock to satisfy a demand spike that could occur in the leadtime within which the next upstream link can reliably replenish the items.

Monitoring the buffer consumption pattern serves as a basis to determine corrective actions against demand changes (Simatupang *et al.*, 2004). For example, emergency replenishment is placed to avoid stockouts if the buffer size drops to within the critical zone.

5. A COLLABORATIVE REPLENISHMENT PROCESS

A collaborative replenishment process can be described within the classical cyclic process of tactical planning, execution, and control as shown in Figure 3. First, the tactical planning process is carried out to specify a set of collaborative objectives, a demand plan, inventory plan, leadtime plan, event plan, targets, and a set of exception plans. The plan is then implemented by the execution process that initiates and terminates a sequence of planned activities or events in three ways: execution of activities as planned, executing of improvement initiatives, and implementation of corrective actions.



Execution of the plan is controlled by monitoring a wide range of measures and indicators. Actual measures are then compared to the targets, and if there is significant deviation, a control response is evoked. A control response can be a selection from the exception plan to minimize the impact of the deviation. Other control responses involve the selection of suitable improvement initiatives such as changing targets, refining the monitoring system, reducing the length and variability of leadtime, and reducing demand variability. The improvement initiative often modifies the plan. The revised plan is then continued until the monitoring process identifies a deviation that justifies a new control response. Monitoring the status of process and performance is not only used to evoke control responses, but also to determine when to execute the next activities specified in the plan. Predefined initiation and termination dates are used to initiate a new planning process. The following paragraphs describe each component of the collaborative replenishment process.

5.1. The tactical planning process

The tactical planning process is a technique for developing a set of objectives to be attained in the future, and an outline of the activities or means used to influence the achievement of those objectives. Plans have a specific time horizon within which the objectives are to be satisfied. The plan for collaborative replenishment comprises seven main components: an overall objective and a set of intermediate objectives, demand plan, inventory plan, leadtime plan, event plan, targets, and a set of exception plans to be applied when actual conditions deviate from those predicted in the plan.

An overall objective of collaborative replenishment is to minimize store stockouts, reduce inventory levels, smooth production requirements, and minimize operating costs. A set of intermediate objectives establishes standards that are used to control the execution of the plan. A demand plan outlines the product plan, sales forecast, and order forecast. An inventory plan specifies optimal deployment of strategic buffers (stocks) to maximize availability at minimum cost. This includes inventory policies that encourage the practice of supply according to demand. A leadtime plan outlines a set of initiatives to reduce leadtimes and leadtime variability. An event plan comprises the order and timing of activities or operations to be undertaken to achieve objectives. Heuristics are often used to determine the sequence and timing of events. Sequencing rule, for instance, is used to determine the chronological order of events during the planning horizon period. Some activities must be undertaken in order for the next activity to proceed. A range of indicators are used to initiate an activity including benchmark dates and performance status.

The targets consist of a set of target thresholds (minimum and maximum values) - such as inventory targets, replenishment time, forecast error, availability, costs, and benchmark dates - that are used to control the execution of the plan through time. Exception plans provide several choices of action in any particular period if conditions vary from those predicted in the plan. These plans are designed to cope with supply and demand variations. The chain members need to ensure that large exception plans are available around critical events to cope with uncertainty, because these events have a major impact on the supply chain profitability. The exception plans can be classified in terms of their influence on supply or demand (such as expediting and price discounts) and their influence on the plan (such as changing the timing or sequence of events).

5.2. The execution process

The execution process aims to activate and terminate the sequence of events such as ordering, shipping, receiving, distributing, and storing - as specified in the plan. This process results in three possible actions. First, the chain member executes events according to the (revised) plan. Second, the chain member might modify the plan and execute the improvement action. Third, the chain member implements the suitable exception plan as conditions deviate.

5.3. The control process

The control process is summarized in Figure 3. First, important events and performance indicators are monitored. Some measures are taken daily, with others less frequent and influenced by conditions rather than time interval. Data is processed and analyzed to some degree and then compared to targets or standards specified in the plan. When actual performance equals or exceeds a standard this identifies a decision point, at which point the chain member must decide between determining a suitable improvement opportunity or a suitable exception plan. The actual conditions at the time are compared with the entry conditions associated with the next activity in the plan. If these match, then the chain member will determine if a suitable improvement opportunity exists. If it does not, the chain member executes the next event or activity in the plan and the execution of the plan continues to be monitored. If an improvement initiative is identified, the plan is modified and the initiative is implemented.

However, if conditions deviate from those predicted due to external factors, an appropriate exception plan may be selected in order to minimize the negative impact of the deviation. The actual conditions are then compared to the entry conditions of the exception plans. The exception plan that complies with all the entry conditions is then implemented. Given the repetitive nature of tactical decisions, the cause of deviation is likely to be known. Diagnosis is used when a novel problem is encountered. The chain members appear to respond to problems in three ways: by modifying the existing plan, initiating a corrective action, and changing targets.

As an illustration of the control process, buffer management is known as an effective inventory buffer control that provides a continuous monitoring approach to demand fluctuation (Goldratt and Cox, 1992). Buffer management is used to protect sales, reduce inventory, and decrease operating expense. The first objective is met by quantifying the buffer according to the level of consumption it services and the time taken to replenish it. This quantity ensures no stockouts when unexpected demand occurs, and should be sufficient to restore to the buffer level in time without loss of sales. Buffer management minimizes the total inventory pipeline of the supply chain, as a shorter replenishment cycle protects an adequate amount of strategic buffer. Buffer management also reduces emergency shipments by increasing the size of the buffer after monitoring an excessive number of penetrations into the expediting zone.

As the buffer size reflects consumption patterns according to demand fluctuations, it can be closely monitored as a basis upon which to determine appropriate actions. Often the buffer size is divided into three zones: green, yellow, and red. The green zone represents the replenishment level of the buffer (i.e., the order-up-to policy). The red zone indicates when levels have fallen such that a high probability of losing sales exists and therefore requires emergency delivery. This zone allows timely recovery for an emergency replenishment without loss of sales. The supplier does not need to worry if the green zone has some of its buffer consumed. If the buffer of the yellow zone between the green and red has begun to be consumed, then the supplier needs to watch the ongoing consumption closely and make plans to order a sufficient quantity to ensure replenishment to the top of the green zone (usually such ordering occurs at the regular time interval).

Should the red zone be penetrated, the supplier should take action to expedite replenishment to the top of the green zone.

6. CONCLUSIONS

Collaborative inventory management encourages the chain members to deploy inventory with speed to provide the right product in the right place at the right time. Each link in the supply chain does not need to hold excessive inventory, as supply-on-demand (the pull system) and reliable replenishment from the previous link ensure product availability without losing sales. This paper has employed the evaporating cloud diagram to explain the logic underlying CIM. It is only by challenging the assumptions underlying current inventory management practice that the new inventory model can make significant progress. A collaborative replenishment process is also proposed as a means to apply CIM. If consists of three cyclic processes: tactical planning, execution, and control.

However, many operational issues of CIM warrant further research. A case study research is required to compare the traditional practice of inventory management with the concept of collaborative inventory management developed in this paper. Inventory control in buffer management appears to combine periodic review and order point. The policy states that if the inventory level drops below a red zone prior the review date, an emergency replenishment is placed. If not, the replenishment quantity is determined at the end of the review period. Future research is required to quantify theoretical levels of buffer size taking into consideration the relationships amongst strategic buffers along the supply chain. Other future research relates to modeling equitable incentive arrangements for CIM and developing an Internet based decision support system for CIM.

7. REFERENCES

Eppen, G.D. & R. Martin. 1988. Determining safety stock in the presence of stochastic lead time and demand. *Management Science*, *34*(11), 1380-1390.

Fisher, M.L. 1997. What is the right supply chain for your product? *Harvard Business Review*, *75*(2), 105-116.

Fuller, J., J. O'Conor, & R. Rawlinson. 1993. Tailored logistics: the next advantage. *Harvard Business Review*, *71*(3), 87-98.

Goldratt, E.M. 1990. *The Theory of Constraints*. North River Press, Croton-on-Hudson, NY, 36-54.

Goldratt, E.M. & J. Cox. 1992. *The Goal: A Process of Ongoing Improvement*. North River Press, Croton-on-Hudson, NY.

Goldratt, E.M., E. Schragenheim & C.A. Ptak. 2000. *Necessary But Not Sufficient*. North River Press, Great Barrington, MA.

Schwarz, L.B. & Z.K. Weng. 1999. The design of a JIT supply chain: the effect of leadtime uncertainty on safety stock. *Journal of Business Logistics*, *10*(1), 141-163.

Simatupang, T.M. & R. Sridharan. 2002. The collaborative supply chain. *International Journal of Logistics Management*, *13*(1), 15-30.

Simatupang, T.M., A.C. Wright & R. Sridharan. 2004. Applying the theory of constraints to supply chain collaboration. To appear in *Supply Chain Management: An International Journal*, *9*(1).

Treacy, M. & F. Wiersema. 1993. Customer intimacy and other value disciplines. *Harvard Business Review*, *71*(1), 84-93.