LSP-CONSTRAINED SUPPLY CHAINS: A DISCRETE EVENT SIMULATION MODEL

Doraid Dalalah

Abstract. In this paper, we present a logistics service provider (LSP) constrained supply chain problem, particularly; we propose a novel supply chain model that consists of three layers of non-cooperative manufacturers, distribution centers and retailers. Products flow from the manufacturers across different warehouses to retailers via LSP. Inventories at warehouses follow smooth and continuous replenishment policy, i.e., perpetual review. The supply chain is represented as an optimization model that maximizes the revenue of manufacturers meets the retailers’ demand and at the same time identifies the necessary warehouses, particularly for supply chains that are affected by leadtime (LT) variation such as fast response industry and short life products. The model solution is adaptive; it determines the best manufacturing rates and identifies the logistic bottlenecks in dynamic supply chain networks. Numerical solutions along with simulation experiments of different supply chain topologies are presented. The simulation results demonstrate the model capability to maximize the revenues by tuning the manufacturing rates and monitoring the work-in-process, products in transit as well as products in inventories.

Mathematics Subject Classification. 68Uxx, 68U20, 90-xx, 76B75.

Received October 3, 2014. Accepted March 10, 2015.

1. INTRODUCTION

Supply chains (SCs) can be understood as the collection of different resources and infrastructure required for the movement of material or products from their sources to end customers and sometimes all the way to disposal. Such movement includes different activities like procurement, manufacturing, transportation, warehousing, customer service, forecast demands and so on [28].

Supply chain performance is strongly affected by its management, in particular with reference to strategy layout [12,26], customers-supplier relationship [31,44], inventory planning [8,13] business process connection [38], order policy [21], topology and internal integration [10,20,29].

Due to outsourcing of manufacturing around the world, current supply chains are increasingly getting more complex, depending on critical infrastructures of roads, railways, airports, etc. In addition to infrastructures, reliable communication systems are crucial for supply chain performance, since conveying information is important between the trading partners. In fact, more and more partners are getting involved in the operation of supply chains a fact which emphasizes the need for reliable communication [36]. Interesting surveys on supply
With the increasing local and global competitiveness, it has been realized that the partners in supply chains should work together in an attempt to minimize the cost of transportation, inventories, etc. Hence, effective management of material and product flows across a supply chain network is crucial to success [11]. The ever changing market fluctuations and the existence of new/old competitors, business providers have to be more adaptive to changes in order to stay competitive and continue to subsist in such an environment. Manufacturing and production establishments encounter such challenges in addition to the need to manage their supply chains and their logistic providers [22].

Coordination among partners in real-time supply chain operation is vital for establishing a global service integration where constraints such as duration, time precedence, capacity and location need to be satisfied [41]. As a matter of fact, supply chains repeatedly illustrate the existence of operational optimization in which a better performance can be achieved via careful management [40].

Typically, managers of supply chains try to maximize the profitable operation of their manufacturing and distribution networks. This may include maximizing gross profit of products or materials distributed through the supply chain minimizing total operating expenses (inventory manufacturing and transportation) and balancing the cost of inventory at all points in the chain with respect to the availability to the customer Delivering products to customers at the lowest total cost and highest profit is one principal goal in supply chain optimization. Such optimization includes tradeoff the costs of transportation, inventory, manufacturing and distributing.

Tradeoffs between reducing lead-times and inventories, uncertainty and sustainability quandaries have been reported in literature [1,7,45]. It was noticed that the supply lead times can be reduced but at the same time inventories are not necessarily reduced. Similarly, the same proposition applies for the Work-in-Process. An expression for the target Work-in-Process when the demand is time-dependent has been proposed in [42].

Evidently, lead time variability is a serious problem in retail industry [16]. Shipping time instability makes retailers face uncertain lead times. In order to induce the retailer to the participation in the coordination plan lead time fluctuations have to be controlled for consistency and reliability of delivery [4,17,30].

In this paper, will present a new supply chain optimization model that introduces both a continuous (perpetual) replenishment policy (CRP) along with the Theory of Constraints to solve LSP capacitated SC. The lead times will be the main trace for the search for the optimal delivery rates. More importantly, our model is suitable for products or materials that can be severely affected by long delivery times such as short shelf-life products as will be illustrated shortly.

The paper is organized as follows: Section 2 presents the motivation and related work to our paradigm. The optimization model is illustrated in Section 3 followed by the numerical solution in Section 4. The algorithmic steps of the solution are presented in Section 5 followed by the simulation experiments in Section 6. Later, the dynamic and adaptive behavior of the model is illustrated in Section 7. Finally the conclusions are given in Section 8.

### 2. Motivation and related work

Inventory replenishment methods usually resolve the following two queries: (a) when a replenishment order should be placed? and (b) how large the replenishment order is supposed to be? The traditional inventory theory classifies replenishment polices as either periodic review systems ((R, S) or (R, s, S) policy) or continuous review systems ((s, S) or (s, Q) policy) [34]. However, studies in such field report that order-based replenishment methods fluctuate owing to downstream supply chain partners’ variation in demand. The fluctuations are amplified across the supply chain network due to further replenishment actions and demands in the supply chain hubs. Named as bullwhip effect, this fluctuation phenomenon is usually found in order based replenishment policies. In most supply chains, bullwhip effect causes excessive inventory, ruins the revenue, and yields imprecise production plans [27]. The class of algorithms that use Continuous Replenishment Policy are
called CRP-type. Such algorithms usually aim at minimizing the bullwhip effect. It is reported that CRP is “an efficient replenishment concept within the Efficient Consumer Response (ECR) arena [47].”

A capacitated flow network is widespread in many application fields. For example, computer networks, the power transmission networks and supply chain networks are all capacitated flow networks. Although capacitated supply chains have been already address in the literature [33], in practice, such networks may suffer a complete or partial failure which led past research interests to focus on the reliability matters [48]. Even for those who did search on optimal replenishment policies they tended to straightforwardly answer the typical questions “when and how much to make/buy” [2].

Not much research exists on perpetual inventory replenishment in supply chain field [47]. A perpetual inventory replenishment is a policy in which buyers and suppliers share inventory information so that they can increase replenishment frequencies and reduce inventories. Perpetual Replenishment Policy is an end-to-end, integrated solution that ensures each store is always stocked with the right merchandise at the lowest possible cost for both retailers and suppliers.

The Theory of Constraints—Supply Chain Replenishment System (TOC-SCRS) proposed by [14] is one of the solutions for the improvement of the bullwhip effect in a multi-echelon supply chain [35, 37, 46]. The TOC-SCRS has been implemented by a growing number of companies. The performance reported by the companies includes reduction of inventory level, lead-time and transportation costs increase in forecast accuracy and customer service levels [15, 18, 43].

At any node in the TOC-SCRS supply chains, two parameters are to be decided, first, the time to reliably replenish (TRR) and the replenishment frequency (RF) [3]. However the TOC method is not flawless, in fact, it was noticed that conflicts between the determination of TRR and the RF may arise. However, the TOC paradigm essentially states that every firm must have at least one constraint [39].

In light of the above discussion and in order to increase the revenues and keep the lead-time constraint inviolate, the presented study differs in various characteristics. First, it introduces both a continuous (perpetual) replenishment policy (CRP) along with the TOC in an effort to minimize the bullwhip effect via a novel replenishment flow-based policy without the aggravation of ordering at distinct times as in traditional periodic review policies. Second, unlike other TOC studies in the extent literature, our study differs in the context that each product has at least one constraint. In contrast to TOC-SCRS and earlier literature exclusively devoted to the periodic-review policies, here, we use a new concept that combines the theory of relaxed constraints and CRP in one suit in order to overcome the bullwhip effect, to deliver according to the LSP capacities and to tune the lead times accordingly.

Third, indeed, our study falls in the category of capacitated supply chain and capacitated networks (i.e., LSP capacitated SC). In fact, it is already realized that the manufacturers can elevate their production rates to an extent beyond the demand rates of the buyer; hence, what transforms the SC to capacitated type is the logistics provider carriage capacity along with the demand. In contrast to existing research, this paper considers the solution of steady state replenishment and inventory levels under the logistic provider’ capacity constraints. Fourth, as a main contribution in this study, lead times will be used as a chief clue in the search for the optimal shipping rates. Buyers and suppliers are supposed to communicate the lead-time of their products. Accordingly, our operational algorithm works in real-time, with adaptability to network changes and capacity as well as demand fluctuations.

Finally, and perhaps most importantly, the presented model is suitable for products or materials that can be severely affected by long delivery times such as dairy products, food, medicine, short shelf-life laboratory products, chemicals, and quick response production [6, 9, 19, 23, 25]. Currently, short shelf-life goods present a challenge for supply chain management owing to shelf-life of the products and strict traceability and delivery requirements. The number of product variants in the short shelf-life category has increased over a relatively short period of time due to an increase in the offering of products such as ready meals and pre-packed meat products. This may also involve medicine, blood units and live tissue transfer and many others. The presented analysis in this paper applies strict constraints on the lead times such that the demand is met under the capacity and...
lead time constraints. Accordingly, the products are delivered with the specified maximum tolerated lead-time under the assumption of reasonable communication system between the suppliers and retailers.

3. The model

In this study, we focus on the logistics part of supply chains. In particular, a control algorithm for the flow of products/material in a time varying supply chain aimed at maximizing the revenues of the manufacturers is proposed. A supply chain that includes manufacturers distribution centers and retailers who are supported by a logistic service provider is considered. Products are held in warehouses at the distribution centers, where continuous replenishment of inventories is achieved via transporting products from the manufacturers to the retailers across a network of distribution centers. The capacities of the LSP will be the driving constraint in our approach. The goal of our study is to meet as much as possible of the retailers demand by simultaneously maximizing the revenues and controlling the backlog of products that are in transit, in inventories and as Work-in-Process (WIP) at the manufacturers. Although we assume that the manufacturers are capable of satisfying any demand size, here, the capacity of the LSP will prevent such production volumes. The product manufacturing rates are restricted since the inventories should not exceed a maximum amount of each product ($I_{inv}$). In our approach, a pseudo-optimal operation plan is achieved by employing the numerical solution in the real operation of the supply chain.

Consider a supply chain consisting of a set of manufacturers $M = \{1, \ldots, M\}$ and a set of products $P = \{I, \ldots, P\}$. Each manufacturer $m, \forall m \in M$ makes a subset of products $p_m$ where $p_m \in P$. Products are transported to retailers through different hubs of distribution centers’ (warehouses) denoted by the set $W = \{1, \ldots, W\}$. The set of retailers is given by $R = \{1, \ldots, R\}$ in which retailer $r$ demands a subset of products out of $P$ where $P_r \in P$ denotes such subset. A logistics service provider takes the role of transporting the different products from the manufacturers to the retailers. The network can admit new manufacturers and/or retailers at any time. Figure 1 shows a simple idea of such configuration. Note that if a retailer orders the same product from two different manufacturers, then each manufacturer will have its own path across the logistic network all the way to the retailer, where each path is treated separately. The sum of the rates of the two paths represents the total quantity arriving to the retailer. For simplicity, let us define the set of products to be manufactured
are given by the vector $\mathbf{I}$, where $\mathbf{I} = \{p_m \cap p_r, \forall m = 1, \ldots, M, \forall r = 1, \ldots, R\}$. The SC warehouses follow a smooth production inventory policy (i.e., the replenishment rate is the same as the delivery rate). The manufacturing rate of product $k$ in $\mathbf{I}$ is denoted by $\mu_k$ while the inventory of the same product at warehouse $w$ is denoted by $I_w^{inv}$ and the total inventory $I^{inv} = \sum_w I_w^{inv}$.

Via an automated information system, the manufacturers can record the arrival time of their products at the retailers’ side and hence they can compute the lead time $Lt_k$ since the issuance of the order.

The retailers convey their demands per period back to the related manufacturer. The demand for product $k$ per time period at retailer $r$ is given by $d_k$ (note only one retailer demands $k$). The products are assumed to be quantified by the same measurement characteristic such as weight, volume packages, containers, etc. Warehouses are connected by a set of links $L = \{l, \ldots, L\}$. The transportation capacity of link $l$ is given by $c_l$ which is assumed to be measured by the same product characteristic per period as well. No matter what product mix is to be transported through the link $l$, the LSP can only transport an amount of $c_l$ of such mix.

The goal of manufacturers is to maximize revenues and to meet as much as possible of the demands set by the retailers within a reasonable amount of time. Hence, the time ($L_t$) necessary to make, to transport and to store a product $k$ of manufacturer $m$ to retailer $r$ is preferred not to exceed a deadline set by the retailer denoted by $Lt_k^{max}$. Here, $Lt_k^{max}$ is maximum tolerable time required to manufacture, to transport and to hold the product in inventories. Accordingly, $Lt_k$ is the sum of the manufacturing time $T_k^{man}$ plus the transportation time $T_k^{trans}$ and the entire inventories’ holding time $T_k^{hold}$, i.e., $Lt_k = T_k^{man} + T_k^{trans} + T_k^{hold}$. Note that for a certain product $k$, the quantities $T_k^{man}$ and $T_k^{trans}$ depend on the capacities which may vary over time while $T_k^{hold}$ depends on the level of inventories held at each warehouse. The quantity $T_k^{hold}$ is the aggregation of the holding times at all warehouses across the path of the product $k$, that is: $T_k^{hold} = \sum_{w \in W(k)} Q_w$ where $Q_w$ is the expected delay time since storage to shipping at warehouse $w$ and $W(k)$ is the set of warehouses traversed by product $k$. For example, if the LSP has as much as necessary capacity to transport product $k$ at a rate of $\mu_k$ (which in essence will be the manufacturing rate in our model) then the amount of product $k$ including the quantities of Work-In-Process ($I_k^{WIP}$), the quantities in transit ($I_k^{trans}$) and the quantities in inventories ($I_k^{inv}$) is estimated by $I_k = I_k^{WIP} + I_k^{trans} + I_k^{inv} = (T_k^{man} + T_k^{trans} + T_k^{hold})\mu_k = Lt_k\mu_k$. Now expressing this in a vector form for the entire $K$ products in $\mathbf{I}$, we have $\mathbf{I} = \mathbf{I}^{WIP} + \mathbf{I}^{trans} + \mathbf{I}^{inv}$, where $\mathbf{I} = [I_1 I_2 \ldots I_K]^T$, $\mathbf{I}^{WIP} = [I_1^{WIP} I_2^{WIP} \ldots I_K^{WIP}]^T$ and so on for the remaining components of $\mathbf{I}$. The products manufacturing/delivery rates are given by the vector $\mu = [\mu_1 \mu_2 \ldots \mu_K]^T$ where $\mathbf{u}$ is a diagonal matrix such that $\text{diag}(\mathbf{u}) = \mu$. Based on the above arguments, the maximum allowable lead-times $Lt^{max}$ will result in a maximum total inventory ($\hat{I}$) that are related to each other by: $\mathbf{u}Lt^{max} = \hat{I}$. Alternatively, this can be written in a matrix form as:

$$Lt^{max} = \mathbf{u}^{-1}\hat{I}$$

where $\hat{I} = \mathbf{I}^{WIP} + \mathbf{I}^{trans} + \mathbf{I}^{inv}$. For instance consider the sub-chain depicted in Figure 2. Let the weight be the common product measure characteristic. Assume further that the steady-state flow of a given product across its path to be 2tons/day and warehouse “1” maintains an average inventory of 3 tons while warehouse “2” maintains an average of 2 tons of the same product. Suppose that at the steady state operation the manufacturer will make 2 tons of the same product per day, then the total expected lead time a ton will spend within the chain
Figure 3. The revenue as a function of the manufacturing rate.

from issuance of the order till the arrival to the related retailer is equal to the expected time in inventories 
\((1.5 + 1 = 2.5\) days\) plus the transportation time \((T_{\text{trans}} = 0.5 + 0.5 = 1\) days\) in addition to the manufacturing 
time \((0.5\) day\), hence the total \(L_t\) time is 4 days for this configuration.

In our model, the product routes to destinations are given by a \((K \times W)\) routing matrix \(A\) which consists 
of the elements denoted by \(r_{k,w}\) where \(r_{k,w} = 1\) if product \(k\) is stored in (passes through) warehouse \(w\) and “0” 
otherwise. Such matrix is predetermined by the LSP.

The objective of our model is to maximize the manufacturers’ revenues while trying to meet as much as 
possible of the retailers’ demands and not to exceed a tolerable inventory limits at the warehouses. Revenues 
on sales (with discounts on larger quantities) can be modeled by a piecewise curve as in Figure 3. Accordingly, 
revenues may be approximated by a differentiable and strictly increasing function. Such revenue pattern can be 
modeled by several functions. One simple and mathematically tractable form is the logarithmic category, which 
symbolizes the logarithmic return or continuously compounded return, also known as force of interest. In this 
proposal, the numerical solution will guide to the best category of such revenue functions as demonstrated in 
Section 4 of this proposal.

Under such assumptions the steady state operation of the supply chain is attained by solving the following 
model:

\[
\max \alpha R \\
\text{s.t.} \\
AQ + T_{\text{man}} + T_{\text{trans}} \leq Lt_{\text{max}} \\
\mu \leq d \\
\mu \geq 0
\]

\[
(3.2)
\]

where \(\alpha = [\alpha_1 \alpha_2 \ldots \alpha_k]\) denotes the weight of the revenue functions of each product, \(R = [R(\mu_1)R(\mu_2)\ldots R(\mu_K)]^T\), where \(R(\mu_K)\) is the revenue function of product \(k\), \(d = [d_1 d_2 \ldots d_K]^T\) is the 
retailers’ demand for all products, the delay of products at each warehouse is given by \(Q = [Q_1 Q_2 \ldots Q_W]^T\). 
Note that the inequalities in the last two constraints in \((3.2)\) refer to element by element comparison of the 
given vectors. Now, for the first constraint, recall that \(Lt_{\text{max}} = u^{-1}1\). which by multiplying by the matrix \(u\) 
results in \(uAQ + uT_{\text{man}} + uT_{\text{trans}}1\). By this notation, the second and third terms in the left hand side of this 
constraint simply represents \(I_{\text{man}}\) and \(I_{\text{trans}}\), where, \(uAQ \leq 1^{\text{inv}}\). Accordingly, our problem reduces to the 
following model:

\[
\max \alpha R \\
\text{s.t.} \\
uAQ \leq 1^{\text{inv}} \\
\mu \leq d \\
\mu \geq 0.
\]

\[
(3.3)
\]
The above problem maximizes the total revenue of the manufacturers subject to the constraints of inventory backlogs and demands. In other words, the LSP capacities and retailer demands are the only constraints in the model while the manufacturers are assumed to have the unrestricted production capability. The decision variables are the components of the manufacturing rates vector \( \mu \). The first constraint which can be written as 
\[
\mathbf{I}^{\text{inv}} \leq \mathbf{I}^{\text{inv}}
\]
is also related as 
\[
L \hat{t}_{k} \leq \mu_{k} \frac{\mathbf{I}^{\text{inv}}}{\mu_{k}} + T^{\text{trans}}_{k} + T^{\text{man}}_{k} = T^{\max}_{k}, \quad \forall k \in \Omega\]
or simply the lead time should not be higher than the specified deadline set by the retailer including the transportation and manufacturing times. However, measuring the lead time in a dynamic environment and continuous replenishment of inventories is not straightforward. Hence, a numerical solution along with real-time simulation is required to address the above model.

Under a steady state operation of the supply chain, the manufacturers are supposed to make and ship as much as possible of their products to meet the demands within the allowable capacities of the logistic service providers. Intuitively, such action will increase the revenues of the manufacturers. Note however that each warehouse will hold an inventory of each product to satisfactorily meet the demand of the upstream warehouses/retailers. A steady state solution of the above model means that the manufacturers are delivering products without violating the LSP transportation capacities.

Now, since we are dealing with non-cooperative manufacturers, each manufacturer seeks to maximize its own revenue while meeting the demand and inventory constraints. The aggregate objective function is separable in terms of the manufacturing/delivery rates i.e., 
\[
\text{revenue} = \alpha_{1} R(\mu_{1}) + \alpha_{2} R(\mu_{2}) + \ldots + \alpha_{K} R(\mu_{K}),
\]
hence, maximizing each manufacturer’s revenue participates in the maximization of the total manufacturers’ revenues. Due to non-cooperative environment in manufacturing, the manufacturers do not communicate their rates with each other; instead, each manufacturer will tune its rate according to the delay in inventory. Note that the higher the value of \( \alpha_{k} \) the more greedy behavior of \( k \), by this we mean the manufacturer \( k \) tends to push more products into the chain, however, this will affect the delay on its products yielding more penalties as will be illustrated in Section 4. The value of \( \alpha_{k} \) has an effect on the convergence of the solution but not the steady state rates. The steady state rates are the same no matter how greedy the manufacturers.

4. Solution of the model

The Lagrangian of the problem in (3.3) above under the first constraint set is given by:

\[
f(\mu, \theta) = \alpha R - \theta \left[uAQ - I^{\text{inv}}\right]
\]

where \( \theta = [\theta_{1}, \theta_{2}, \ldots, \theta_{K}] \) is a vector of Lagrange multipliers. The above expression can also be expressed as:

\[
f(\mu, \theta) = \sum_{k \in \Omega} \alpha_{k} R(\mu_{k}) - \sum_{k \in \Omega} \theta_{k} \left(\mu_{k} \sum_{w \in W(k)} Q_{w} - I^{\text{inv}}_{k}\right)
\]

\[
= \sum_{k \in \Omega} \left(\alpha_{k} R(\mu_{k}) - \theta_{k} \left(\mu_{k} \left(Lt_{k} - T^{\text{trans}}_{k} - T^{\text{man}}_{k} - I^{\text{inv}}_{k}\right)\right)\right.
\]

Our problem can be presented as a function of the dual vector \( \theta \) as:

\[
\max_{0 \leq \mu \leq d} f(\mu, \theta)
\]

A dual vector \( \theta \) exists so as to min \[0 \leq \theta \max_{0 \leq \mu \leq d} f(\mu, \theta)\] Note that the solution of (4.3) is either at an interior point or at the limit “\( d \)” Now, the Lagrangian of problem (4.3) above is given by:

\[
F(\mu, \theta, \lambda, \gamma) = \alpha R - \theta \left[uA \left(Lt - T^{\text{trans}} - T^{\text{man}}\right) - I^{\text{inv}}\right] + \lambda |d - \mu| + \gamma \mu
\]
where $\lambda$ and $\gamma$ are row vectors of Lagrange multipliers. By simplifying the matrix multiplication we get the following:

$$F(\mu, \theta, \lambda, \gamma) = \sum_{k \in \Omega} \left( \alpha_k R(\mu_k) - \theta_k \cdot \left( \mu_k( Lt_k - T_{k}^{\text{trans}} - T_{k}^{\text{man}}) - \hat{t}_{k}^{\text{inv}} \right) \right) + \sum_{k \in \Omega} \lambda_k \left( d_k - \mu_k \right) + \sum_{k \in \Omega} \gamma_k \mu_k. \quad (4.5)$$

The Karush–Khan–Tucker optimality conditions of the above expression results in the following:

$$\alpha_k R'(\mu_k) - \theta_k \left( Lt_k - T_{k}^{\text{trans}} - T_{k}^{\text{man}} \right) - \lambda_k + \gamma_k = 0,$$

$$\lambda_k \geq 0,$$

$$\gamma_k \geq 0,$$

$$\lambda_k \left( \mu_k - d_k \right) = 0,$$

$$\gamma_k \mu_k = 0,$$

$$\forall k \in \Omega. \quad (4.6)$$

The above will yield a set of the following possible solutions from which the optimal solution can be a combination of any as a function of the multiplier $\theta$. Particularly: $\mu_k(\theta_k) = 0$,

$$\mu_k(\theta_k) = d_k \quad \text{or} \quad \mu_k(\theta_k) = R_k^{-1} \left( \theta_k \frac{Lt_k - T_{k}^{\text{trans}} - T_{k}^{\text{man}}}{\alpha_k} \right) \quad \forall k \in \Omega.$$

Consider the first candidate in which the manufacturing rate is equal to zero, here, since the components of the objective function are strictly increasing as a function of the rates and are separable with respect to the delivery rates then a zero rate unlikely will participate in the maximization of the objective function in problem (3.3).

Now if the demand for all the products is so high under the hypothetical assumption of unlimited production capacity of the manufacturers, the only constriction will be the LSP transportation capacity and the storage capacities which are implicitly involved in the first constraint. Under this situation with such strictly increasing objective function, the first constraint in (3.3) will be active when the demand is so high, otherwise the second constraint will be active if the demand is low. Note that when the retailers’ demands are so high, the optimal solution of the problem in (4.3) will be pushed into an interior point of its feasible space, where inventories will build up due to the delay experienced while waiting for the LSP transportation service. Inventory at warehouses will vanish in the case of high outgoing transportation capacity. Hence, products will exhibit more delay when the demand is small. Accordingly, the possible solutions are expressed as follows:

$$\mu_k^*(\theta_k) = \begin{cases} 
\mu_k(\theta_k) = d_k & \text{if } Lt_k = T_{k}^{\text{trans}} + T_{k}^{\text{man}} \\
\mu_k(\theta_k) = R_k^{-1} \left( \theta_k \frac{Lt_k - T_{k}^{\text{trans}} - T_{k}^{\text{man}}}{\alpha_k} \right) & \text{if } Lt_k > T_{k}^{\text{trans}} + T_{k}^{\text{man}}
\end{cases} \quad (4.7)$$

where $\delta$ is an estimate of the inventory charges remunerated by the manufacturer making $k$. The above solution states that the optimal manufacturing rate is either equal to the demand in the case that the LSP is capable of transporting products without delay, otherwise, the limits on the inventory in the first constraint of (3.3) will be responsible for determining the optimal production rates.

To put the above solution in more closed form, we already have established that when $Lt_k = T_{k}^{\text{trans}} + T_{k}^{\text{man}}$, then $\mu_k(\theta_k) = d_k$ and at the same time $\mu_k(\theta_k) = R_k^{-1} \left( \theta_k \frac{Lt_k - T_{k}^{\text{trans}} - T_{k}^{\text{man}}}{\alpha_k} \right) = R_k^{-1} (0)$, which results in the boundary condition: $R_k^{-1} (0) = d_k$. 


Figure 4. The function $R'_k(\delta_k)$ of product $k$ with the conditions $R'_k(0) = d_k$ and $R'_k(\infty) = 0$ satisfied.

Now if the LSP has low transportation capacities, products will remain in inventories for longer time periods, hence higher charges which can be calculated as a function of $\delta$. Large amounts of inventories remaining for long time periods entail that the manufacturers should lower their manufacturing rates. Hypothetically, extremely accumulating inventories should force manufacturers to lower/stop production, that is: $R'_k(\infty) = 0$. These two boundary conditions help identify the revenue function stated earlier. Again, several functions can be employed for such boundary conditions, here, one of the mathematically tractable functions is expressed as $\mu_k(\theta_k) = \frac{1}{\alpha_k(\theta_k(L_t - T^{\text{trans}}_k - T^{\text{man}}_k) + \delta_k)}$. Note if the lead time is so high, then the manufacturing rates should approach "0". On the other hand, when $L_t = T^{\text{trans}}_k + T^{\text{man}}_k$ i.e., no delay is observed, the rate will be exactly equal to $d_k$ In fact, a strictly defined $R'_k$ function will strictly determine the revenue function and a strictly defined revenue function will determine the $R'_k$ function. That is:

$$\theta_k = R'_k(\delta_k) = R'_k\left(\frac{L_t - T^{\text{trans}}_k - T^{\text{man}}_k}{\alpha_k}\right) = \frac{1}{\theta_k(L_t - T^{\text{trans}}_k - T^{\text{man}}_k) + \frac{1}{\alpha_k}}$$

$$= \frac{1}{\alpha_k(\theta_k(L_t - T^{\text{trans}}_k - T^{\text{man}}_k) + \frac{1}{\alpha_k}) + \frac{1}{\alpha_k}} = \frac{1}{\alpha_k\delta_k + \frac{1}{\alpha_k}}.$$ (4.8)

Figure 4 presents a typical $R'_k$ function with the given boundary conditions satisfied. Solving for $R$ from $R'_k(\delta_k)$ results in:

$$R(\mu_k) = \frac{1}{\alpha_k}\log\mu_k - \frac{\mu_k}{\alpha_k d_k}.$$ (4.9)

Such revenue function is strictly increasing as a function of $\mu$ within the interval $0 < \mu_k \leq d_k$. Clearly, more greedy behavior of the manufacturers is expected when the demand for a product is high, where the larger the demand, the more greedy a manufacturer is expected to be.

5. The algorithmic solution

The optimal point of the problem in (4.3) is affected by the demand size. If the demand is so high then the optimal rates will be tuned to match the LSP transportation capacities (interior point solution). If the demands are way less than the LSP capacities, then the demands will be satisfied and the solution is said to be on the borders of the problem constraints (active constraints). The gradients of (4.3) are simply calculated by:

$$\frac{\partial L(\mu, \theta)}{\partial \theta_k} = -\mu_k(L_t - T^{\text{trans}}_k - T^{\text{man}}_k) - \hat{i}^{\text{inv}}_k \quad \forall k \in \Omega.$$ (5.1)
By minimizing the problem (4.3) with respect to the vector \( \theta \), the gradient projection algorithm can be used to solve the problem in (3.3). That is:

\[
\theta_k^{\text{new}} = \max \left\{ \theta_k^{\text{old}} + \Delta \left( \mu_k \left( Lt_k - T_k^{\text{trans}} - T_k^{\text{man}} \right) - \hat{I}_k^{\text{inv}} \right) , 0 \right\}
\]

where \( \Delta \) is a small step size. Our model assumes that the lead time of each product can be recorded concurrently once the product arrives to its retailer via an information system which is an available service in current logistics advances. Consequently, the numerical solution of the above model is attained by conducting the following steps:

\[
\begin{align*}
\theta_k^{\text{new}} &= \max \left\{ \theta_k^{\text{old}} + \Delta \mu_k \left( Lt_k - T_k^{\text{trans}} - T_k^{\text{man}} \right) - \hat{I}_k^{\text{inv}} , 0 \right\} \\
\mu_k \left( \theta_k^{\text{new}} \right) &= d_k \quad \text{if } Lt_k = T_k^{\text{trans}} + T_k^{\text{man}} \\
\mu_k \left( \theta_k^{\text{new}} \right) &= \frac{1}{\theta_k \times (Lt_k - T_k^{\text{trans}} - T_k^{\text{man}}) + \frac{1}{\alpha_k}} \quad \text{if } Lt_k > T_k^{\text{trans}} + T_k^{\text{man}}.
\end{align*}
\]

Note that the solution in (5.3) requires that the lead time of each product to be available at the time of its arrival to the intended retailer which can only be available via real operation or simulation of the chain. The solution above has been implemented using discrete event simulation package.

6. Experimental results

In our experimental part of this study, we have built the necessary modules using ARENA discrete event simulation package (R13) to model different supply chain configurations. To demonstrate the idea, first we consider a simple supply chain consisting of a single manufacturer and a retailer with two warehouse hubs as shown in Figure 5a. Using the tonnage as the common characteristic the retailer’s demand is assumed to be 8 tons/day while the capacity of the LSP between the two hubs is 5 tons/day. The maximum inventory that can be held at any of the two warehouses is 10 tons. From the optimality conditions, we have \( \frac{\partial L(\mu, \theta)}{\partial \theta_k} = \mu_k \left( Lt_k - T_k^{\text{trans}} - T_k^{\text{man}} \right) - \hat{I}_k^{\text{inv}} = 0 \) or equivalently for this instance \( \mu_1 \left( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \right) = \hat{I}_1^{\text{inv}} \). One product is considered in this experiment and hence, \( \hat{I}_1^{\text{inv}} = 10 \) tons, which means that the inventory downstream of the related manufacturer should not exceed this amount. Further, maintaining a steady state inventory of this amount entails that the LSP is transporting at its tolerable transportation capacity, i.e., \( \mu_1 = 5 \) tons/day and hence \( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} = 2 \). The dual feasible conditions can be used to find the theoretical value of the Lagrange multiplier that is: \( \frac{\partial L(\mu, \theta)}{\partial \mu_1} = \alpha_1 R_1 - \theta_1 \left( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \right) = 0 \), where in our simulation the revenue curve is given by: \( R(\mu_1) = \log \mu_1 - \frac{\mu_1}{d_1} \) when \( \alpha_1 = 1 \). Substitute the revenue in the expression above we get: \( \left( \frac{1}{\mu_1} - \frac{1}{d_1} \right) - \theta_1 \left( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \right) = 0 \). Plugging the values of \( \mu_1, d_1 \) and \( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \) results in \( \frac{1}{\mu_1} - \frac{1}{d_1} - \theta_1 \left( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \right) = 0 \). Plugging the values of \( \mu_1, d_1 \) and \( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \) results in \( \frac{1}{\mu_1} - \frac{1}{d_1} - \theta_1 \left( Lt_1 - T_1^{\text{trans}} - T_1^{\text{man}} \right) = 0 \).

The general layout in ARENA is presented in Figure 5b. After conducting a simulation run of the same example, the same rate and multiplier values are attained at the steady state operation. Figures 5c and 5d show the rate and the multiplier as a function of the chain operational time. Note that after almost 50 days the chain will reach a steady state without receiving a feedback from the warehouses or retailers except the lead-time.

The algorithm starts with a manufacturing rate that is equal to the demand of each retailer. If the chain cannot take such production, inventories will build up until the manufacturers tune their rates according to the observed lead-times of their products.

In our next simulation experiment we consider a supply chain consisting of two manufacturers, 5 warehouses and 4 retailers. Manufacturer 1 produces two types of products which are shipped to retailer 1 and 3 respectively. Manufacturer 2 makes one product for retailer 2 and another for 4 as shown in Figure 6. Assuming the tonnage as the common characteristic between the different products, the maximum tolerable amount in inventories of product 1 is set to 15 tons, that of 2 is set to 10 tons and of product 3 is set 20 tons and finally of product 4 is set to 5 tons. The demands for the 4 products are 15, 20, 20 and 12, respectively per period. The LSP
transportation capacities are listed in Figure 6 per period. Accordingly, the data of the model are: $M = \{1, 2\}$, $P = \{1, 2, 3, 4\}$, $R = \{1, 2, 3, 4\}$, $\alpha = [1, 1, 1, 1]$, i.e., equal greed weights.

Upon conducting a simulation run for this model, the following rates were found to maximize the manufacturers’ revenues as shown in Table 1.

The given manufacturing rates are the optimal rates that will maximize the aggregate manufacturers’ revenues. The results of the above example indicate that warehouse 1 and warehouse 2 as well as 5 need not to be replenished with any products since the steady state operation entails that products passing such warehouses will be shipped right away without any delay. Concluding that the hubs at warehouse 1, 2 and 5 are not necessary; instead products can be shipped directly without stopping at such hubs. Further, some products, may not need hubs at some distribution centers, for instance, product 2 does not need a hub at warehouse 1, 3 and 5 which lessens the need for loading and unloading of products. Clearly, the total tonnage in inventories within the SC is determined by the right hand side of the first constraint in problem (3.3), which here equals to 45 tons of the different products.

The next experiment consists of three manufacturers, 4 retailers and a set of 8 warehouses. Seven different products are manufactured and shipped from their manufacturers to miscellaneous hubs of warehouses for the
Table 1. The results of the layout in Figure 6.

<table>
<thead>
<tr>
<th>Product Manufacturer Retailer</th>
<th>Data</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Rate</td>
<td>Lead time (Days)</td>
</tr>
<tr>
<td>k</td>
<td>m</td>
<td>r</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Sums</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A simulation model has been built and run for this layout. The steady state results are shown in Table 2. For this example, no greedy manufacturers exists, that is $\alpha_k = 1$. The optimal manufacturing rates will yield the given lead times. Although it may be possible to have less lead times for some products via altering the manufacturing rates, however, doing so, will decrease the total revenue since our solution is Pareto efficient, i.e., increasing one rate results in decreasing another at full operational capacity of the LSP.

A simulation model has been built and run for this layout. The steady state results are shown in Table 2. For this example, no greedy manufacturers exists, that is $\alpha_k = 1$. The optimal manufacturing rates will yield the given lead times. Although it may be possible to have less lead times for some products via altering the manufacturing rates, however, doing so, will decrease the total revenue since our solution is Pareto efficient, i.e., increasing one rate results in decreasing another at full operational capacity of the LSP.

Note that when the demands are high compared to the available transportation capacities, the first constraint in problem (3) will be active and the demands will not be completely met (i.e., slack in the second constraint).
In such scenario, the LSP will ship at its maximum capacity to meet as much as possible of the retailers demand resulting in bottlenecks due to incapability to cope with the demands. Such situation will yield accumulation in some warehouses because the manufacturers are loading the SC at a rate higher than that which can be put up with the LSP.

In this chain, the shipping bottlenecks are the links 2, 6, 8 and 10 resulting in accumulating inventories in the warehouses 1, 2, 5 with the totals of 27, 20 and 22, respectively. Such result mean that other warehouses do not have to hold inventories of products (i.e., loading and unloading of products at such warehouses does not add any significance to the operation of the supply chain). These warehouses do not need to be visited by the products being shipped. In other words, transportation vehicles can just keep moving downstream to the next warehouse station.

With the findings above, the following simplified layout can achieve exactly the same as that in Figure 7 with the same performance, steady state rates and revenues without any further delay or higher lead times.

### 7. Dynamic adaptive supply chain

The model and solution algorithm presented in this paper are characterized by their responsiveness to changes within the supply chain components. In fact, the algorithm can cope with admitting new products, retailers, changes in the LSP capacities and random demands.

To demonstrate we consider a SC segment by modeling a single link which admits three different products at different times with a demand of 80 tons/period, followed by an abrupt change in the transportation capacity
of the link and finally random retailer’s demand. The transportation capacity is initially set to 60 tons/period. The SC starts with only one manufacturer/period, followed by the other two which are 15 periods apart. At time 45, the transportation capacity shifts from 60 to 240 tons/period for a duration of 15 periods followed by a random retailer demand given by the distribution $\text{norm}(100, 8)$ for the three products. The real time production/transportation rates according to this scenario are presented in Figure 9. When only one product is being manufactured, the optimal manufacturing rate is 60 tons/day, when the second product is involved in the production: the optimal rates of both stabilize at 30 tons/period which will be reduced to 20 when the third product is included. Note that when the logistics provider increases its transportation capacity to 240 the three products will be shipped with equal thirds of this capacity, i.e., 80 tons/period. Finally, when a random demand is encountered for the three products, the rates still track the available logistic provider capacity. The rates and their moving averages are illustrated in Figure 10 for the period from 60 to 75 periods. Although the average demand is 100 in this time interval, however, the LSP maximum capacity is 240 tons/period, i.e., 80 for each product. Our model entails equal shares of this capacity when similar demand averages and revenue weights are used.
Figure 9 shows that the algorithmic solution can cope with such dynamic environment and can present new solution for each scenario. Evidently, due to real-time feedback, the search for new optimal rates will take some time; however, simulation packages such as ARENA can present such results with fractions of a second saving the effort of real-time operation.

Note that if one manufacturer tends to exceptionally increase its rate via pushing its products to the network, this will result in inventory buildups, resulting in longer lead times and higher charges. Our solution works in a distributed manner without having the manufacturers communicate with each other. A greedy behavior will affect the revenue of any manufacturer that tends to push more products into the chain.

8. Conclusions

In this study, we have proposed a numerically tractable methodology to address the problem of optimally controlling material flow in supply chains of non-cooperative manufacturers subject to inventory and demand constraints. The model seeks to maximize revenues, control inventories and specify significant warehouses through identifying the best manufacturing rates.

The presented model can adapt to changes concurrently under real operation of the supply chain such as random demand levels, admitting new manufacturers, products, retailers and new topology as well as LSP capacity changes. The proposed model considers layers of manufacturers, distribution centers and retailers who are involved in the activity of making and transporting products demanded by the retailers. According to the demand of the retailers, the capacity of the LSP and the limits on the held inventory, manufacturers will tune their production rates of the different products so as to maximize their revenues while meeting such constraints.

The numerical solution is adopted at the manufacturers’ side which works by receiving feedback from the retailers on the arrival time of the products and accordingly, the manufacturers can compute their production rates. At the same time by following the stated rates of the solution, warehouses maintain constant level of products to meet as much as possible of the retailers’ demand.

To assess the descriptive merits of our model, simulation experiments were conducted for different supply chain topologies using ARENA 2013 Discrete Event Simulation package. The results show that it would be possible to maximize the revenues, keep inventories within stated limits and meet as much as possible of the
retailers demand by following the stated numerical steps. The presented model is adaptive and suites industries in which the delivery times are short and critical such as such as dairy products, food, medicine and short shelf life laboratory products.

REFERENCES


