Analysis of the effects of Lead-Time variation and Lateral Transshipment insertion in a Supply Chain Network

Mohamed Salim Amri Sakhri
Mounira Tlili
Ouajdi Korbaa

Received: date / Accepted: date

Abstract In a supply chain, inventory is the single largest source of costs for a company. This is due to the various physical and informational activities that accompany inventory management, primarily the holding and transportation of inventory. Companies are looking to streamline these activities and minimize the associated costs. One of the most coveted models to jointly solve these two problems is the Inventory Routing Problem (IRP), which will be the focus of this study. This paper addresses the case of a deterministic replenishment demand in a distribution network consisting of a supplier and a number of customers to be served by a single vehicle over a finite planning horizon. We will first study the impact of increasing supplier lead times on network costs. Then, we will study the effects of the Lateral Transshipment (LT) technique on the overall network cost. A mathematical model is developed and solved by an exact method. The results obtained will show that LT is an effective tool capable of improving the total network cost and balancing the customers’ inventory level.

Keywords Inventory Routing Problem · Replenishment Lead-Time · Lateral Transshipment · Supply chain network · Exact methods · Optimization

1 Introduction

Logistics is a constantly evolving concept because of its direct impact on business growth. It is based on logistical practices that allow to control the different costs of the supply chain, namely the costs of ordering, distribution, storage and return of stock. Indeed, logistics allows better cost management by analyzing, planning and anticipating any event that may occur during the supply process. One of the most well-known logistics problems is the Inventory Routing Problem (IRP). This problem has been the subject of much research over the past decade. It combines two important supply chain activities, namely inventory management and product transportation.

The first definition of IRP is given by [1]. They considered IRP as a process of periodic distribution of a single product from a supplier to a set of customers with deterministic demand, using an infinite and homogeneous fleet of vehicles. The vehicles start and end their tours at the supplier location. Next, the IRP was defined by [2] as an extension of the classical Vehicle Routing Problem (VRP) while controlling the inventory level at the different customers and suppliers. This problem consists of making simultaneous vehicle routing and inventory management decisions.

Many variants of IRP have emerged since its inception and the original definition has been reconstructed. The variant of IRP considered in this article looks at the possibility of having stocks not only with the supplier but also from an exchange between customers using the technique of lateral transshipment (LT) as an additional means of replenishment. In other words, LT means that in case of need or to improve its performance, the customer can cover its necessary demand from other customers [3]. LT is very often used when the
supplier’s production unit or warehouse is geographically distant from the customers. In this case, balancing the stock between nearby customers is an advantage for the distribution network.

TL is used to improve profitability and customer service. This process can reduce replenishment time and overall network costs, align decision making processes and improve supply chain performance. This technique can be used primarily in the presence of a supplier to avoid unnecessary holding costs and in the absence of a supplier to avoid stock-outs. It also offers customers with high holding costs an alternative supply solution, allowing them to rely on customers with low holding costs during periods of non-replenishment.

The first investigation into the study of the replenishment network with the LT option was carried out by [4]. Their models aimed to minimise only the inventory costs of the LT-based network in a context of continuous customer demand, without considering routing costs. This same objective has been the subject of several works such as those of [5–10].

Recently, researchers have become increasingly interested in the use of the LT technique in different variants of VRP. In this article, we will only focus on research conducted on the use of LT in VRP problems. Their investigations will be the subject of the next section.

In this article, the problem addressed is firstly the replenishment of a number of customers from a supplier. The supplier has a stock capable of covering the total demand of the different customers in the network. Deliveries are made by a single vehicle of limited capacity and over a periodic planning horizon. Customers cannot be out of stock. Next, we will modify our model by extending the supplier’s delivery time. Finally, we will introduce the LT technique for an inter-customer stock transfer by the supplier’s vehicle during the non-replenishment period.

The aim of this work is to minimise the costs of holding and transporting stocks, both for replenishment and transshipment. In this paper, both the supplier and the customers have the possibility to set delivery times. In the first model developed, the supplier imposes specific deadlines for replenishment. In the second model, customers outside the replenishment periods can use LT to balance their stocks and optimise their holding costs. In the latter model, customers can also impose their stock transfer schedules, which adds a distinctive element to the problem.

The main contributions of this document can be summarised in the following points:

- Proposal of a new supply chain model taking into account customer demands, inventory transport and holding costs, variation of supplier lead times in a periodic planning horizon,
- Possibility of stock exchange only between customers by LT technique during the non-replenishment period,
- During the replenishment period, customers are visited only once by the supplier’s vehicle to deliver a quantity of stock capable of covering a horizon that exceeds a single period,
- During transshipment periods, customers are only visited once by the supplier’s vehicle for the stock pickup or delivery,
- Complex combinations of stock deliveries: a restriction that allows only one type of stock transfer during each type of period, e.g., replenishment is performed during the replenishment period and LT is performed during the transshipment period,
- Optimize the total cost of the network, i.e., the costs of holding and transporting inventory through the various nodes of the supply chain under new, more complex constraints,
- Solve the problem with an exact predefined method.

The rest of the paper is organised as follows. In the next Sect. 2, we present the different works carried out in the IRP research axis with the use of lateral transfer (IRP-LT). In Sect. 3, we formally describe the studied problems and present the mathematical formulation of the developed IRP variants. Then, we discuss in Sect. 4 the exact method used to solve our model. In Sect. 5, we present the experimental results of our computational tests and analyse them. The last Sect. 6 concludes our work and gives some perspectives for future research.

2 Literature Review

Our research focuses on the variant of the inventory routing model known as IRP-LT which is attracting increasing interest. Research on these models is very recent and growing, the best known are [11–17]. This work will be the subject of this section. In the following, we will briefly review the studies that have been carried out on the IRP-LT mainly related to our approach.

Coelho et al., [11] studied a variant of IRP, in which transshipments are allowed if necessary. Transshipment is performed if a subsequent delivery exceeds the capacity of the vehicle or if an inventory constraint is violated. Two methods were developed to solve their problem: the Branch and Cut (B&C) algorithm and the Adaptive Large Neighborhood Search (ALNS) algorithm. To evaluate the performance of their model, they used instances of [18] and analysed the solutions...
obtained. They showed the advantage of using transshipment and its ability to significantly reduce the total cost of the replenishment network.

Rabbani et al, [12] proposed a mixed integer linear programming (MILP) for a multi-item IRP, considering financial and lateral transshipment decisions. The network is composed of a firm and agents. The agents’ demand is collected either from the company or from other agents via LT. Two types of homogeneous fleet vehicles with limited capacity are used. The first one transports products from the company to the agents and no other type of delivery should be made during this time. The second type of vehicle is used between the agents when the TL is economical. A hybrid metaheuristic combining a genetic algorithm and a particle swarm optimisation method is developed to solve their problem. Tests on randomly generated instances were performed. The results showed a great improvement by introducing LT in the model structure.

Jemai et al, [13] studied an IRP in a supplier integration context, in the case of a vendor-managed inventory with a consignment stock (CS) policy. They assumed that the supplier could use transshipment for direct customer replenishment if this option improved its economic performance. They proposed three scenarios to analyze the performance of their model: The first uses transshipment, the second uses a partial delivery solution, and the third uses a full delivery solution with an additional fleet of vehicles. The results of their experiments showed that the approach that reduces distribution costs the most is the one where the supplier can meet the entire demand only with the integration of transshipment.

Haliza et al, [14] studied a location inventory routing problem with transshipment (LIRPT). Four logistics decisions were considered: location, vehicle routing, inventory management, and transshipment. Their study aimed to minimize the total costs of the network. The objectives considered were to include the transshipment process in the LIRP model and to evaluate the effects of this inclusion on the performance of the LIRP. They assumed that one of the customers was going to be considered as the transshipment point in the logistics network. A preliminary study of the LIRP model to find the best performance of the LIRP and to select the possible transshipment customers to extend the LIRP to the LIRPT. To select the transshipment point, they used the p-median method. Computational tests were presented, and the results obtained indicated that the LIRPT improved the solutions of the LIRP.

Peres et al, [16] presented a multi-product IRP with planned transshipment. The transshipment is performed by the supplier’s vehicles, which means that they are subject to the same capacity, time and cost restrictions. They introduced an exact formulation and developed a metaheuristic which is a hybrid Randomized Variable Neighborhood Descent (RVND) for solving the problem. They applied their methods on a real case that deals with large scale retailing in Brazil. The experimental results show that transshipment significantly reduces the cost of the network producing a 9% reduction in logistics costs and 70% reduction in inventory levels compared to the existing operation.

Rabbani et al, [15] studied a Supply Chain Network Design problem (SCND) able to meet lead-times directly and explicitly in a multi-echelon, multi-period network with recourse to LT if necessary. They felt that unmet customer demand was penalized. They defined a maximum lead-time for each customer that should not be exceeded. They defined a maximum lead time for each customer that should not be exceeded and considered the LT option at the customer level as well as direct and indirect shipments from distribution centers to consumers. They proposed an algorithm based on graph theory and an exact solution approach calculated by the GAMS solver to solve the SCND problem. The results indicate that the LT can increase customer satisfaction and reduce the total cost of inventory management.

Al-e-Hashem et al, [17] have developed a multi-product IRP model with transshipment in a green context. They considered an uncertain demand in a bi-objective stochastic programming framework. Two objectives have been addressed. The first one is to minimize the supply chain costs. The second is minimizing the total amount of greenhouse gas (GHG) emissions produced by the vehicles and products disposed of. The transshipment option was introduced to control transportation costs, reduce GHG emissions and absorb uncertainty. In order to solve their problem, an efficient hybrid algorithm combining the L-method (a stochastic optimization approach) and compromise programming (an approach for multi-objective optimization) is proposed. The results show that their research leads to a reduction of costs and environmental concerns and highlights the role of the transshipment option as a lever to improve economic and environmental performance and absorb demand fluctuations.

The originality and the good results of the above mentioned work make it worthwhile to engage in this research area. This paper extends the traditional multi-period IRP to create a new model capable of meeting customer demands with varying delivery times and improving supply chain performance using the LT technique. In this work, LT is used as a preemptive option to balance inventory among customers over the non-
replenishment period and intervene when it can generate savings on the total network cost.

The problem addressed in this paper is very different from previous contributions and very original because it presents a real aspect that we find in our daily life. The error in sales forecasts puts us in front of a bad distribution of stocks between the sites of the same company. This problem can be solved by the balancing proposed by the LT. The above considerations give reasons to deal with this problem and justify this research.

The research cited above studied different compositions of distribution networks using LT. Some of them had similarities, including aspects of our research on which we had developed our mathematical model. Indeed, the model we developed was based on the model of [18] in the composition and organization of the network. We also adapted the hypothesis of [12] for the planning of deliveries. For the formulation of LT, we used the one announced by [11]. In the next section, we will present the model we have developed.

3 Problem Formulation

3.1 Problem description

In this article we study a two-tier supply chain. The first contains a supplier and the second consists of a set of customers. Customer demand is deterministic and independent for a single product type. The planning horizon extends over three replenishment periods. The idea of this work is to study the effect of lead-time variation (L) on the network cost of the supply chain. We will also study the impact of the introduction of the LT technique in this same network.

Two types of delivery will be the subject of this work, the first is supplier to customer and the second is inter-customer. In the first type of delivery, only one vehicle is operational, it starts its round at the supplier and ends it there. The total demand from customers is less than the capacity of the vehicle in each period. The supplier’s production capacity is limited but covers the customers’ demand in each period. No other type of delivery is allowed during the replenishment period.

The second type of delivery is for customers only, when they are outside the replenishment periods. In this case, if the transshipment option is economical and reduces the total cost of the network, it is recommended to use it. The stock transfer is carried out by the supplier’s vehicle. Customers are divided into two groups. Customers who have surplus stock and can transfer the stock are called pick-up nodes, and customers who may be out of stock are transshipment nodes.

In what follows, we will begin by defining the parameters of the IRP model. Then, we represent the basic model of [18] which we will transform. Then, we extend the formulation of the proposed model by increasing the lead-time from one to two periods. Finally, we introduce the formulation of the LT technique which will be done in the non-replenishment period. Indeed, this problem has not been sufficiently explored in previous research. However, there are other important problems in the literature, closely related to our studied IRP, which will help us to formulate our models.

3.2 Presentation of the IRP

Our problem can be defined as a Mixed Integer Programming (MIP) model in the sense that the supplier must control the level of stock at each customer and jointly decide on the choice of delivery routes. This problem can be represented by a graph $G(N, E)$, where $N = N' \cup \{0\}$, $N'$ is the set of vertices of network customers such that $(N' = \{1, ..., n\})$ and and the vertex 0 is the supplier. $E$ is the set of edges that defines the routes connecting the vertices. We will start by giving the different notations that will be used to formulate our model.

3.2.1 Indices

$i$ and $j$: the supplier or customer index,
$t$: Period index $(t = 1, ..., T)$

3.2.2 Parameters

$C_{ij}$: cost of edge $(i, j)$ traversed by the vehicle $v$ at the period $t$

$I^0_i$: the initial inventory level at the customers $i$

$I^0_0$: the initial inventory level at the supplier

$U_i$: maximum inventory level at the customers $i$

$M_i$: minimum inventory level at the customers $i$

$Q$: the capacity of the vehicle

$s$: the quantity of products available at the supplier at each period $t$
3.2.3 Decision variables

\( X_{ij}^t \): equal to 1 if the edge \((i, j) \in E\) is traversed by the supplier’s vehicle at the period \(t\), 0 otherwise

\( Y_i^t \): equal to 1 if the customer \(i \in N^0\) is visited by the supplier’s vehicle at the period \(t\), 0 otherwise

\( I_{it}^t \): the inventory level at node \(i \in N\) at the period \(t\)

\( Z_{it}^t \): the quantity carried by the supplier’s vehicle after visiting the node \(i\), in each period \(t\)

\( p_i^t \): the quantity of product picked up from the node \(i \in N\) at the period \(t\)

\( d_i^t \): the quantity of product transshipped to the customer \(i\) at the period \(t\)

3.3 Assumptions

The first model studied is inspired by the model of [18] for the single-vehicle IRP and the second by [11] for the IRP-LT. In our model, six assumptions are initially made:

- The supplier’s production capacity is limited, but sufficient to meet the deterministic demand of each customer,
- Customers’ inventory capacities are limited and out of stock is not allowed;
- Only one type of delivery is allowed during each period,
- The Customer’s replenishment or transshipment requests must be delivered in full, at each visit;
- Pick-up customers are those with \(g_i = 3\), their maximum stock level can cover three consumption periods, knowing that \(g_i\) represents the number of time units needed to consume the quantity \(U_i\);
- LT is optional and used by customers who have excess inventory or incur higher holding costs for those at risk of out of stock or who favour reducing network costs.

3.4 Formulation

3.4.1 Replenishment model formulation

In this subsection, we present the mathematical formulation of our proposed IRP. We start with the objective function. It aims at minimising the total operational costs of the network, namely the inventory costs at the supplier and the customers as well as the route costs over the time horizon:

\[
\min \sum_{t=1}^{T} h_0 I_{0t}^t + \sum_{t=1}^{T} \sum_{i=1}^{N'} h_i I_{it}^t + \sum_{t=1}^{T} \sum_{i=1}^{N'} \sum_{j=1}^{N} C_{ij} X_{ij}^t
\]  

(1)

These operational costs are the sum of supplier and customer inventory holding costs plus transportation costs over the time horizon. The constraints of this model will be used in the next section to develop our solution method. To better understand the model, we will discuss constraints that have the same functionality in a separate subsection.

3.4.2 Inventory Level Constraints

The constraints of inventory level at the supplier at the period \(t\):

\[
I_{0t}^t = I_{0t}^{t-1} + s - p_0^t \quad \forall t \in T
\]  

(2)

\[
p_0^t = \sum_{i=1}^{N'} d_i^t \quad \forall t \in T
\]  

(3)

Stock-out constraint at the depot: This constraint guarantees that for each time \(t \in T\) the inventory level at the depot is sufficient to ship the total quantity delivered to the customers:

\[
I_{0t}^t \geq \sum_{i=1}^{N'} d_i^t \quad \forall t \in T
\]  

(4)

The constraints of inventory level at the customers at the period \(t\):

\[
I_{it}^t = I_{it}^{t-1} - r_i + d_i^t \quad \forall i \in N', \forall t \in T
\]  

(5)

Constraints enforce the inventory level to stay between lower bound and upper bound:

\[
I_{it}^t \geq M_i \quad \forall i \in N', \forall t \in T
\]  

(6)

\[
U_i \geq I_{it}^t \quad \forall i \in N', \forall t \in T
\]  

(7)
3.4.3 Order-Up-To Level Constraints

The following constraints ensure the OU policy requirements imposing that, if a customer is visited, the quantity delivered is such that the maximum inventory level is reached:

\[ d_i^t \geq U_i \sum_{j=1}^{N'} X_{ij}^t - I_{i}^{t-1} \quad \forall i \in N', \forall t \in T \]  

(8)

Capacity constraint at the customers: This constraint guarantees that for each customer \( i \) in \( N' \) the maximum quantity \( U_i \) is never exceeded:

\[ d_i^t \leq U_i - I_{i}^{t-1} \quad \forall i \in N', \forall t \in T \]  

(9)

3.4.4 Replenishment Constraints

In our model, the lead time \( L \) is equal to two periods. Indeed, no replenishment is carried out during the second period (12). Therefore, customer demand for the second period must be covered during the first period (11). During the last period, the replenishment only covers the demand for that period (13). The following constraints represent the distribution of inventory quantities delivered versus demand over the planning horizon.

\[ p_0^1 \geq \sum_{i=1}^{2} \sum_{i=1}^{N} d_i^t \]  

(11)

\[ p_0^2 = 0 \]  

(12)

\[ p_0^3 = \sum_{i=1}^{N} d_i^t \]  

(13)

3.4.5 Vehicle Routing Constraints

The flow conservation constraint:

\[ \sum_{i=0}^{N} X_{ij}^t = \sum_{i=0}^{N} X_{ji}^t \quad \forall j \in N', i \neq j, \forall t \in T \]  

(14)

The vehicle capacity constraint:

\[ \sum_{i=1}^{N'} d_i^t \leq Q \quad \forall t \in T \]  

(15)

The next constraint ensures that a single vehicle is available:

\[ \sum_{i=0}^{N} X_{i0}^t \leq 1 \quad \forall t \in T \]  

(16)

Sub-tour elimination constraints for each vehicle route at each period:

\[ Z_{ij}^t + QX_{ij}^t \leq Q - d_j^t \quad \forall i,j \in N', i \neq j, \forall t \in T \]  

(17)

\[ d_i^t \leq Z_i^t \leq Q \quad \forall i \in N', \forall t \in T \]  

(18)

3.4.6 Non-negativity and Integrality Constraints

These constraints will ensure the stability of the model:

\[ X_{ij}^t \in \{0,1\} \quad \forall i,j \in N, i \neq j, \forall t \in T \]  

(19)

\[ Y_i^t \in \{0,1\} \quad \forall i \in N', \forall t \in T \]  

(20)

\[ p_i^t = 0 \quad \forall i \in N', \forall t \in T \]  

(21)

\[ d_i^t \geq 0 \quad \forall i \in N', \forall t \in T \]  

(22)

\[ Z_i^t \geq 0 \quad \forall i \in N', \forall t \in T \]  

(23)

3.4.7 Transshipment Constraints

In this subsection, we will define LT and then present the formulation of its constraints which will be introduced as an optimization tool for the previous model. Indeed, Jabbarzadeh et al. [19] defined LT as a proactive strategy to reduce the impact of disruptions in a supply chain. This managerial practice redistributes products from customers with extra inventory to others who are out of stock. Fig. 1 shows an example of a LT path construction process. The network consists of one supplier and ten customers. Customers are served by the supplier’s vehicle. The black arrows represent paths where the vehicle is empty, the green arrows represent pickup paths, and the red arrows represent delivery paths.

Fig. 1 Example of IRP-LT.
the vehicle leaves the supplier empty, and the last visit must be to a transshipment node \(i\) before returning empty to the supplier’s location. Equalities (4) and (21) will be updated by adding the LT process. Constraint (21) remains valid only when \(t \in \{1, 3\}\), otherwise \(p_{i}^{t} (\in N')\) can be non-zero if transshipment is performed. The constraint (4) of the inventory level at the customers becomes the following:

\[
I_{i}^{t} = I_{i}^{t-1} - r_{i} + d_{i}^{t} (\in D) - p_{i}^{t}(\in P) \quad \forall i \in N' \tag{24}
\]

The equation below ensures that the quantities picked must equal the quantities transshipped.

\[
\sum_{i}^{P} p_{i}^{t} = \sum_{i+1}^{D} d_{i}^{t} \quad \tag{25}
\]

Based on the definition provided by [21], each customer is assigned an action, so that if \(p_{i}^{t} > 0\), the customer is assigned to pickup and the quantity picked up is equal to \(|I_{i}^{t} - r_{i}|\), and if \(d_{i}^{t} < 0\), the customer is assigned to delivery and the quantity transshipment is equal to \(|I_{i}^{t} - r_{i}|\). The maximum quantity picked up based on the customer’s inventory levels is determined by the following inequality:

\[
\sum_{i}^{P} p_{i}^{t} \leq \min \left(\sum_{i}^{P} I_{i}^{t} - r_{i}, \sum_{i+1}^{D} |I_{i}^{t} - r_{i}|\right) \quad \tag{26}
\]

The quantity picked up must not exceed the capacity of the vehicle, which is guaranteed by the following constraint:

\[
\sum_{i=1}^{P} p_{i}^{t} \leq QY^{t}_{i} \quad \tag{27}
\]

4 Presentation of the Resolution Method

Our current work is devoted to the study of two common real-life problems consisting of lead time variation and the introduction of lateral transshipment. We will use an exact method to solve small and medium sized instances that is Branch and Cut (B&C) algorithm. We choose to use the CPLEX solver to solve our IRP. In this section, we will briefly present the B&C method included in CPLEX and the concept of upper and lower bounds.

4.1 Introduction of the B&C algorithm of CPLEX

Among the most powerful and widely used algorithms for solving the MIP models used by CPLEX, we mention the Branch and Cut algorithm (B&C). The B&C algorithm is an exact method used to obtain solutions with better accuracy. This method is used by many researchers to optimally solve the IRP problem or its variants. We will give a general description of its branching and cutting process implemented in CPLEX:

- The problem breaks down into a series of sub-problems.
- CPLEX builds a tree structure in which each sub-problem is a node, it represents a Linear Program (LP) and the root of the tree is the continuous relaxation of the original MIP problem (Fig. 2).
- If the solution to the relaxation has one or more fractional variables, then CPLEX tries to find cuts.
- Cutting planes are constraints that intersect areas of the feasible relaxation region that contain fractional solutions (Fig. 3).
- If, after adding the cutting planes, the relaxed solution contains integer variables with fractional values, then CPLEX branches to a fractional variable to generate two new sub-problems.
- CPLEX repeats this process until it obtains a complete solution or an unworkable solution.

By increasing the size of our problem, its complexity also increases and an optimal solution becomes difficult to obtain with an exact method and especially if we impose a limit on the computation time. In this case, an upper (lower) bound must be computed to limit the obtained solutions to a well defined interval.

4.2 The upper bound

The B&C method included in CPLEX is useful to search the upper (lower) bound. In fact, an upper (lower) bound is determined for each sub-problem by solving the linear relaxation of the problem. The upper bound search is a fundamental point of any implicit enumeration algorithm. A good upper bound makes it possible to prune quickly several branches of the search tree. At each iteration, we check if the current solution of LP is an incidence vector of a Hamiltonian Cycle. In this case, if the value of the solution is lower than the global upper bound, this variable is updated, and the solution is saved as the best solution obtained. Otherwise, we try to exploit the solution of the LP to create a high-quality Hamiltonian Cycle. Sometimes, the LP solution provides useful information about the probability of belonging to an edge of the optimal solution.

5 Computational Experiments

5.1 Implementation and Benchmarks

We implement our problem using CPLEX. 12.5.1.0. and run on a PC with a 2.40 GHz Intel® Core™ i7-5500U.
CPU @ processor and 8 GB of RAM. To evaluate the performance of the selected solver, we used instances of the benchmark defined for the single vehicle IRP created by [18]. The characteristics of these instances are described, briefly, as follows:

- Time horizon $T = (3)$;
- Number of customers: 10 sets of small instances, each set consisting of 5 instances of 5, 10, 15, ..., 50 customers;
- Product quantity $r_i$ consumed by the customer $i$ at each period: constant over time and randomly generated as an integer number in the interval [10, 100];
- Product quantity $s$ made available at the supplier at each period: $\sum_{i=1}^{N'} r_i$;
- Maximum inventory level $U_i$ at customer $i$, it is given by $r_i * g_i$ with $g_i$ is selected randomly from {2, 3} and represents the number of time units needed to consume the quantity $U_i$;
- Starting inventory level $I_0^s$ at the supplier: $\sum_{i=1}^{N'} U_i$;
- Starting inventory level $I_0^i$ at the customer $i$: $U_i - r_i$;
- Inventory cost of customer $i \in N'$ is $h_i$: randomly generated in the intervals $[0.01, 0.05]$;
- Inventory cost at the supplier $h_0 = 0.03$;
- A single vehicle is used whose capacity $Q$ is given by: $3/2 \cdot \sum_{i=1}^{N'} r_i$;
- Transportation cost $C_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where the points $(x_i, y_i)$ and $(x_j, y_j)$ are obtained by randomly generating each coordinate as an integer number in the interval $[0, 500]$;
- Replenishment policy: Order-up-to Level (OU) policy, in which every visit to a customer brings its inventory to the maximum level, more details on the OU policy in [22];
- Transshipment policy: The complete pooling (CP) policy. According to this policy, even if the stock is not enough to fill the entire customer demand, a customer agrees to transfer all its available stock in case of need, more details on PC policy can be provided in [20].

5.2 Computational Results

In this study, the maximum stock level at each customer covers for some customers two consumption periods and for others three periods, which is explained by the parameter $g_i \in \{2; 3\}$, (subsection 5.1). This information limited the scope of our model, which can only be tested in the case of $L < 3$. In their work, [18] solved a multi-period IRP in the case of $L = 1$. Here, we will extend $L$ to two periods, applied on the same instances.

Table 1 gives the average results of the multi-period IRP with $L = 1$ and $L = 2$, tested on small and medium instances of [18] using the same resolution method (B&C), with each instance row corresponding to the average results of five instances with the same number of customers. We set a time limit of two hours beyond which the best result obtained is considered the solution of the tested instance. We imposed this time limit to compare our results with those of [18] who used it in their tests.

The first column is relative to all the instances. The second column gives the average results provided in [18] for a single vehicle IRP with $L = 1$. Columns 3 to 6 present the results of our experiments on the IRP model with $L = 2$. Column 3 gives the average results of our developed problem. Column 4 gives the average CPU (in seconds) required to obtain these results. Column 5 shows the average percentage difference between the cost of the existing solution (upper bound) and the cost of the best LP relaxation, when the optimal solution is not obtained within the computational time. Finally, column 6 shows the average rate of change in the total cost of the supply network, from an IRP with $L = 1$ to an IRP with $L = 2$.

In Table 1, we first examine the effects of the variation in $L$ on the total cost (inventory and transport costs), with all other parameters held fixed. We note that this variation affects all the performances of the system. Indeed, customers try to build up their stock to cover their demand in the first two demand periods, which has a direct effect on the increase in stock movements. This is because the growth in demand leads to the need for additional customer stock to cover the supplier’s idle periods. This increases the quantity stored in the customers’ depots, which increases the cost of holding.

The number of customers with priority demand will change. Therefore, the transport map will be changed, which should include more customers in the first and third delivery period. By increasing the supplier’s replenishment lead time, this action has a visible influence on the objective function, which increases, and the results have completely changed due to the increased holding costs. As a result, customers decide to order more products to avoid the possibility of stock-outs. The last column for each set of customers shows the average percentage change in total cost when we go from $L = 1$ to $L = 2$. We notice that, in all sets of instances, when we increase the replenishment lead time, the total cost of the distribution network also increases. This increase averaged 26.38%.

As the number of customers increases, the average computation time, standard deviations and performance of the CPLEX solver deteriorate. Moreover, the quality of the CPLEX solutions does not match that of the best solutions of the IRP variant in any of the time restriction classes. The fifth column under each class of time constraints shows the average percentage difference between the lower and upper bounds. CPLEX has its limits which depend on the growth of the instances and the number of variables. The upper bound will be considered in some cases due to the excessively long CPLEX processor time for problems with more than 20 customers.

However, CPLEX determines the optimal solution values in 16 instances. In 14 instances, the solver obtains results with a deviation from optimality of less than 6%. These results can be considered as solutions to our problem and have upper bounds for each set of instances. These results prove the interest of the B&C method applied to the IRP with $L = 2$, as a viable alternative to obtain excellent quality results for small instances. For the remaining instances (up to 50 customers), the maximum gap between the optimal solu-
Table 1: Average solution values for the IRP with $L=1$ and $L=2$.

| Model | IRP ($L=1$) | IRP ($L=2$) | %Eval. |
|-------|--------------|--------------|--------|
| Instance | $z_1$ | $z_2$ | CPU (s) | %GAP | $z_1$ to $z_2$ |
| Small-5 | 1418.76 | 1791.83 | 0.46 | 0.00 | 26.30 |
| Small-10 | 2228.67 | 2566.23 | 10.17 | 0.00 | 15.15 |
| Small-15 | 2493.47 | 3157.80 | 580.40 | 0.00 | 26.64 |
| Small-20 | 3053.02 | 3944.57 | 1993.59* | 4.89 | 29.20 |
| Small-25 | 3451.15 | 4320.23 | 4158.36* | 8.10 | 25.18 |
| Small-30 | 3643.22 | 4453.52 | 1764.39* | 3.93 | 22.24 |
| Small-35 | 3846.87 | 5022.88 | 2005.02* | 5.97 | 30.57 |
| Small-40 | 4125.70 | 5599.97 | 1439.82* | 6.91 | 31.13 |
| Small-45 | 4270.61 | 6267.08 | 1458.13* | 11.00 | 30.27 |

*: premature stop with an end-of-calculation-time status

Tions and the obtained solutions (upper bound) does not exceed 11%.

IRP involves satisfying customer demands over a discrete time horizon in separate periods. Each customer’s demand is met either by an order from the supplier or by the stock accumulated in previous periods. Another alternative, established during the supplier’s idle period, is to add the use of LT strategy if necessary. This involves considering the possibility of intervention by customers with excess stock where this technique can improve network performance.

Table 2 shows the average results of the IRP problem, when $L=2$ with and without the LT option, on the same instances as in Table 1. Column 1 represents the number of customers in each set of instances. Column 2 gives the average results obtained with our first optimisation scenario (IRP with $L=2$). Columns 3 to 6 present the results of our experiments on the IRP-LT model. Column 3 gives the average results obtained for the IRP-LT. Column 4 shows the average processing time (in seconds) of the IRP-LT. Column 5 shows the average deviation between the cost of the incumbent solution (upper bound) and the cost of the best LP relaxation, when the optimal solution is not obtained within the system memory. Column 6 shows the average deviation of the optimal cost of the IRP-LT from the IRP with $L=2$.

In Table 2, the solutions of the IRP model are compared with those of the IRP-LT, using small instances of [18]. The results in Table 2 show that the LT intervention either generates the same solutions as those obtained for the previous IRP or minimises the different network costs. From this first observation, it can be deduced that LT can only be considered as advantageous. LT can be considered as an alternative for planned delivery in the second period. The application of LT allows the adjustment of the inventory status at the customers, so that the total cost has decreased by 9.92%. The improvement in results can be explained by better management of the quantity of stock allocated to each customer. The above discussion leads us to conclude that this technique is effective in solving IRP problems with constant demands.

However, in some cases LT has no effect and supplier replenishment is considered the optimal delivery method. This can be seen in the solution instances provided, consisting of ten customers. Furthermore, as the number of customers increases, the benefits of using LT increase. The last column shows that the average percentage of total cost decreases by more than 12% in all instances with fifty customers. In conclusion, LT improved the second model solutions presented in nine of the ten sets of instances examined.

The efficiency of CPLEX is also demonstrated by the CPU measurement. Regarding the average processing times needed to obtain the best solution for each run, the CPU varies from 0.41 seconds on average for instances of five customers to 2426.24 seconds for instances of fifty customers, using the same execution platform. The robustness of the solver to solve the two IRP scenarios developed (with and without LT) decreases when the number of customers increases, and in this case the upper bound is considered as the best solution.

This situation is defined by the increasing gap between the best and the average objective, and by the increasing standard deviation of the computation times. The computation times are considered satisfactory, taking into account both the scale of the instance and the structures of the problem explored. The percentage deviation is that between the best given solution (upper bound) and the corresponding lower bound. The results show that CPLEX solved three sets of instances in an optimal way. For the rest of the sets, the solution provided in some instances is given with an acceptable percentage of error. Indeed, the maximum error rate observed on the 50-customer instances was 11.20%, which
can be explained by the lack of memory in the system or the time limit imposed from the beginning.

Another comparison between the solutions provided by the IRP-LT model, and those of [18] can be proposed in this study to evaluate the effect of LT. Table 3 provides the average calculation results of the IRP model proposed in [18] and the developed IRP-LT. Column 1 is assigned to the number of customers of each set of instances. Column 2 shows the results provided by [18]. Column 3 gives the results of the IRP-LT. Column 4 shows the average gap between the optimal solutions of [18] and the IRP-LT.

### Table 3 Average solution values for the IRP of [18] and the developed IRP-LT.

| Model       | Archetti et al. [18] | IRP-LT | %Eval. |
|-------------|----------------------|--------|--------|
| Instance    | z₁                   | z₃     |        |
| Small-5     | 1418.76              | 1729.84| 21.93  |
| Small-10    | 2228.67              | 2566.23| 15.15  |
| Small-15    | 2493.47              | 2943.25| 18.04  |
| Small-20    | 3053.02              | 3639.01| 19.19  |
| Small-25    | 3451.15              | 3940.67| 14.18  |
| Small-30    | 3643.22              | 4093.21| 9.88   |
| Small-35    | 3846.87              | 4492.44| 16.78  |
| Small-40    | 4125.70              | 4595.23| 11.38  |
| Small-45    | 4270.61              | 4866.41| 14.42  |
| Small-50    | 4810.87              | 5508.38| 14.50  |

*: premature stop with an end-of-calculation-time status

Comparing the results of [18] with those provided by our IRP-LT, we realise that the use of LT cannot replace the periodic deliveries from the supplier. However, LT can minimize the loss due to the increase of L by decreasing the total cost of the distribution network by 58.92%, when $L = 2$. This observation reinforces our argument on the usefulness of using LT to improve our research results under such conditions. In conclusion, LT is able to optimize the results of a supply chain with $L \neq 1$ and its effectiveness is more visible when the number of customers is large.

### 6 Conclusion

In this paper, we have examined new IRP structures, in which we try to satisfy customer demands and find the optimal quantity of products to ship for each delivery period. We also determine the best delivery routes by an optimisation process for each period under several constraints. In this context, we studied the effects of increasing the supplier’s lead time and adding the LT technique to the supply chain network.

Throughout this work, we developed mathematical models and tested the performance of each model using instances of [18]. We first compared the results of the new models (with and without LT), then with those of [18]. Finally, we proved the effectiveness of using LT to improve the performance of the supply chain network.

In future research, we can test our model on larger benchmark instances, in terms of number of customers and replenishment periods, or using real cases. We can try to solve the IRP-LT heuristically and see the benefits of such methods. We can propose other replenishment policies to test and compare the results with those of this work. We can also propose another network structure with more than two levels and using multi-vehicle delivery.

### References

1. W.J. Bell and L.M. Dalberto and M.L. Fisher and A.J. Greenfield and R. Jaikumar and P. Kedia and R.G. Mack and P.J. Prutzman (1983) Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer, Interfaces, 13 (6), 4–23. https://doi.org/10.1287/inte.13.6.4
2. A. Federgruen and P.H. Zipkin (1984) A combined vehicle routing and inventory allocation problem, Operations Research, 32, 1019–1037. https://doi.org/10.1287/opre.32.5.1019
3. Y. Liao and J. Li and X. Hu and Y. Li and W. Shen (2020) Application of Lateral Transshipment in Cost Reduction of Decentralized Systems, Sustainability, 12 (12), 1–20. https://doi.org/10.3390/su12125081
4. D. Gross (1963) Centralized inventory control in multi-location supply systems. In: H.E. Scarf and D.M. Gilford and M.W. Shelly, (eds.) Multistage Inventory Models and Techniques, Stanford University Press, Stanford, California, 47–84.

5. M. Tlili and M. Moalla and J.P. Campagne (2012) The transshipment problem in a two-echelon, multi-location inventory system with lost sales, International Journal of Production Research, Applied Mathematical Modelling, 50 (13), 3547–3559. https://doi.org/10.1080/00207543.2012.671590

6. X. Li and L. Sun and J. Gao (2013) Coordinating preventive lateral transshipment between two locations, Computers Industrial Engineering, 66 (4), 933–943. https://doi.org/10.1016/j.cie.2013.05.002

7. R. Patriarca and F. Costantino and G. Di-Gravio (2016) Inventory model for a multi-echelon system with unidirectional lateral transshipment, Expert Systems with Applications, 65, 372–382. https://doi.org/10.1016/j.eswa.2016.09.001

8. J. Meissner and O.V. Senicheva (2018) Approximate dynamic programming for lateral transshipment problems in multi-location inventory systems, European Journal of Operational Research, 265 (1), 49–64. https://doi.org/10.1016/j.ejor.2017.06.049

9. Y. Li and Y. Liao and X. Hu and W. Shen (2019) Lateral Transshipment with Partial Request and Random Switching, Omega, 92, 102134. https://doi.org/10.1016/j.omega.2019.102134

10. Y. Wang and Z.S. Dong and S. Hu (2020) A stochastic prepositioning model for distribution of disaster supplies considering lateral transshipment, Socio-Economic Planning Sciences, 74, 100930. https://doi.org/10.1016/j.seps.2020.100930

11. L.C. Coelho and J.F. Cordeau and G. Laporte (2012) The inventory-routing problem with transshipment, Computer and Operations Research, 39, 2537–2548. https://doi.org/10.1016/j.cor.2011.12.020

12. M. Rabbani and M. Baghersad and R. Jafari (2013) A new hybrid GA-PSO method for solving multi-period inventory routing problem with considering financial decisions, Journal of Industrial Engineering and Management, 6 (4), 909–929. https://doi.org/10.3926/jiem.629

13. Z. Jemai and Y. Rekik and R. Kalai (2013) Inventory routing problems in a context of vendor-managed inventory system with consignment stock and transshipment, Production Planning Control: The Management of Operations, 24 (8-9), 671–683. https://doi.org/10.1080/09537287.2012.666844

14. A. Haliza and H. Paezah and M.Y. Zaitul and A. Melisa and R.S. Sarifah (2014) Location Routing Inventory Problem with Transshipment (LRIP-T), In Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management, Bali, Indonesia, 1595–1605.

15. M. Rabbani and A. Sabbaghnia and M. Mobini and J. Razmi (2018) A graph theory based algorithm for a multi-echelon multi-period responsive supply chain network design with lateral-transshipments, Operational Research, 18, 1–21. https://doi.org/10.1007/s12351-018-0425-y

16. I.T. Peres and H.M. Repolho and R. Martinelli and N.J. Monteiro (2017) Optimization in inventory-routing problem with planned transshipment: A case study in the retail industry, International Journal of Production Economics, 193, 748–756. https://doi.org/10.1016/j.ijpe.2017.09.002

17. S.M. AleHashem and Y. Rekik and E.M. HoseinhaJou (2019) A hybrid L-shaped method to solve a bi-objective stochastic transshipment-enabled inventory routing problem, International Journal of Production Economics, 209, 381–398. https://doi.org/10.1016/j.ijpe.2017.06.020

18. C. Archetti and L. Bertazzi and G. Laporte and M.G. Speranza (2007) A branch-and-cut algorithm for a vendor-managed inventory-routing problem, Transportation Science, 41, 382–391. https://doi.org/10.1287/trsc.1060.0188

19. A. Jabbarzadeh and M. Haughton and A. Khosrojerdi (2018) Closed-loop Supply Chain Network Design under Disruption Risks: A Robust Approach with Real World Application, Computers & Industrial Engineering, 116, 178–191. https://doi.org/10.1016/j.cie.2017.12.025

20. A.C. Van-Wijk and J.J. van Houtum (2018) Optimal lateral transshipment policies for a two location inventory problem with multiple demand classes, European Journal of Operational Research, 272(2), 481–495. https://doi.org/10.1016/j.ejor.2018.06.033

21. H. Hernández-Pérez and I. Rodríguez-Martín and J.J. Salazar-González (2008) A hybrid GRASP/VND heuristic for the one-commodity pickup-and-delivery traveling salesman problem, Computers & Operations Research, 36 (5), 1639–1645. https://doi.org/10.1016/j.cor.2008.03.008

22. E.O. Pacheco and P.S. Cannella and R. Lüders and A.P. Barbosa-Povoa (2017) Order-up-to-level policy update procedure for a supply chain subject to market demand uncertainty, Computers Industrial Engineering, 113, 347–355. https://doi.org/10.1016/j.cie.2017.09.015