Mathematical formulations for a two-echelon inventory routing problem

Katyanne Farias * Khaled Hadj-Hamou ** Claude Yugma ***

* Université Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, F-38000 Grenoble, France, (e-mail: katyanne.farias-de-araujo@grenoble-inp.fr).
** Université de Lyon, INSA Lyon, DISP, F-69621 Villeurbanne, France, (e-mail: khaled.hadj-hamou@insa-lyon.fr)
*** Center of Microelectronics in Provence, Site Georges Charpak, École des Mines de Saint-Étienne, F-13541 Gardanne, France, (e-mail: claude.yugma@emse.fr)

Abstract: Integrated supply chain management is a current practice in the search for overall cost minimization. The Vendor Management Inventory is one of those practices, in which the supplier is responsible for managing the inventory of the customers in accordance with pre-determined management policies. The problem that considers transportation decisions in addition to inventory management is referred to as the Inventory Routing Problem (IRP). The IRP considers a system with one supplier that manages the inventory level of a set of customers and defines when and how much products to supply and how to combine customers in routes while minimizing storage and transportation costs. We present an extension of this problem that considers a two-echelon system with indirect deliveries and route decisions at both levels. In this variant, the customers demands have to be met by deliveries through distribution centers with minimum total cost. We propose two mathematical formulations for the two-echelon IRP under different inventory policies. Computational experiments on a set of randomly generated instances evaluate the proposed formulations. The obtained results show that both formulations are able to solve to the proven optimality small-scale instances and one of the models found feasible solutions for almost all instances under the considered inventory policies.

Keywords: Inventory routing problem, two-echelon system, optimization problems, complex systems, distribution systems, mathematical models.

1. INTRODUCTION

The Vendor-Managed Inventory (VMI) is an approach that has been increasingly considered in the literature since a joint management is more efficient than individualized management in terms of global cost. The VMI is an inventory management practice in which the data of customers inventory are shared with the vendor that makes replenishment decisions to the customers based on inventory and supply chain policies (Angulo et al. 2004). Accordingly, anticipating the access to information increases the chance of avoiding stockouts.

The Inventory Routing Problem (IRP) can be seen as an extension of the VMI that also includes route decisions. The classic IRP involves one supplier that defines the delivery quantities, when the deliveries are made and how to combine the deliveries in routes with the objective of meeting the demands of customers over a time horizon at a minimum total cost. The sequential management of transportation and inventory can generate lower quality solutions due to the trade-off between these two components. On the one hand, if only the transportation is optimized, it potentially leads to solutions with high inventory levels, implying high inventory management costs. On the other hand, optimizing inventory management can require frequent deliveries, generating high transportation costs.

Two replenishment policies are often considered in IRPs: the Maximum Level (ML), where the delivered quantities must respect inventory and vehicle capacities; and the Order-Up-to level (OU), in which, besides respecting the capacities, whenever a replenish is made the inventory level must reach the inventory capacity. Archetti et al. (2007) are probably the first authors to exactly solve the IRP under both ML and OU policies by a Branch-and-Cut (B&C) algorithm. Solyalı and Süräl (2011) presented a similar formulation and a second one that uses a shortest-path network representation for inventory replenishment decisions of customers and solved the problem by a B&C. Archetti et al. (2014) addressed to the multi-vehicle IRP and compared several formulations including valid inequalities through B&C algorithms. More recently, Desaulniers et al. (2016) proposed a branch-price-and-cut for the IRP under the ML policy. For the same problem, Santos et al. (2016) presented a hybrid algorithm based on an Iterated Local Search metaheuristic with a randomized Variable Neighborhood Descent. Coelho et al. (2013) present several variants of the IRP, applications and resolution methods.
Due to the growth of on-line purchases and the imposition of regulatory traffic restrictions in urban areas, a new supply dynamics has become necessary in recent years. In this context, the implementation of Distribution Centers (DC) has proven to be an effective alternative to meet the deadline of customers and reduce traffic, pollution, and noise in urban areas. Trentini et al. (2015) present several deployments of DC in Europe. According to Interface Transport (2004), in the cities of Monaco and La Rochelle, France, the use of DCs contributed to a reduction of about 38% and 48% in the emission of polluting gases and 40% and 61% in the noise levels, respectively.

Therefore, DCs have been introduced on the outskirts of the city to coordinate freight flows inside and outside the urban areas. In the literature, several supply chain configurations have been studied, particularly with regard to two-echelon vehicle routing problems, as highlighted by Cuda et al. (2015). However, most of the works do not consider the inventory management in the problem or deal with very specific conditions. Given the foregoing, we define a Two-Echelon Inventory Routing Problem (2E-IRP), which is an extension of the multi-vehicle IRP with a finite time horizon. In this problem, the customers must be served by the supplier strictly through DCs and routes must be defined in both echelons over a given time horizon. Both the ML and the OU replenishment policy are applied. We introduce two Mixed Integer Linear Programming (MILP) formulations to solve the problem, one with undirected variables indexed on the routes and the other one with flow variables. We conducted computational experiments on a set of randomly generated instances.

The only work identified by us in the literature dealing with a two-echelon IRP was recently proposed by Guimarães et al. (2019). They presented a multi-depot two-echelon IRP under the ML and the OU policies, where each DC has a fleet of homogeneous vehicles that pickup products from a supplier and/or deliver products to customers at a single time period without route decisions between DCs and suppliers. The authors proposed a B&C and a matheuristic algorithm using a subproblem formulation and an adaptive large neighborhood search to solve the problem.

Accordingly, the following differences between the problem addressed by Guimarães et al. (2019) and the problem proposed by us are highlighted: we consider two different fleets of homogeneous vehicles, one available for the supplier and one shared by the DCs, while the authors consider one fleet of homogeneous vehicles for each DC to pickup from the supplier and delivery to customers possibly at the same time period; we consider route decisions in both echelons, while Guimarães et al. (2019) allow routes only on the second echelon; lastly, the authors consider several suppliers with a not constrained availability of products, and we simplify it to one supplier. With regard to the solving methods, Guimarães et al. (2019) proposed the B&C algorithm based on a route indexed formulation. Otherwise, we introduce a formulation with flow variables in addition to a formulation with variables indexed on the routes.

The paper is organized as follows. Section 2 describes the 2E-IRP. Section 3 presents the MILP formulations. The computational results are presented in Section 4. Finally, Section 5 provides the concluding remarks.

2. PROBLEM DESCRIPTION

The 2E-IRP can be shown to be NP-hard via reduction to IRP, which is a special case of 2E-IRP when there is only one DC. In the 2E-IRP, the supplier must meet the demands of customers for a single product through DCs. Direct deliveries from the supplier to a customer are not allowed. Thus, the supplier must provide the DCs and each DC must meet the demands of a given subset of customers. As discussed previously in Section 1, this distribution configuration in which customers are previously assigned to a DC, based on their geographical location for example, occurs in many European cities.

A fleet of homogeneous primary vehicles is available for deliveries from the supplier to DCs and a different fleet of homogeneous secondary vehicles transports the shipments from the DCs to the customers. A primary vehicle can visit several DCs on a route starting and ending at the supplier, as long as its capacity is respected. Similarly, a second vehicle can visit several customers of a DC on a route starting and ending at the DC, respecting its capacity. As considered for IRPs, DCs and customers can be visited at most once per time period. Stockout is not allowed.

Each DC and customer has a maximum inventory level that must be respected. The inventory capacity of the supplier is not restrictive. We address two variants of 2E-IRP in the context of replenishment policies: when the ML policy is applied for DCs and customers and when the OU is considered at both echelons. In this problem, delivery quantities and routes at each time period should be defined in both echelons so that the final customer demands are met during each time period. The objective is to find a solution with minimal total cost. Figure 1 illustrates the 2E-IRP.

3. MATHEMATICAL FORMULATIONS

We propose two Mixed integer linear programming (MILP) formulations for the 2E-IRP: one route index MILP formulation with undirected variables for routing decisions, referred to as FI, and one model with flow variables, denoted as FI II. Given a set of vertex \( V = V^0 \cup V^1 \cup V^2 \), where \( V^0 = \{0\} \) represents the supplier, \( V^1 = \{1, \ldots, m\} \) is the set of \( m \) DCs and \( V^2 = \{m+1, \ldots, m+n\} \) is the set of \( n \) customers. Each DC \( u \in V^1 \) must meet the demands of the subset \( V^2_u \subseteq V^2 \) of customers associated with it. The supplier, in turn, must meet the demands of the customers through DCs. The problem involves two echelons \( e \in \{1, 2\} \), where the first echelon concerns to the deliveries from the supplier to the DCs and the second echelon refers to the deliveries from each DC to its customers.

The objective of this problem is to meet demands of the customers over a time horizon \( T \) at a minimum total cost. Therefore, each customer \( i \in V^2 \) has a demand \( d^i_t \) per time period \( t \in T \). The total cost is composed by: vehicle fixed costs \( f^e \), for each echelon \( e \in \{1, 2\} \), that is considered each time a primary or secondary vehicle is used; transportation costs \( c_{i,j} \), if a vehicle travels from \( i \) to \( j \), equal to a unit cost multiplied by the distance; order costs \( g_{i,j} \in V^1 \cup V^2 \),...
each time a DC or a customer receives an amount of the product; and inventory costs \( h_i, i \in V \), per each period that a unit of inventory is held.

The inventory capacities of DCs and customers \( (C_i, i \in V^1 \cup V^2) \) and vehicles capacities \( (Q^1, Q^2) \) for primary and secondary vehicles, respectively) must be respected. Each echelon \( e \in \{1, 2\} \) has a number of \( \kappa^e \) available vehicles. Initial inventory levels are known in advance. A quantity \( r^e \) of product is made available at the supplier at each time period \( t \in T \).

### 3.1 Formulation I

The 2E-IRP can be defined on an undirected graph \( G = (V, E) \), where \( E = E^1 \cup E^2 \) is the set of edges of the first and second echelons. Thus, the edge set \( E^1 = \{(i, j) | i, j \in V^1 \cup V^1, j < i \} \) contains the edges between the supplier and the DCs and the edge set \( E^2 = \{(i, j) | u \in V^1, i, j \in V^2_u \cup \{u\}, j < i \} \) contains the edges linking each DC and its customers. In addition, each echelon \( e \in \{1, 2\} \) has a set \( \mathcal{K}^e = \{0, \ldots, \kappa^e - 1\} \) of possible routes associated with.

The model FI uses the following variables: (i) the integer variables \( x_{i,j}^{k,t} \), representing the number of times that the edge \((i, j) \in E^e \) of the echelon \( e \in \{1, 2\} \) is traveled by a vehicle in the route \( k \in \mathcal{K}^e \) at time \( t \in T \); (ii) the binary variables \( y_{i,j}^{k,t} \), taking value 1 if vertex \( i \in V^{e-1} \cup V^e \) is visited in route \( k \in \mathcal{K}^e \) at time \( t \in T \) for the echelon \( e \in \{1, 2\} \), and 0 otherwise; (iii) the continuous variables \( q_{i,j}^{k,t} \), that represent the quantity delivered to \( i \in V^e \), \( e \in \{1, 2\} \), via the route \( k \in \mathcal{K}^e \) at time \( t \in T \); and (iv) the continuous variables \( I^e_t \) that models the inventory level of \( i \in V \) at the end of the time period \( t \in T \).

The objective function (1) aims at minimizing the sum of fixed vehicle costs, transportation costs, order costs and inventory holding costs. Constraints (2), (3) and (4) define the inventory level of the supplier, of the DCs and of the customers for each time period, respectively. The inventory capacities of DCs and customers are respected through constraints (5). Constraints (6) impose that if a quantity is delivered to a customer or a DC through a route at a time period, the respective assignment variable must be equal to one. The OU policy is introduced by constraints (7). The capacities of primary and secondary vehicles are respected, respectively, through constraints (8) and (9). Constraints (10)–(13) are degree and subtour elimination
constraints at both echelons. Constraints (14) oblige that DCs and customers are visited at most once at each period. Constraints (15) restrict the number of routes per echelon and time period to the number of available vehicles at each echelon. Constraints (16)–(20) define the variables domain.

3.2 Formulation II

The 2E-IRP can also be modeled by making use of a directed graph $G = (V, A)$, where $A = A^1 \cup A^2$ contains the arcs of the first and second echelon. Therefore, the arc set $A^1 = \{(i, j) | i, j \in V^1 \cup V^2, j \neq i\}$ is composed by the arcs between the supplier and the DCs and the arc set $A^2 = \{(i, j) | u \in V^1, i, j \in V^2 \cup \{u\}, j \neq i\}$ connects each DC and its customers.

The variables of this model do not consider route indexes. Similarly to formulation FI, model FII uses the following variables: (i) the binary variables $x^1_{i,j}$, assuming value 1 if the arc $(i, j) \in A$ is traveled by a vehicle at time $t \in T$, and 0 otherwise; (ii) the continuous variables $q^1_{i,j}$ that represent the quantity delivered to $i \in V^1 \cup V^2$ at time $t \in T$; and (iii) the continuous variables $I^1_t$ representing the inventory level at each time period. In addition, FII contains (iv) the binary variables $U^1_t$, taking value 1 if $i \in V^1 \cup V^2$ is visited at $t \in T$, 0 otherwise, and (v) the continuous variables $w^1_{i,u}$ that represent the freight flow passing through the arc $(i, j) \in A$ at time $t \in T$.

\begin{align*}
\text{FII:} & \quad \min \left[ \sum_{i \in T} \left( \sum_{u \in V^1} \left( f^1_i x^1_{0,u} + \sum_{j \in V^2} f^1_j x^1_{u,j} \right) + \right. \right. \\
& \quad \left. \left. \sum_{(i,j) \in A^1} c_{i,j} x^1_{i,j} + \sum_{i \in V^2} g^1_i \right) \right] + \sum_{t \in T} \sum_{i \in V^1} h^1_i t^1_i \quad (21) \\
\text{s.t.} & \quad t^0 = t^1_0 + r^t - \sum_{u \in V^1} q^1_{u,0} \quad t \in T \quad (22) \\
& \quad t^1_t = t^1_{t-1} + q^1_{t-1} - q^1_{t} \quad t \in V^2 \quad (23) \\
& \quad t^1_t = t^1_{t-1} + q^1_i - d^1_i \quad i \in V^2, t \in T \quad (24) \\
& \quad q^1_e \leq C_i - t^1_{i-1} \quad e \in \{1, 2\}, i \in V^2, t \in T \quad (25) \\
& \quad q^1_i \leq \min\{Q^2, C_i\} U^1_t \quad i \in \{1, 2\}, i \in V^2, t \in T \quad (26) \\
& \quad q^1_i \geq C_i U^1_t - t^1_{i-1} \quad i \in \{1, 2\}, i \in V^2, t \in T \quad (27) \\
& \quad q^1_{u,v} \leq \min\{Q^2, C_u\} \sum_{v \in V^1 \cup \{0\}} x^1_{v,u} \quad u \in V^1, t \in T \quad (28) \\
& \quad q^1_{u,v} \leq \min\{Q^2, C_v\} \sum_{j \in V^2 \cup \{0\}} x^1_{u,j} \quad j \neq i \quad (29) \\
& \quad \sum_{u \in V^1 \cup \{0\}} x^1_{0,u} \leq 1 \quad u \in V^1, t \in T \quad (30) \\
& \quad \sum_{j \in V^2 \cup \{0\}} x^1_{i,j} \leq 1 \quad u \in V^1, i \in V^2, t \in T \quad (31) \\
& \quad \sum_{u \in V^1 \cup \{0\}} x^1_{u,0} \quad u \in V^1 \cup \{0\}, t \in T \quad (32) \\
& \quad x^1_{i,j} = \sum_{u \in V^1} x^1_{i,u} + \sum_{j \in V^2} x^1_{u,j} \quad i \in V^1 \cup \{u\}, t \in T \quad (33) \\
& \quad x^1_{i,j} \leq \sum_{u \in V^1} x^1_{i,u} \quad i \in V^1 \cup \{u\}, t \in T \quad (34) \\
& \quad x^1_{i,j} \leq \sum_{u \in V^2} x^1_{i,u} \quad i \in V^1 \cup \{u\}, t \in T \quad (35) \\
& \quad \sum_{v \in V^1 \cup \{0\}} w^1_{u,v} - q^1_{u,v} = \sum_{v \in V^1 \cup \{0\}} w^1_{u,v} \quad u \in V^1, t \in T \quad (36) \\
& \quad \sum_{v \in V^2 \cup \{0\}} w^1_{i,v} - q^1_{i,v} = \sum_{v \in V^2 \cup \{0\}} w^1_{i,v} \quad i \in V^1, t \in T \quad (37) \\
& \quad w^1_{u,v} = 0 \quad \{ u = 0, i \in V^1, t \in T \} \quad (38) \\
& \quad w^1_{i,j} \leq Q^2 x^1_{i,j} \quad i \in V^1 \cup \{0\}, j \in T \quad (39) \\
& \quad U^1_t \in \{0, 1\} \quad i \in V^1 \cup \{0\}, t \in T \quad (40) \\
& \quad \sum_{i \in T} \sum_{u \in V^1} x^1_{i,u} \leq 1 \quad u \in V^1, i \in V^2, t \in T \quad (41) \\
& \quad \sum_{u \in V^1} x^1_{u,v} \leq 1 \quad i \in V^1 \cup \{0\}, i \in V^2, t \in T \quad (42) \\
& \quad w^1_{i,j} \geq 0 \quad \{ e = 1, i \in V^1 \cup \{0\}, j \neq i, t \in T \} \quad (43) \\
& \quad \sum_{i \in T} \sum_{u \in V^1} x^1_{i,u} \leq 1 \quad u \in V^1, i \in V^2, t \in T \quad (44) \\
\end{align*}

The objective function (21) and constraints (22)–(25) and (27) have the same meaning of (1), (2)–(5) and (7) from FII, but written in terms of flow variables in FII. Constraints (26) link the decision variables of the delivery quantity to the binary variable that indicates if a DC or a customer is visited at a time period. Constraints (28) and (29) impose, respectively, that if a DC or a customer is supplied at a time period, it must be part of a route at the same time period. Constraints (30) and (31) define that DCs and customers are visited at most once per time period. Constraints (32) and (33) are the flow conservation for the first and the second echelon, respectively. The number of vehicles available at the first and the second echelon is respected through constraints (34) and (35). Constraints (36)–(38) are the subtour elimination constraints and constraints (39) establish that the vehicles capacity is not exceeded. Constraints (40)–(44) specify the domains of the variables.

4. COMPUTATIONAL EXPERIMENTS

The computational experiments were conducted on a personal computer equipped with an 2.40 GHz Intel Core i5-6300U processor and 8 GB of RAM, with a maximum running time of one hour. The methods were implemented in C++ language and solved using CPLEX 12.7.1 and a single thread.

In order to evaluate the performance of formulations FII and FII, a set of instances has been generated based on the instances of the IRP proposed by Archetti et al. (2007). The geographic configuration of the instances respects the arrangement of customers and DCs implemented in many cities of Europe. For each combination of time horizon $|T| = \{3, 6, 10\}$ and number of DCs $m = \{3, 4, 5\}$, five instances have been randomly created, resulting in a total
of 45 instances. The number of customers $n_u$ associated with each DC $u \in V^1$ was randomly generated from the interval $[5, 10]$. Thus, the total number of customers is $n = \sum_{u \in V^1} n_u$. The demands $d_i$, $i \in V^2$, and the quantity made available at the supplier per time period are constant over the time horizon. The data were randomly generated for the first echelon and by

$$\sum_{i=1}^{\alpha} h_i = 0.01;$$

$$g_i = 5l_i,$$

where $l_i$ is selected from $\{6, 7, 8, 9, 10\}$ for $i \in V^1$ and from $\{1, 2, 3, 4\}$ for $i \in V^2$.

The zones of customers for each DC are not overlapped. Thus, the coordinates $(x_i, y_i)$ of the supplier, the DCs and the customers are randomly selected from the interval $[0, 10000]$, where each customer is at a distance less than or equal to 500 from its DC. The transportation costs are equal to the rounded Euclidean distances multiplied by $\alpha^1$ for the first echelon and by $\alpha^2$ for the second echelon, where $\alpha^1$ and $\alpha^2$ are respectively selected from the interval $[0.01, 0.05]$ and $[0.06, 0.10]$.

Computational experiments were conducted for the 2E-IRP under the ML and OU policies. For the ML policy, we consider the model FI without constraints (7) and the formulation FII without constraints (27).

Table 1 shows the results obtained by formulations FI and FII under the ML and OU policies for each group of instances. The first column indicates the replenishment policy. The second column gives the instance group name in the format $TmM$, where $t$ is the time horizon and $m$ is the number of DCs. The following columns give for each instance group, the results obtained by the models FI and FII, respectively, in terms of: the number of times a feasible solution was found respecting the maximum running time of one hour ($\#\text{Feasible}$); the number of times a proven optimal solution was found within one hour of execution ($\#\text{Optimal}$); the average gaps between the upper bound and the lower bound (Gap); the average gaps between the best upper bound (between FI and FII) at the end of the execution and the LB of the root node (GapLR); the average running times (Times (s)) in seconds; and the average number of open nodes ($\#\text{Nodes}$).

The gap of the root node analyzes the lower bound of the first node given by the CPLEX with automatic cuts. The gap of the linear relaxation evaluates the lower bound of the linear relaxation without cuts. The gaps GapROOT and GapLR are calculated using the best upper bound between the upper bounds found by the models FI and FII to analyze the quality of the lower bounds using the same basis for comparison. The abbreviation “elim” indicates that the running time limit has been reached. We consider, for each execution, a gap equal to zero when a proven optimal solution is found and a gap equal to one when no feasible solutions found. The best results from the two models are highlighted in bold.

All instances of the first three instance groups, i.e., those with a time horizon equal to 3, when submitted to the ML policy, were solved to the proven optimality by both mathematical formulations. However, the model FII proved the optimality in a shorter running time. The model FI opened a smaller number of nodes on the branch-and-bound tree, but presented a longer running time than FII due to the larger number of constraints and variables, requiring thus more time to solve each node. The formulation FII presented better average gaps at the root node and the gap of the linear relaxation is on average slightly better than those by FII for all instance groups.

Still under the ML policy, for the group of instances with the time horizon equal to 6, the formulation FI proved the optimality of four instances, while FII did not prove the optimality of any instance. However, the model FII found feasible solutions for all instances while FI did not find feasible solutions for two instances of the group 6T5M. Therefore, for this instance group the gap obtained by FII is much lower than the Gap of FI. For instances with 10 time periods, the model FII was able to find feasible solutions for all instances, while the model FI did not find feasible solutions for 6 out of 15 instances, which is evidenced in the gaps of instance groups 10T4M and 10T5M. The number of open nodes by FI is lower due to the largest size of this model in terms of constraints and variables.

When the OU policy is considered, the average gaps at the root node obtained by FII are lower on all instance groups. The model FII found proven optimal solutions for all instances with a time horizon equal to 3. Contrarily, the formulation FI did not prove the solution optimality of 6 instances on this group. When an optimal solution was proven by both models, FII always presented a lower running time. For the group of instances with a time horizon equal to 6 and 10, the feasible solutions found by both methods were not proven to be optimal. In general, the formulation FII found a larger number of feasible solutions and presents gaps significantly lower than the ones obtained by FI.

5. CONCLUSION

We proposed a two-echelon inventory routing problem under the two classic replenishment policies: the ML and the OU policies. This problem is inserted in a current context of supply chains, in the search for the reduction of the overall costs of the system. We introduced two MILP formulations (FI and FII) to solve the problem. In general,
the model FII presented better results compared to the FI. However, both models were able to solve to the proven optimality only small-scale instances with a time horizon equal to 3.

The next step of this research will consist in evaluating the performance of a Branch-and-Cut algorithm based on the FI formulation in addition to a heuristic method to generate high quality initial solutions. Given the exponential number of subtour elimination constraints, it may be advantageous to dynamically add some of them only when they are violated. We also intend to develop a matheuristic solution method for solving big-scale instances and adjust the problem to real cases, where we try to generalize the problem to multi products and multi suppliers.

**REFERENCES**


