Available online at www.sciencedirect.com

ScienceDirect

European Journal of Operational Research 189 (2008) 1022–1040

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/ejor

A new model and hybrid approach for large scale inventory routing problems

Yugang Yu, Haoxun Chen *, Feng Chu

Institute of Computer Science and Engineering of Troyes, Industrial System Optimization Group, University of Technology of Troyes, 10010 Troyes, France

Received 1 October 2005; accepted 1 February 2007
Available online 15 July 2007

Abstract

This paper studies an inventory routing problem (IRP) with split delivery and vehicle fleet size constraint. Due to the complexity of the IRP, it is very difficult to develop an exact algorithm that can solve large scale problems in a reasonable computation time. As an alternative, an approximate approach that can quickly and near-optimally solve the problem is developed based on an approximate model of the problem and Lagrangian relaxation. In the approach, the model is solved by using a Lagrangian relaxation method in which the relaxed problem is decomposed into an inventory problem and a routing problem that are solved by a linear programming algorithm and a minimum cost flow algorithm, respectively, and the dual problem is solved by using the surrogate subgradient method. The solution of the model obtained by the Lagrangian relaxation method is used to construct a near-optimal solution of the IRP by solving a series of assignment problems. Numerical experiments show that the proposed hybrid approach can find a high quality near-optimal solution for the IRP with up to 200 customers in a reasonable computation time.

Keywords: Logistics; Inventory routing problem; Inventory; Transportation; Lagrangian relaxation

1. Introduction

The inventory routing problem (IRP) is concerned with the distribution of a single product from a single facility (depot) to a set of geographically dispersed customers over a given planning horizon (Campbell et al., 2002; Campbell and Savelsbergh, 2004). It is to determine the delivery quantity for each customer and a set of feasible vehicle routes for the delivery of the quantities in each period, subject to the vehicle capacity constraints and the customers' product requirements and inventory capacity constraints, so that a total inventory and transportation cost is minimized. The problem arises in many distribution systems, especially in vendor-managed inventory systems.

* Corresponding author. Present address: ISTIT-OSI, University of Technology of Troyes, 12 Rue Marie Curie, BP 2060, Troyes 10010, France. Tel.: +33 325 71 56 42; fax: +33 325 71 56 49.

E-mail addresses: ygyums@gmail.com (Y. Yu), haoxun.chen@utt.fr (H. Chen), feng.chu@utt.fr (F. Chu).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2007.02.061

The coordination of inventory and transportation decisions in multiple periods is the key for the optimization of the IRP, because: 1) Inventory costs and transportation costs are at odds. It is usually cheaper in terms of transportation cost to ship a full truckload of a product to customers. However, the customer demand of the product may be far less than a full truckload. When the product is delivered in a full truckload, it may be stocked for a long time before it is consumed, leading to a higher inventory cost. Thus, a good trade-off between inventory and transportation costs should be made in order to minimize the total cost. This can be achieved by coordinating inventory and transportation decisions in multiple periods. 2) The utilization of the vehicle resource can be well balanced by delivering more than the demand of a product to customers in some periods with lower demand.

The IRP has found its applications in many industrial sectors, including oil and gas delivery (Adelman, 2004; Reiman, 1999). It has been attracting the attention of many researchers in academic communities. In the literature, the IRP has been presented with different time horizons and assumptions about the nature of demand. Single-period deterministic, multiple period deterministic, single-period stochastic, and multiple period stochastic models and solution methods have been proposed.

The single-period IRP with deterministic demand is actually a classical vehicle routing problem (VRP). Dror and Trudeau (1989) first introduced the split delivery VRP (SDVRP) by relaxing a constraint of the VRP that every customer is served by only one vehicle. They showed how this relaxation could lead to important savings, both in the total distance traveled and in the number of vehicles used. The SDVRP remains NP-hard despite this relaxation (Dror and Trudeau, 1990). This leads to further studies in the development of approximate approaches for the problem (Belenguer et al., 2000; Ho and Haugland, 2004; Lee et al., 2006).

For the multiple period IRP with deterministic demand, Campbell et al. (2002) and Campbell and Savelsbergh (2004) developed a two-phase approach, where a delivery schedule is first created by solving an integer programming problem, followed by the construction of a set of delivery routes by using heuristics. For the IRP with constant demand, Anily and Bramel (2004) proposed a fixed partition policy, where all customers are partitioned into disjoint and collectively exhaustive sets and the customers in each set are served independently of the customers in the other sets. They derived a lower bound on the cost of an optimal fixed partition policy. A probabilistic analysis of this bound demonstrates that the policy is asymptotically 98.5%-effective. The same problem was also studied by Aghezzaf et al. (2006) and Zhao et al. (2007). The former proposed a column generation-based approximation method with the resulted routing sub-problems solved by using a savings-based heuristic, whereas the latter used a tabu search method to find the customers' optimal partition regions for the fixed partition policy. For other work on deterministic multiple period IRP, readers can refer to Campbell et al. (2002), Campbell and Savelsbergh (2004) and Schwarz et al. (2004).

Federgruen and Zipkin (1984) first studied the single-period inventory routing problem with stochastic demand. They showed how some well-known heuristics for the deterministic VRP can be modified to handle the problem modeled as a stochastic nonlinear integer program. Federgruen et al. (1986) studied the problem in the distribution of perishable products and showed that significant cost savings can be achieved by using an integrated inventory planning and routing approach. Since more opportunities may be generated by coordinating deliveries in multiple periods (Campbell and Savelsbergh, 2004), the multiple period IRP with stochastic demand was studied by Adelman (2004), Dror and Ball (1987), Dror et al. (1985), Jaillet et al. (2002), Kleywegt et al. (2002, 2004), Kumar et al. (1995) and Trudeau and Dror (1992). Despite the previous research efforts, the stochastic IRP remains notoriously intractable and the structure of an optimal policy is not known even for the problem with a single-customer (Adelman, 2004; Reiman, 1999). The readers who want to know more about the stochastic IRP please refer to the literature reviews given in Campbell et al. (2002), Campbell and Savelsbergh (2004) and Kleywegt et al. (2002, 2004).

In this paper, we consider the multiple period deterministic inventory routing problem with split delivery (IRPSD) where the demand of each customer in each period over a given planning horizon is assumed to be known and must be satisfied without backorder. The delivery of each customer in each period can be split and performed by multiple vehicles. Allowing split delivery increases the flexibility of distribution, which may further reduce transportation costs (Dror and Trudeau, 1989). However, due to the complexity of the IRP, the split delivery was not considered in the pervious literature (Campbell et al., 2002; Campbell and Savelsbergh, 2004; Kleywegt et al., 2002, 2004) except for Chandra and Fisher (1994), Fumero and Vercellis (1999), and Yu et al. (2005).

We use a deterministic model rather than a stochastic model for the IRP because of the high complexity of the stochastic IRP mentioned above. In practice, future customer demands may be estimated (forecasted) with allowable errors, especially for near future demands. These demand estimates can be used as inputs of our model. Our planning approach also adopts a rolling horizon framework as in Jaillet et al. (2002). That is, although in each period, the inventory and routing plans are generated for the whole time horizon, only the plans for the current period are really implemented. As time progresses, the time horizon is moved forward and the unimplemented plans for future periods will be updated using new demand forecast data.

The previous work directly relating to our present research includes Chandra and Fisher (1994), Fumero and Vercellis (1999), and Yu et al. (2005). Chandra and Fisher (1994) proposed an integrated multi-period multi-product production and distribution system model based on a multi-stop routing problem formulation with additional setup constraints and split delivery relaxation. They presented a two-phase solution procedure: the multi-item lot-sizing sub-problem is first solved, and the distribution plan is then generated by using a heuristic algorithm. However, Yu et al. (2007) found that a feasible solution of their model might not define a feasible solution for the original production and distribution problem they considered. In fact their model only offers a lower bound for the original problem. Moreover, the subtour elimination constraints in their model are quite complex, which makes it difficult to cope with large scale problems. Fumero and Vercellis (1999) developed a multi-period integrated model for a single plant logistical system, in which multiple items are manufactured and delivered to customers. Production, inventory, and routing decisions are considered in the same model, which is solved by using Lagrangian relaxation. However, although the vehicles considered are homogeneous, they introduced in their model decision variables for each specific vehicle, which makes the number of decision variables increase considerably as the problem size increases. Another problem of their model is that the transportation cost for any empty vehicle traveling from one customer to another is zero so that some vehicle routes in an optimal solution of the model may contain a partial route in which an empty vehicle first goes to a customer with minimum direct return cost to the depot rather than directly returns to the depot. The same problem exists in our previous work (Yu et al., 2005), which adopted the same transportation cost structure of Fumero and Vercellis (1999). The model to be proposed in the present paper will eliminate such partial routes.

In order to solve large scale IRPSD, instead of providing an exact model for the problem, we propose an approximate model, whose solution only defines the quantity delivered to each customer, the quantity transported through each directed arc and the number of times that each directed arc is visited by vehicles, where a directed arc connects two customers or a customer and the central depot in the corresponding transportation network. The approximate model makes the formulation of a much larger problem with the same number of variables as used in other models possible. For example, to formulate the IRPSD problem with 100 customers and 30 vehicles per period, the model of Fumero and Vercellis (1999) requires nearly 6×10^5 variables, whereas with the same number of variables our model can formulate the problem with almost 500 customers. In addition, new subtour elimination constraints for the vehicle routing sub-problem in each period are introduced in our model. Compared with that of Chandra and Fisher (1994), our model uses much fewer but tighter subtour elimination constraints. If N customers are considered, only N subtour elimination constraints are required in our model, whereas 2^N constraints are required in their model.

The present work also improves our previous model presented in Yu et al. (2005) by introducing a set of constraints to avoid empty partial routes and other valid constraints to reduce the solution space. However, with the new constraints, the Lagrangian relaxation approach developed in Yu et al. (2005) becomes sensitive to problem parameters. To overcome this difficulty, we develop a new and more robust Lagrangian relaxation approach by constraint reformulation and by using a different constraint relaxation framework. Since the model's solution obtained by the Lagrangian relaxation may be not feasible for the original IRPSD problem, a procedure based on sequentially solving a set of assignment problems is proposed to transform it into a feasible solution of the IRPSD. The assignment problems assign the quantity transported through each directed arc and the number of times that each directed arc is visited by vehicles in the corresponding transportation network to a set of feasible vehicle routes. Finally, a simple local search is used to further improve the quality of the routes, leading to a near-optimal solution of the IRPSD.

The remainder of the paper is organized as follows: Section 2 describes the problem and outlines our solution methodology. Section 3 presents an approximate model of the IRPSD and its properties. Section 4 is dedicated to the Lagrangian relaxation approach for the model. The construction of a feasible solution of the model and its transformation to a near-optimal solution of the IRPSD are presented in Section 5. Computational results are given in Section 6. Section 7 concludes the paper with some remarks.

2. Problem description and solution methodology outline

We study the multiple period IRPSD with a set of customers, a central depot, and a fleet of homogeneous vehicles with limited capacity, for which in each period the depot has to deliver sufficient units of a product to each customer to completely fulfill its demand, which is deterministically known. No backorder is allowed at each customer but each customer may hold a local inventory used to meet future demands with a holding cost. The delivery to each customer in each period can be performed by multiple vehicles. The objective is to minimize over a given time horizon the total inventory holding and transportation cost, which is the sum of the inventory holding costs of all customers and the transportation costs for all deliveries to the customers.

The inventory cost of each customer linearly depends on its inventory level at the end of each period. The transportation cost includes not only fixed usage costs, which are related to vehicle insurance, depreciation, and drivers' rewards, but also variable shipping costs depending on both transported quantity and traveled distance. This transportation cost structure, first adopted by Fumero and Vercellis (1999), not only can approximately model purely distance proportional cost components in classical VRP but also reflects the transportation pricing in practice.

The solution methodology for finding a near-optimal solution of the IRPSD is outlined in Fig. 1. 1) An approximate model of the IRPSD is developed based on some important properties of the problem. 2) Lagrangian relaxation is used to decompose the model into an inventory sub-problem and a routing sub-problem which are solved by a linear programming algorithm and a minimum cost flow (MCF) algorithm, respectively. 3) A feasible solution of the model is constructed based on the results of the Lagrangian relaxation. 4) The solution, which may be not feasible for the original IRPSD, is transformed into a feasible one by solving a series of assignment problems. 5) The feasible solution is further improved by a simple local search to obtain a near-optimal solution of the IRPSD. 6) In order to evaluate the quality of the final solution, another Lagrangian relaxation approach with exact decomposition is used to calculate a lower bound of the IRPSD.

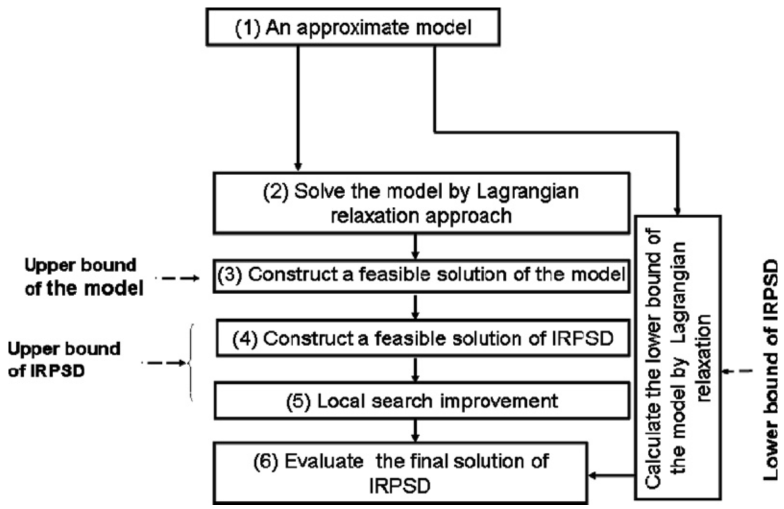

Fig. 1. The solution methodology for IRPSD.

3. Model

The notation used in the model is given as follows.

Indices
$t = 1, \ldots, T$ period index
$i,j = 0, 1, \ldots, N$ customer or depot index, where $i,j = 1, \ldots, N$ represent N customers, and 0 represents the central depot

Parameters
C vehicle capacity
c_{ij} shipping cost per unit of product along arc (i,j) where $c_{ij} = c_{ji}$ and the triangle inequality, $c_{ik} + c_{kj} \geqslant c_{ij}$, holds for any i, j, k with $k \neq i$, $k \neq j$
c_{i0}^b cost of traveling with an empty vehicle from customer i back to the depot in period t
f_t fixed vehicle cost per tour in period t
h_{it} holding cost per unit of product at customer i in period t
I_{i0} initial inventory level of customer i at the beginning of period 1
M size of the vehicle fleet (number of available vehicles)
r_{it} demand of customer i in period t
V_i inventory capacity of customer i

Variables
d_{it} delivery quantity to customer i in period t
I_{it} inventory level of customer i at the end of period t
q_{ijt} quantity transported through the directed arc (i,j) in period t
x_{ijt} number of times that the directed arc (i,j) is visited by vehicles in period t

With the notation, an approximate model of the IRPSD, denoted by P, is given as
Model P:

$$Z = \min \sum_{t=1}^{T} \sum_{i=1}^{N} h_{it} I_{it} + \sum_{t=1}^{T} \sum_{j=0, j\neq i}^{N} \sum_{i=0}^{N} c_{ij} q_{ijt} + \sum_{t=1}^{T} \sum_{i=1}^{N} f_t x_{i0t} + \sum_{t=1}^{T} \sum_{i=1}^{N} c_{i0}^b x_{i0t} \tag{1}$$

subject to

$$I_{it} = I_{i,t-1} + d_{it} - r_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{2}$$

$$I_{i,t-1} + d_{it} \leqslant V_i, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{3}$$

$$\sum_{j=0, j\neq i}^{N} x_{ijt} = \sum_{j=0, j\neq i}^{N} x_{jit}, \quad i = 0, \ldots, N, \ t = 1, \ldots, T, \tag{4}$$

$$\sum_{j=0, j\neq i}^{N} q_{jit} - \sum_{j=0, j\neq i}^{N} q_{ijt} = d_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{5}$$

$$\sum_{i=1}^{N} q_{0it} = \sum_{i=1}^{N} d_{it}, \quad t = 1, \ldots, T, \tag{6}$$

$$\sum_{i=1}^{N} x_{i0t} \leqslant M, \quad t = 1, \ldots, T, \tag{7}$$

$$q_{ijt} \leqslant C \cdot x_{ijt}, \quad i,j = 0, \ldots, N, \ i \neq j, \ t = 1, \ldots, T, \tag{8}$$

$$I_{it} \geqslant 0, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{9}$$

$$x_{ijt} \geqslant 0 \quad \text{and integer } i = 0, \ldots, N, \ j = 0, \ldots, N, \ i \neq j, \ t = 1, \ldots, T, \tag{10}$$

$$d_{it} \geqslant 0, \quad i = 1, \ldots, N, \ q_{ijt} \geqslant 0, \ i = 0, \ldots, N \ j = 1, \ldots, N, \ j \neq i, \ t = 1, \ldots, T. \tag{11}$$

The objective function (1) includes both inventory costs of each customer and transportation costs (fixed and variable costs). Constraints (2) are the inventory balance constraints of each customer. Constraints (3) are the inventory capacity constraints of each customer. Constraints (4) ensure that the number of vehicles leaving from a customer or the depot is equal to the number of its arrival vehicles. Constraints (5) are the product flow conservation equations, assuring the flow balance at each customer and eliminating all subtours. Constraints (6) assure the collection of accumulative delivery quantity at the depot. Constraints (7) assure that the number of vehicles used for delivery in each period does not exceed the size of the vehicle fleet. Constraints (8) model the vehicle capacity and logical relationship between q_{ijt} and x_{ijt}. Constraints (9) ensure that each customer's demand is completely fulfilled without backorder.

Theorem 1. *Consider the SDVRP with N customers for which a positive quantity d_i must be delivered to customer i. Let x_{ij} and q_{ij} be the number of times that the directed arc (i,j) is visited by vehicles and the quantity transported through the directed arc (i,j), respectively, $i,j = 0,1,\ldots,N$, $i \neq j$, then the following constraints (12) and (13)*

$$\sum\nolimits_{j=0,j\neq i}^{N} q_{ji} - \sum_{j=0,j\neq i}^{N} q_{ij} = d_i, \quad i = 1,\ldots,N, \tag{12}$$

$$q_{ij} \leqslant C \cdot x_{ij}, \quad i = 0,\ldots,N, \; j = 1,\ldots,N, \; i \neq j, \tag{13}$$

1) *are tighter than the constraints (14):*

$$\sum_{j\in U_0\setminus S} \sum_{i\in S} x_{ji} \geqslant \sum_{i\in S} d_i/C \quad \forall S \subseteq \{1,\ldots,N\} \tag{14}$$

proposed by Chandra and Fisher (1994), *where* $U_0 = \{0,1,\ldots,N\}$;

2) *give a necessary condition for defining a feasible solution of the SDVRP*;

3) *eliminate all subtours of the SDVRP.*

Proof. 1) For $\forall S \subseteq \{1,\ldots,N\}$, $\sum_{i\in S}(\sum_{j=0,j\neq i}^{N} q_{ji} - \sum_{j=0,j\neq i}^{N} q_{ij}) = \sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ji} - \sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ij} \leqslant \sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ji}$.

Considering constraints (12), we have

$$\sum_{j\in U_0\setminus S} \sum_{i\in S} q_{ji} - \sum_{j\in U_0\setminus S} \sum_{i\in S} q_{ij} = \sum_{i\in S} d_i \quad \text{and} \quad \sum_{j\in U_0\setminus S} \sum_{i\in S} q_{ji} \geqslant \sum_{i\in S} d_i.$$

From constraints (13) and $q_{ij} \leqslant C \cdot x_{ij}$, we have $\sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ji} \leqslant C \cdot \sum_{j\in U_0\setminus S}\sum_{i\in S} x_{ji}$.

$$\text{So } C \cdot \sum_{j\in U_0\setminus S} \sum_{i\in S} x_{ji} \geqslant \sum_{i\in S} d_i, \quad \text{that is,} \quad \sum_{j\in U_0\setminus S} \sum_{i\in S} x_{ji} \geqslant \sum_{i\in S} d_i/C.$$

2) The result is obvious and its proof is omitted.

3) (by contradiction) subtours are the tours that do not depart from and return to the depot. The subtours can be traced by the arcs with $x_{ij} > 0$. If there is a subtour satisfying the condition (12), select all customers in the subtour to compose a set S. For this subtour, $x_{ji} = x_{ij} = q_{ji} = q_{ij} = 0$ $j \in U_0 \setminus S$, $i \in S$ and $\exists x_{ji} > 0$ $i,j \in S$. According to constraints (12), we have $\sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ji} - \sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ij} = \sum_{i\in S} d_i$ and $\sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ji} = \sum_{j\in U_0\setminus S}\sum_{i\in S} q_{ij} = 0$, thus $\sum_{i\in S} d_i = 0$. That is, $d_i = 0$, $\forall i \in S$. Since the SDVRP is a minimum problem, we have $x_{ij} = q_{ij} = 0$ $\forall i,j \in S$ according to constraints (13). This contradicts the existence of a subtour defined by S. □

Note: Constraints (12) and (13) together are neither a sufficient condition for defining a feasible solution of the SDVRP, nor a necessary condition for its subtour elimination. However the model P provides a lower bound for the IRPSD since any solution of the IRPSD is also a solution of the model.

The following theorem and property will be used for deriving a tighter model P' of the IRPSD with a reduced solution space containing an optimal solution of P.

Theorem 2. *If c_{ij} $i,j = 1,\ldots,N$, $i \neq j$ satisfy the triangle inequality and model P is feasible, then*
1) *P has an optimal solution with no two routes having more than one common customer.*

$$2)\ x_{ijt} \in \{0,1\}, x_{i0t}, x_{0it} \geqslant 0 \quad \text{and integer } i,j = 1,\ldots,N,\ i \neq j. \tag{15}$$

This theorem was proved by Dror and Trudeau (1990) for the SDVRP with standard VRP cost structure. It is also applicable to our model P. The proof of the theorem is thus omitted here.

Property 1. *For model P, if the same conditions of Theorem 2 hold, then,* 1) $x_{0it} \geqslant \lceil q_{0it}/C \rceil$ $i = 1,\ldots,N$ *for any feasible solution of P*; 2) *At least one optimal solution* (x^*, q^*) *of P satisfies* $x^*_{0it} = \lceil q^*_{0it}/C \rceil$ $i = 1,\ldots,N$; *and*

$$3)\ (x^*_{0it} - 1)C \leqslant q^*_{0it}, \quad i = 1,\ldots,N. \tag{16}$$

Proof. 1) From vehicle capacity constraints $q_{0it} \leqslant x_{0it}C$, $i = 1,\ldots,N$, we have $x_{0it} \geqslant C/q_{0it}$. Because x_{0it} $i = 1,\ldots,N$ are integers, $x_{0it} \geqslant \lceil C/q_{0it} \rceil$ holds.

2) If in an optimal solution (x^*, q^*) of model P there is a customer $i, i \in \{1,\ldots,N\}$ and a period t, $t \in \{1,\ldots,T\}$ not satisfying $x^*_{0it} = \lceil q^*_{0it}/C \rceil$, then $x^*_{0it} \geqslant \lceil q^*_{0it}/C \rceil + 1$, because $x_{0it} \geqslant \lceil q_{0it}/C \rceil$ according to 1). Consequently, according to constraints (4) there is a customer k, $k \in \{1,\ldots,N\}$ $\quad k \neq i$ with $x^*_{ikt} = 1$, and $q_{0it} \geqslant q_{ikt}$ since model P is feasible. Modifying the solution with $x_{ikt} = 1 - 1 = 0$, $x_{0it} = x^*_{0it} - 1$, $\quad x_{0kt} = x^*_{0kt} + 1$, $q_{0it} = q^*_{0it} - q^*_{ikt}$, $\quad q_{ikt} = 0 \quad q_{0kt} = q^*_{0kt} + q^*_{ikt}$ and with no change of other variables, we obtain a new solution (x, q) satisfying all constraints of P, which is also a feasible solution of the model. According to the triangle inequality of c_{ij}, $i,j = 1,\ldots,N$, the transportation cost of the new solution is reduced by $(c_{0i} + c_{ik} - c_{0k})q^*_{ikt} \geqslant 0$. At the same time, the total number of times that all arcs are visited by vehicles in the corresponding transportation network is reduced by 1, i.e., $\sum_{j=0,j\neq i}^{N}\sum_{i=0}^{N} x_{ijt} = \sum_{j=0.j\neq i}^{N}\sum_{i=0}^{N} x^*_{ijt} - 1$. The above procedure can be repeated if there is still a customer and a period not satisfying $x_{0it} = \lceil q_{0it}/C \rceil$. Since $\sum_{j=0,j\neq i}^{N}\sum_{i=0}^{N} x^*_{ijt}$ is finite, this procedure cannot be repeated infinitely. It implies that an optimal solution of P with no customer violating the condition $x^*_{0it} = \lceil q^*_{0it}/C \rceil$ will be finally obtained when the procedure terminates.

3) Since $x^*_{0it} = \lceil q^*_{0it}/C \rceil$, we have $x^*_{0it} \leqslant q^*_{0it}/C + 1$. The condition $(x^*_{0it} - 1)C \leqslant q^*_{0it}$ is thus satisfied. □

More valid constraints can be identified under a mild assumption to eliminate all routes with an empty vehicle traveling from one customer to another in any solution of model P. For this purpose, suppose that in model P all quantity related parameters, such as I_{i0}, r_{it}, C and V_i, are integers. With this assumption, it is very rare that in an optimal solution the quantity transported on each arc $q_{ijt} < 1$. Without loss of optimality, we assume that either $q_{ijt} = 0$ with $x_{ijt} = 0$, or $q_{ijt} \geqslant 1$ with $x_{ijt} \geqslant 1$. It follows that

$$x_{ijt} \leqslant (q_{ijt} - 1)/C + 1, \quad i = 0,1,\ldots,\ N, j = 1,\ldots,N,\ i \neq j,\ t = 1,\ldots,T. \tag{17}$$

With constraints (17), x_{ijt} must be 0 when $q_{ijt} = 0$. This eliminates a potential problem that empty vehicles may travel between two customers in an optimal solution.

With the above theorem and property, we can obtain a tighter model of the IRPSD with additional constraints (15) to replace (10), constraints (16) and (17). However, in the model, x_{i0t} and x_{0it} are integers. In order to effectively solve the model by using a Lagrangian relaxation approach described in the next section, we introduce a tighter model P' after transforming x_{0it} from integer variables into binary variables by introducing the following additional parameters and variables.

M_i maximal number of vehicles left from the depot to customer i. These vehicles are indexed by $m = 1,\ldots,M_i$, $M_i \leqslant M$

p_{mit} quantity transported by the mth vehicle on directed arc $(0, i)$ in period t, with $\sum_{m=1}^{M_i} p_{mit} = q_{0it}$

y_{mit} $y_{mit} = 1$ if the mth vehicle left from the depot to customer i, 0 otherwise; $\sum_{m=1}^{M_i} y_{mit} = x_{0it}$

The model P' can now be formulated as follows:

Model P′:

$$Z = \min \sum_{t=1}^{T}\sum_{i=1}^{N} h_{it}I_{it} + \sum_{t=1}^{T}\sum_{j=1,j\neq i}^{N}\sum_{i=1}^{N} c_{ij}q_{ijt} + \sum_{t=1}^{T}\sum_{i=1}^{N} c_{0i}\left(\sum_{m=1}^{M_i} p_{mit}\right) + \sum_{t=1}^{T}\sum_{i=1}^{N} f_t\left(\sum_{m=1}^{M_i} y_{mit}\right) + \sum_{t=1}^{T}\sum_{i=1}^{N} c^b_{i0}x_{i0t} \tag{18}$$

subject to constraints (2), (3), (7), (9), (17) and

$$\sum_{j=0,j\neq i}^{N} x_{ijt} = \sum_{m=1}^{M_i} y_{mit} + \sum_{j=1,j\neq i}^{N} x_{jit} \quad i=1,\ldots,N,\ \ t=1,\ldots,T \quad \text{(for customers)}, \tag{19a}$$

$$\sum_{i=1}^{N} x_{i0t} = \sum_{i=1}^{N}\sum_{m=1}^{M_i} y_{mit}, \quad t=1,\ldots,T \quad \text{(for the depot)}, \tag{19b}$$

$$\sum_{m=1}^{M_i} p_{mit} + \sum_{j=1,j\neq i}^{N} q_{jit} - \sum_{j=0,j\neq i}^{N} q_{ijt} = d_{it}, \quad i=1,\ldots,N,\ \ t=1,\ldots,T, \tag{20}$$

$$\sum_{i=1}^{N}\sum_{m=1}^{M_i} p_{mit} = \sum_{i=1}^{N} d_{it}, \quad t=1,\ldots,T, \tag{21}$$

$$q_{ijt} \leqslant x_{ijt}C, \quad i,j=1,\ldots,N,\ \ i\neq j,\ \ t=1,\ldots,T \quad \text{(for customers)}, \tag{22a}$$

$$\sum_{m=1}^{M_i} p_{mit} \leqslant C\cdot\sum_{m=1}^{M_i} y_{mit}, \quad i=1,\ldots,N,\ \ t=1,\ldots,T \quad \text{(for the depot)}, \tag{22b}$$

$$C\cdot\left(\sum_{m=1}^{M_i} y_{mit} - 1\right) \leqslant \sum_{m=1}^{M_i} p_{mit}, \quad i=1,\ldots,N,\ \ t=1,\ldots,T, \tag{23}$$

$$y_{mit}, x_{ijt} \in \{0,1\}, \quad x_{i0t} \geqslant 0 \text{ and integer } i,j=1,\ldots,N,\ \ i\neq j,\ \ m=1,\ldots,M_i,\ \ t=1,\ldots,T, \tag{24}$$

$$d_{it}, \quad p_{mit}, \quad q_{ijt} \geqslant 0, \quad i=1,\ldots,N,\ \ j=0,\ldots,N,\ \ j\neq i,\ \ m=1,\ldots,M_i,\ \ t=1,\ldots,T, \tag{25}$$

where constraints (19a) and (19b) are the reformulation of constraints (4); constraints (20) and (21) are the reformulations of constraints (5) and (6), respectively; constraints (22a) and (22b) are the reformulation of constraints (8); constraints (23) are the reformulation of constraints (16); constraints (24) and (25) are the reformulations of constraints (10) and (11), respectively. Note that in model P', vehicle specific variables are introduced only for the directed arcs from the depot to a customer, the number of vehicle specific variables in the model is thus much less than that of Fumero and Vercellis (1999) especially when the vehicle fleet size M is large. The number of integral variables in model P' is

$$N\times(N-1)\times T + \left(\sum_{i=1}^{N} M_i\right)\times T + N\times T \leqslant N\times N\times T + M\times N\times T,$$

whereas the number of integral variables in their model is $M\times(N+1)\times N\times T$. For example, if $N=100$, $T=10$, and $M=40$, model P' has less than 140,000 integral variables, whereas their model has 4,040,000 integral variables which can be used to model a problem of more than $N=500$ customers with our model P'.

4. Lagrangian relaxation approach

4.1. Relaxation framework

Model P' is NP-hard since it is more difficult than general SDVRP. This inspires us to seek an approximate approach to solve the model. In this section we present a Lagrangian relaxation (LR) approach for finding a near-optimal solution of model P'.

The constraints that complicate the resolution of this model are constraints (17), (22a), (22b) and (23), which couple q_{ijt}, p_{mit} with integral variables x_{ijt} and y_{mit}. They can be relaxed by introducing non-negative Lagrange multipliers. However, this kind of relaxation can not lead to a relaxed problem whose solution contains useful information for constructing a satisfactory feasible solution of the model. In order to have an effective relaxed problem, the constraints (22a) and (22b) are substituted by the following equivalent ones.

$$q_{ijt}(1-x_{ijt})=0,\quad i,j=1,\dots,N,\ i\neq j,\ t=1,\dots,T, \tag{26a}$$
$$q_{ijt}\leqslant C,\quad i,j=1,\dots,N,\ i\neq j,\ t=1,\dots,T, \tag{26b}$$
$$p_{mit}(1-y_{mit})=0,\quad m=1,\dots,M_i,\ i=1,\dots,N,\ t=1,\dots,T, \tag{27a}$$
$$p_{mit}\leqslant C,\quad i=1,\dots,N,\ m=1,\dots,M_i,\ t=1,\dots,T. \tag{27b}$$

Without loss of optimality, the following constraints similar to (17) are also considered

$$y_{mit}\leqslant (p_{mit}-1)/C+1,\quad i=1,\dots,N,\ m=1,\dots,M_i,\ t=1,\dots,T. \tag{28}$$

By introducing Lagrange multipliers $\lambda=\{\lambda_{ijt}\}_{N\times(N-1)\times T}$, $\alpha=\{\alpha_{it}\}_{N\times T}$, $\bar{\alpha}=\{\bar{\alpha}_{it}\}_{M_i\times N\times T}$, $\gamma=\{\gamma_{ijt}\}_{N\times(N-1)\times T}$ and $\bar{\gamma}=\{\bar{\gamma}_{ijt}\}_{M_i\times N\times T}$ to relax constraints (17), (23), (28), (26a) and (27a), respectively, we obtain the relaxed problem of P′, denoted by RP, as

Model RP:

$$Z_{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}}(q,p,x,y)=\min Z+\lambda^T g_1+\alpha^T g_2+\bar{\alpha}^T g_3+\gamma^T g_4+\bar{\gamma}^T g_5, \tag{29}$$

subject to constraints (2), (3), (7), (9), (19a),(19b),(20), (21), (24), (25), (26b) and (27b), where

$$g_1(q,x)=\{x_{ijt}-(q_{ijt}-1)/C-1\},\quad g_2(p,y)=\left\{C\left(\sum_{m=1}^{M_i}y_{mit}-1\right)-\sum_{m=1}^{M_i}p_{mit}\right\},$$
$$g_3(p,y)=\{y_{mit}-(p_{mit}-1)/C-1\},\quad g_4(q,x)=\{q_{ijt}(1-x_{ijt})\},\quad g_5(p,y)=\{p_{mit}(1-y_{mit})\}.$$

The relaxed problem, however, cannot be exactly decomposed into sub-problems because of the coupling terms $q_{ijt}x_{ijt}$ and $p_{mit}y_{mit}$ in its objective function. Nevertheless, it can be approximately solved by alternatively solving its two sub-problems, one with given $\{q_{ijt},p_{mit}\}$ and the other with given $\{x_{ijt},y_{mit}\}$.

The sub-problem with (x,y) given by $(\widehat{x},\widehat{y})=(\{\widehat{x}_{ijt}\},\{\widehat{y}_{mit}\})$, denoted by INV$(\widehat{x},\widehat{y})$ determines the q, p values and can be formulated as

$$Z^1_{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}}(q,p,\widehat{x},\widehat{y})=\min\sum_{t=1}^{T}\sum_{i=1}^{N}h_{it}I_{it}+\sum_{t=1}^{T}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{i=1}^{N}\left(c_{ij}-\frac{\lambda_{ijt}}{C}+\gamma_{ijt}(1-\widehat{x}_{ijt})\right)q_{ijt}$$
$$+\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M_i}(c_{0i}-\alpha_{it}-\frac{\bar{\alpha}_{mit}}{C}+(1-\widehat{y}_{mit})\bar{\gamma}_{mit})p_{mit}, \tag{30}$$

subject to constraints (2), (3), (9), (20), (21), (25), (26b) and (27b). In the sub-problem, $\widehat{x}_{ijt}$ and $\widehat{y}_{mit}$ are the predetermined values of variables x_{ijt} and y_{mit}, respectively.

From constraints (2), we have $I_{it}=I_{i0}+\sum_{\tau=1}^{t}(d_{i\tau}-r_{i\tau})$. The sub-problem INV can then be reformulated as

$$Z^1_{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}}(q,p,\widehat{x},\widehat{y})=\min\sum_{t=1}^{T}\sum_{i=1}^{N}h_{it}\left(I_{i0}+\sum_{\tau=1}^{t}(d_{i\tau}-r_{i\tau})\right)$$
$$+\sum_{t=1}^{T}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{i=1}^{N}\left(c_{ij}-\frac{\lambda_{ijt}}{C}+\gamma_{ijt}(1-\widehat{x}_{ijt})\right)q_{ijt}$$
$$+\sum\nolimits_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M_i}\left(c_{0i}-\alpha_{it}-\frac{\bar{\alpha}_{mit}}{C}+(1-\widehat{y}_{mit})\widehat{\gamma}_{mit}\right)p_{mit}, \tag{31}$$

subject to constraints (20), (21), (25), (26b), (27b) and

$$I_{i0}+\sum_{\tau=1}^{t}d_{i\tau}-\sum_{\tau=1}^{t-1}r_{i\tau}\leqslant V_i,\quad i=1,\dots,N,\ t=1,\dots,T, \tag{32}$$
$$I_{i0}+\sum\nolimits_{\tau=1}^{t}(d_{i\tau}-r_{i\tau})\geqslant 0,\quad i=1,\dots,N,\ t=1,\dots,T, \tag{33}$$

where constraints (32) and (33) are the reformulations of constraints (3) and (9), respectively.

The sub-problem with (q,p) given by $(\widehat{q},\widehat{p}) = (\{\widehat{q}_{ijt}\},\{\widehat{p}_{ijt}\})$, denoted by $\mathrm{ROU}(\widehat{q},\widehat{p})$, determines the x and y values and can be formulated as

$$Z^2_{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}}(\widehat{q},\widehat{p},x,y) = \min\sum_{t=1}^{T}\sum_{i=1}^{N}(c^b_{i0}+f_t)x_{i0t} + \sum_{t=1}^{T}\sum_{j=1,j\neq i}^{N}\sum_{i=1}^{N}(\lambda_{ijt}-\gamma_{ijt}\widehat{q}_{ijt})x_{ijt} + \sum\nolimits_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M}(\alpha_{it}C+\bar{\alpha}_{mit}-\bar{\gamma}_{mit}\widehat{p}_{mit})y_{mit}, \tag{34}$$

subject to constraints (7), (19a), (19b) and (24). In the sub-problem, $\widehat{q}_{ijt}$ and $\widehat{p}_{mit}$ are the predetermined values of variables q_{ijt} and p_{mit} respectively.

For ROU, it can be further decomposed into T sub-problems, one for each period, given by

$$Z^{2t}_{\lambda_{(t)},\alpha_{(t)},\bar{\alpha}_{(t)},\gamma_{(t)},\bar{\gamma}_{(t)}}(\widehat{q}_{(t)},\widehat{p}_{(t)},x_{(t)},y_{(t)}) = \min\sum_{i=1}^{N}(c^b_{i0}+f_t)x_{i0t} + \sum_{j=1,j\neq i}^{N}\sum_{i=1}^{N}(\lambda_{ijt}-\gamma_{ijt}\widehat{q}_{ijt})x_{ijt} + \sum\nolimits_{i=1}^{N}\sum_{m=1}^{M_i}(\alpha_{it}C+\bar{\alpha}_{mit}-\bar{\gamma}_{mit}\widehat{p}_{mit})y_{mit} \tag{35}$$

subject to

$$\sum\nolimits_{j=0,j\neq i}^{N}x_{ijt} = \sum_{m=1}^{M_i}y_{mit} + \sum_{j=1,j\neq i}^{N}x_{jit}, \quad i=1,\ldots,N \quad \text{(for customers)}, \tag{36a}$$

$$\sum\nolimits_{i=1}^{N}x_{i0t} = \sum_{i=1}^{N}\sum_{m=1}^{M_i}y_{mit} \quad \text{(for the depot)}, \tag{36b}$$

$$\sum\nolimits_{i=1}^{N}x_{i0t} \leqslant M, \tag{37}$$

$$x_{ijt}\in\{0,1\}, \quad j\neq i,\ \ y_{mit}\in\{0,1\},\ \ x_{i0t}\geqslant 0 \text{ and integer},\ \ i=1,\ldots,N. \tag{38}$$

Let $D(\mu)$ be the optimal objective value of RP for any Lagrange multiplier vector $\mu = \{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}\}$. The Lagrangian dual problem of RP, denoted by DP, can be formulated as

Model DP:

$$\max_{\mu\geqslant 0} D(\mu), \tag{39}$$

where $D(\mu) = \max\{Z_\mu(q,p,x,y) = Z_{\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}}(q,p,x,y) \mid \text{s.t. } (2),(3), (7), (9), (19a), (19b), (20), (21), (24), (25), (26b) \text{ and } (27b)\}$.

4.2. Resolution of the relaxed problem

To solve the Lagrangian dual problem DP, for any given Lagrange multiplier vectors $\lambda,\alpha,\bar{\alpha},\gamma,\bar{\gamma}$, the corresponding sub-problems INV and ROU need to be solved. The sub-problem INV is a linear program, which can be easily solved by using the simplex algorithm. To improve the performance of the Lagrangian relaxation approach, one additional technique is adopted. That is, a very small perturbation, $\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M_i}10^{-5}\cdot m\cdot p_{mit}$, is added to the objective function (31) to distinguish $p_{1it},\ldots,p_{M_iit}$. This perturbation will make $p_{1it},\ldots,p_{M_iit}$ in a non-increasing order, i.e., $p_{1it}\geqslant p_{2it}\geqslant\cdots\geqslant p_{M_iit}$, in any solution of sub-problem INV at each iteration of the Lagrangian dual maximization.

The sub-problem ROU can be transformed into a min-cost flow problem using a similar method from Fumero and Vercellis (1999). In the method, the depot, indexed 0, is duplicated by adding an artificial depot indexed by $0'$. The two depots are then connected with an arc of capacity equal to the vehicle fleet size M. Moreover, the original arcs incident to node 0 (i.e. $y_{mit}\in\{0,1\}, x_{i0t}\ i=1,\ldots,N$) are divided into two groups which are associated with the two depots respectively so that depot 0 has only entering arcs (i.e. x_{i0t}, $i=1,\ldots,N$) and depot $0'$ has only leaving arcs (i.e. $y_{mit}\in\{0,1\}$, $i=1,\ldots,N$), with an exception of the newly added arc, denoted by $x_{0,0',t}$, connecting the two depots. In this way, we can obtain a modified model, denoted by ROU′, of the above ROU by replacing constraints (36b) and (37) with the following three constraints:

$$\sum_{i=1}^{N} x_{i0t} = x_{0,0',t} \quad \text{(for the depot 0)}, \tag{40}$$

$$x_{0,0',t} = \sum_{i=1}^{N}\sum_{m=1}^{M_i} y_{mit} \quad \text{(for the depot 0}'\text{)}, \tag{41}$$

$$x_{0,0',t} \leqslant M \text{ and integer}. \tag{42}$$

Note that the integer requirements of the decision variables of ROU′ are not necessary since its constraint matrix is totally unimodular and its all right-hand side constants are integers (Wolsey, 1998). In other words, solving ROU′ as a linear program using the simplex method always yields an integral solution. Moreover, ROU′ is a minimum cost flow (MCF) problem that can be efficiently solved by using the out-of-kilter algorithm, Klein, Jewell, Busacker & Gowan's method, etc. (Wolsey, 1998). These algorithms run in polynomial time and are more efficient than the simplex algorithm. In order to improve the performance of the Lagrangian relaxation approach, a very small perturbation, such as $\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M_i}10^{-5}\cdot m\cdot y_{mit}$, is again added to the objective function (34) to distinguish $y_{1it},\ldots,y_{M_iit}$.

4.3. Resolution of the dual problem by surrogate subgradient method

The Lagrangian dual problem is usually solved by using the subgradient (SG) method. However the application of the SG method requires that the relaxed problem is optimally solved at each iteration of the dual optimization. Since the relaxed problem we consider in Section 4.1 is only approximately solved, the SG method is no longer applicable.

Fortunately, a subgradient-like method called surrogate subgradient (SSG) method was recently developed to solve the Lagrangian dual problem in case of an approximate resolution of the relaxed problem (Zhao et al., 1999). When the relaxed problem is approximately solved, the method still ensures the convergence of the dual problem to its optimal solution under some conditions.

SSG is similar to SG except for the definition of the subgradient and the step sizing scheme for the update of Lagrange multipliers. For our dual problem DP, the surrogate subgradient is given by

$$g(q,p,x,y) = \{g_1(q,x), g_2(p,y), g_3(p,y), g_4(q,x), g_5(p,y)\},$$

where (q,p,x,y) is the solution of the relaxed problem RP obtained by using the method described in Sections 4.1 and 4.2.

Using an adaptive step sizing scheme, the procedure of the SSG method is given as follows.

Step 0. Initiation: $k=0$, $\mu^0=0$, and $\theta=1$, where k is the iteration index, θ is a parameter for step sizing.

Step 1. Solve inventory sub-problem INV(x^{k-1}, y^{k-1}) and routing sub-problem ROU(q^{k-1},p^{k-1}) with the Lagrange multipliers μ^k, where (x^{k-1},y^{k-1}) and (q^{k-1},p^{k-1}) are the solution of sub-problem ROU and the solution of INV obtained at the last iteration (iteration $k-1$), respectively. At $k=0$, neither of the two sub-problems depends on (x^{k-1},y^{k-1}) or (q^{k-1},p^{k-1}) and they can be solved independently.

Step 2. Set step size s^k as $s^k=\beta(D^*-Z^k)/\|g^k\|^2$, where β is a step size with $0<\beta<1$; $Z^k=Z_{\mu^k}(q^k,p^k,x^k,y^k)$ is the surrogate dual at the current iteration k; $g^k=\{g_1(q^k,x^k),g_2(p^k,y^k),g_3(p^k,y^k), g_4(q^k,x^k), g_5(p^k,y^k)\}$ is the surrogate subgradient at the current iteration k; D^* is the optimal objective value of the dual problem, which is estimated by $(1+\frac{\omega}{\theta^\rho})Z^{[k]}$, where $Z^{[k]}$ is the best surrogate dual obtained prior to iteration k, $\theta=\max(1,\theta-1)$ if $Z^k>Z^{[k]}$ and $\theta=\theta+1$ otherwise. Parameters ω and ρ are taken as $\omega\in[0.1,1.0]$, $\rho\in[1.1,1.5]$.

Step 3. Update the Lagrange multipliers in iteration $k+1$: $\mu^{k+1}=\max\{\mu^k+s^kg^k,0\}$.

Step 4. Check the stopping criterion: 1) Z^k is not improved for a given number of iterations, or 2) A given maximal total iteration number is reached. If the criterion is met, stop and output all required results. Otherwise, set $k=k+1$ and go to Step 1.

The convergence conditions for SSG are:

$$Z_{\mu^0}(q^0, p^0, x^0, y^0) < D^*, \tag{43}$$

$$Z_{\mu^k}(q^k, p^k, x^k, y^k) < Z_{\mu^k}(q^{k-1}, p^{k-1}, x^{k-1}, y^{k-1}), \tag{44}$$

where (q^k, p^k, x^k, y^k) is the solution of the relaxed problem at multiplier vector μ^k. D^* is the optimal objective value of the dual problem.

It is easy to show that the first condition (43) always holds and the second condition (44) holds if we replace "$<$" by "$\leqslant$" in the inequality. The case "$=$" rarely happens. If it happens, the relaxed sub-problems INV and ROU can be solved once again with the updated values of $\{x, y\}$ and $\{q, p\}$ obtained respectively from the resolution of the sub-problems in the last time to improve the current solution (q^k, p^k, x^k, y^k) until the second condition holds.

Since the surrogate dual is not a Lagrangian dual in a strict sense, its value may exceed the minimum objective value of the original problem P'. Consequently, the best surrogate dual obtained by SSG in the Lagrangian relaxation approach is not a lower bound of P'.

5. Construction and evaluation of feasible solutions

5.1. Construction of a feasible solution of model P'

If we substitute d, q, p in model P' by the solution of Lagrangian relaxed problem RP in Section 4.2, we can obtain a minimal cost flow problem which can be decomposed into T sub-problems, one for each period, as sub-problem ROU. A feasible solution of model P' can then be obtained by solving the minimal cost flow problem.

In order to obtain a better upper bound (a better feasible solution) of P', we can construct a feasible solution in last several iterations or in every iteration of the Lagrangian relaxation approach. The best feasible solution is selected as the final solution of P'.

5.2. Transformation of the solution of P' into a feasible solution of the IRPSD

The feasible solution d, q, p, x, y of P' obtained in Section 5.1 may not define a feasible solution for the original IRPSD because P' is only an approximate model of IRPSD. In order to obtain a feasible solution of the IRPSD, a transformation procedure is proposed in the following to trace (construct) in every period a set of feasible routes based on the solution of model P'. Because the same procedure is applied for the feasible route tracing in each period, we will omit the period index – subscript t – in relevant variables and parameters in the following discussion.

Given a solution d, q, p, x, y of model P', a directed transportation network (DTN) can be defined for every period, as illustrated in Fig. 2, where two customer nodes i and j are connected by a directed arc (i,j) if $x_{ij} = 1$, customer node i and the depot node 0 are connected by a directed arc $(i, 0)$ for x_{i0} times if $x_{i0} \geqslant 1$, and the depot node 0 and customer node i are connected by a directed arc $(0, i)$ for x_{0i} times if $x_{0i} \geqslant 1$, where $x_{0i} = \sum_{m=1}^{M_i} y_{mi}$. The directed arcs associated with $\{x_{ji} | x_{ji} > 0, j = 0, 1, \ldots, N\}$ are called incoming arcs of customer node i, and the directed arcs associated with $\{x_{ij} | x_{ij} > 0, j = 0, \ldots, N\}$ are called outgoing arcs of customer node i. In the DTN, the number of vehicles arriving at a customer node i and that departing from the node are equal according to constraints (36a) and (36b), and $\{q_{ji} | q_{ji} > 0,\ j = 0, 1, \ldots, N\}$ (with $q_{0i} = \sum_{m=1}^{M_i} p_{mi}$) and $\{q_{ij} | x_{ij} > 0, j = 0, \ldots, N\}$ forms the inflows and outflows of the node, respectively.

In the DTN, if the number of incoming arcs or the number of outgoing arcs of each customer node is one, that is, each customer's delivery is performed by a single vehicle, then a set of feasible routes which constitute a feasible solution of the IRPSD can be naturally traced from the solution of P'. Otherwise, there exists at least one customer node whose number of incoming arcs or number of outgoing arcs is greater than 1, as nodes 2 and 3 in Fig. 2. In this case, some customers are common customers of multiple vehicles and two conditions must hold for tracing feasible routes: 1) Any incoming arc of each customer node must be matched with one of its outgoing arc; 2) for each pair of matched arcs, the flow of the incoming arc must be no less than the flow of the outgoing arc. If 1) or 2) cannot be satisfied, we have to modify the solution of P' to obtain a feasible solution of the IRPSD.

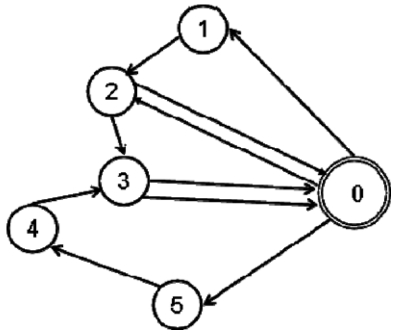

Fig. 2. Directed transportation network.

In the following, we present a procedure for constructing a feasible solution of the IRPSD from a solution of P'. This procedure assigns the arcs of the DTN to a set of routes by solving a series of assignment problems defined for each customer node on the DTN. Each assignment problem matches the incoming arcs of a customer node with its outgoing arcs. Because the order of solving these assignment problems is critical for their successful resolution, which leads to a feasible solution of the IRPSD, we first number all nodes on the DTN starting from the depot node and following the directions of the arcs, where the depot node is numbered as 0. A customer node can be numbered next if and only if its all preceding nodes on the DTN have already numbered. The successful numbering of all customer nodes requires that the DTN has no subtour. Although subtours may appear in a solution of P' obtained by using the SSG method, but they can be eliminated with the triangle inequality property of c_{ij}.

After numbering the customer nodes, a set of vehicle routes are traced following the order of customers numbered. During the route tracing, if a customer node has more than one incoming or outgoing arcs, its incoming arcs must be matched with its outgoing arcs in order to trace feasible routes. For this purpose, an assignment problem is solved to determine the matching of the incoming arcs of the customer with its outgoing arcs.

Considering the arc matching of customer node i, if the flow of an incoming arc is no less than the flow of its matched outgoing arc, the corresponding match is a feasible match. Otherwise, the match is infeasible. The primary objective of the assignment problem is to minimize the number of infeasible matches by penalizing them. For the assignment problem of customer node i, if its solution contains infeasible matches, then the values of some variables of x_{ji}, q_{ji}, $j = 0, 1, \ldots, N$ and x_{ij}, $q_{ij}j = 0, \ldots, N$ have to be adjusted so that the feasible route tracing procedure can continue. The secondary objective of the assignment problem is thus to minimize the number of such variables.

Based on the above analysis, the assignment problem for each customer node can be formally defined. Let $l = 1, \ldots, L$ be the incoming/outgoing arc index of customer node i where $L = \sum_{j=0}^{N} x_{ij}$ is the number of vehicles arriving at or departing from the customer node; $q_+^l(i)$ be the flow of the lth incoming arc of customer node i; $q_-^l(i)$ be the flow of the lth outgoing arc of customer node i; $\bar{q}_-^{\min}(i)$ be the minimal outflow of customer node i, that is, $\bar{q}_-^{\min}(i) = \min\{q_-^l(i) | l = 1, \ldots, L\}$; $C_{lm}^A = \overline{M}$ if $q_+^l(i) < q_-^m(i)$ where $\overline{M}$ is a very large positive number, $C_{lm}^A = 0$ otherwise; ${}^l C_{lm}^B = \bar{q}_-^{\min}(k) - q_+^l(i)$ if $q_+^l(i) < q_-^m(i)$ and $\bar{q}_-^{\min}(k) - q_+^l(i) > 0$ where k is the immediate successor of customer node i connected with its mth outgoing arc, $C_{lm}^B = 0$ otherwise. Decision variable $u_{ln} = 1$ if incoming arc l is matched with outgoing arc n,$u_{ln} = 0$ otherwise.

The assignment problem for customer node i $(i = 1, \ldots, N)$ can then be formulated as

Model AP:

$$\min \sum_{l=1}^{L} \sum_{n=1}^{L} (C_{ln}^A + C_{ln}^B) u_{ln}, \tag{45}$$

$$\sum_{l} u_{ln} = 1, \quad n = 1, \ldots, L, \tag{46}$$

$$\sum_{n} u_{ln} = 1, \quad l = 1, \ldots, L, \tag{47}$$

$$u_{ln} \in \{0, 1\}, \quad l, n = 1, \ldots, L. \tag{48}$$

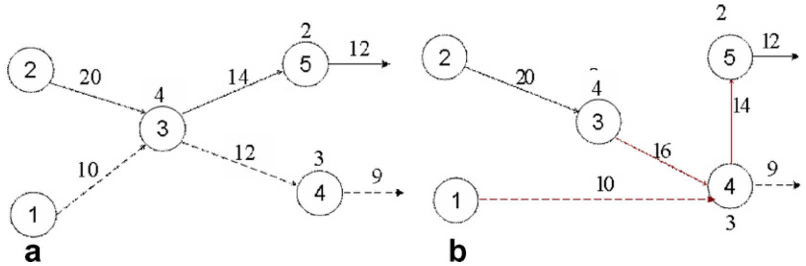

Fig. 3. Adjustment of the solution of P' in tracing feasible routes.

The assignment problem can be easily solved by using a specific algorithm in polynomial time (Dell'Amico et al., 2001; Martello and Toth, 1987).

Once an optimal solution of the AP is obtained, a match of the incoming arcs and the outgoing arcs of customer node i is determined. If the minimum objective value of the AP is zero, the incoming arcs of the customer node are well matched with its outgoing arcs so that the route tracing procedure can proceed to the next customer node. Otherwise, $\sum_{l=1}^{L}\sum_{n=1}^{L}(C_{ln}^{A}+C_{ln}^{B})u_{ln}>0$. In this case, the values of some variables in the solution q, p, x, y of P' have to be adjusted to continue tracing feasible routes.

In the following, we give an example to illustrate this solution adjustment. The example is shown in Fig. 3a , where the number on each directed arc (i, j) and the number on each node i are q_{ij} and d_i, respectively. The values of q, p, x, y are adjusted to those in Fig. 3b by replacing a partial path $1 \rightarrow 3 \rightarrow 5$ by $1 \rightarrow 4 \rightarrow 5$, leading to a feasible match of the incoming/outgoing arcs of customer node 3. The match of incoming arc $1 \rightarrow 3$ with outgoing arc $3 \rightarrow 5$ and the match of incoming arc $2 \rightarrow 3$ with outgoing arc $3 \rightarrow 4$ are suggested by the solution of the AP for customer node 3. However, the first match is infeasible because the flow of incoming arc $1 \rightarrow 3$ is less than the flow of outgoing arc $3 \rightarrow 5$. The two arcs are then deleted and replaced by arcs $1 \rightarrow 4$ and $4 \rightarrow 5$. After the arc replacement, the corresponding flow values are adjusted accordingly.

The procedure for tracing a set of feasible routes (a feasible solution) of the IRPSD starts from the customer node numbered as 1. For each customer examined, an assignment problem is solved to determine a match of its incoming and outgoing arcs. If the minimum objective value of the AP is zero, the procedure will proceed to the next customer node (the node numbered just after the current node). Otherwise, the solution of P' is adjusted based on the solution of the AP and in a way such that a feasible match of the incoming and outgoing arcs of the current node is obtained before the procedure proceeds to the next customer. The procedure terminates when all customer nodes have been examined, leading to a feasible solution (a set of feasible routes) of the IRPSD.

5.3. Local search improvement of the feasible solution

In the last subsection, a feasible solution of the IRPSD is constructed from the solution of P'. In most cases, the feasible solution is quite close to an optimal solution of the IRPSD. However, in some cases, the feasible solution can be further improved by local search. Two local search operators are used in our algorithm: one is to reduce the variable transportation cost by relocating the delivery of a customer served by a vehicle to another vehicle who serves the same customer; the other is to reduce the fixed transportation cost by merging two partially loaded routes (vehicles) into one route to reduce the number of routes.

5.4. Evaluation of solution quality

To evaluate the quality of the solution obtained in the last subsection, a lower bound for the objective value of the IRPSD is required. If such a lower bound is found, the gap between the upper bound provided by the solution and the lower bound, i.e., (the upper bound – the lower bound)/the upper bound × 100%, can be used to evaluate the quality of the solution.

To obtain the lower bound, we consider a model, denoted by P1, which is derived from P by adding additional constraints (15)–(17). It is obvious that any feasible solution of the IRPSD is also a feasible solution of model P1, and a lower bound of P1 is also a lower bound of the IRPSD. The lower bound of P1 can also be computed by using a Lagrangian relaxation approach (called LR2 below). However, this approach is different from the Lagrangian relaxation approach used for model P' (called LR1 below) in Section 4. In the approach LR2, the constraints (8), (16) and (17) of model P1 are relaxed by introducing non-negative Lagrange multipliers, leading to a decomposable relaxed problem. The relaxed problem can be optimally solved. The corresponding dual problem can then be solved by using the subgradient method, which provides a lower bound of model P1. If the gap between this lower bound and the upper bound is small, the solution of the IRPSD obtained in the last subsection is close to its optimal solution.

6. Numerical experiments

In this section, the performance of our proposed approach is evaluated by using randomly generated instances in different scenarios with different problem sizes, time horizons and other parameters. The parameters of the base scenario are generated in the following way: the total number of customers and the depot is taken as 100; the length of the time horizon is taken as $T = 5$, which corresponds to five working days in each week; $C, f_t, h_{it}, I_{i0}, V_i$ and r_{it} are randomly and uniformly generated from the intervals [100, 300], [400, 700], [0.5, 2], [50, 100], [400, 800], [50, 400] respectively; In order to make the problem feasible, M is uniformly generated from the interval $[1, 1.5] \times \max\{\sum_{t=1}^{\tau}\sum_{i=1}^{N} r_{it}/(\tau C)|\tau = 1,\ldots,T\}$. For c_{ij}, to ensure that the triangle inequality condition is satisfied, the coordinates of all customers and the central depot are first generated from a 10×10 square, and c_{ij} is then set as the physical distance between different customers i and j. c_{i0}^b is set to $10 \times c_{i0}$.

The algorithm is coded in C++ using the callable library of Lingo 6.0. A feasible solution of model P' is constructed based on the solution of its relaxed problem in each iteration. The final feasible solution of the IRPSD is obtained by transforming the best feasible solution of P' into a set of feasible routes. The numerical test was performed on a Pentium IV 1.73 GHz PC with 1 GB RAM. For each instance the termination conditions for LR2 and LR1 are 300 and 150 iterations, respectively. For LR2, M_i is set to 5. Ten instances are generated for the base scenario. For each instance model P' has 103,455 variables with 51,480 integral variables. With the notation in Table 1, the results of the instances are given in Table 2.

From Table 2, we can see: 1) the average gap between the upper bound and the lower bound is 6.71% with the largest gap 8.21%; 2) all instances are solved in a reasonable time, with the average computation time of

Table 1
Notation used in numerical results

CT	Computation time (minutes: seconds)
UB	Upper bound of the IRPSD found by LR1
LB	Lower bound found by LR2
Gap	Value of (UB − LB)/UB × 100%
N_0	Total number of customers and the depot

Table 2
The results of the instances of the base scenario

Instances	LB (10^5)	UB (10^5)	Gap (%)	CT
1	9.97	10.78	7.48	15:01
2	10.88	11.48	5.29	16:30
3	9.29	10.12	8.21	16:05
4	8.96	9.55	6.24	15:22
5	9.27	10.10	8.20	14:41
6	11.05	11.74	5.86	13:53
7	11.89	12.69	6.34	15:52
8	11.35	12.04	5.70	14:53
9	9.86	10.52	6.27	14:21
10	9.97	10.78	7.48	13:54
Average	10.25	10.98	6.71	15:03

15 minutes and 3 seconds. Note that some instances with only 10 customers could not be solved optimally by using LINGO 6.0 software after 24 hours of computation!

In order to evaluate the performance of our method for different scenarios, we also tested other five scenarios: the problem size is increased from $N_0 = 100$ to $N_0 = 200$; the time horizon is increased from 5 days to 10 days; the fixed vehicle cost per tour f_t is generated from the uniform distribution [800, 1600] instead of the uniform distribution [400, 700]; the holding cost h_{it} is generated from the uniform distribution [1, 4] instead of the uniform distribution [0.5, 2]; and the vehicle capacity is generated from the uniform distribution [50, 150] instead of the uniform distribution [100, 300]. For each scenario, 10 instances were randomly generated and tested. The results for these scenarios are given in Tables 3–7 respectively.

Some important observations can be obtained from the results:

1) From Tables 2–7 , the average gaps between the upper bound and the lower bound for all scenarios are less than 7% with the largest gap being 8.70% in Table 3 and the lowest gap being 5.15% in Table 6, which shows that our algorithm not only can obtain a near-optimal solution, but also is robust in the sense that its results are insensitive to the changes of the problem parameters.
2) All the instances can be solved in a reasonable time, with the average computation time of the base instances being only 15 minutes and 3 seconds. Even for larger instances tested with $N_0 = 200$ whose model P' contains 408,950 variables with 203,980 integral variables, our algorithm can obtain satisfactory solutions within 100 minutes of computation time on an ordinary computer. Moreover, with the increase of the problem size from $N_0 = 100$ to 200 or the time horizon from $T = 5$ to 10, the increase of the average computation time likely has an approximate linear relationship with the increase of the number of decision variables.

Table 3
The results of the instances with $N_0 = 200$

Instances	LB (10^5)	UB (10^5)	Gap (%)	CT
1	21.90	23.53	6.89	92:05
2	18.38	19.62	6.34	82:33
3	19.37	21.10	8.23	80:00
4	25.62	27.39	6.44	75:18
5	20.91	22.25	6.01	78:02
6	17.70	18.94	6.54	81:07
7	17.81	19.51	8.70	74:51
8	20.18	21.53	6.30	74:09
9	16.67	17.97	7.20	80:24
10	19.62	21.00	6.57	74:35
Average	19.82	18.88	6.92	79:19

Table 4
The results of the instances with time horizon of 10 days

Instance	LB (10^5)	UB (10^5)	Gap (%)	CT
1	18.70	20.10	6.97	40:15
2	25.50	27.14	6.04	45:17
3	21.32	23.16	7.93	39:12
4	18.82	20.12	6.48	39:53
5	21.38	22.79	6.18	38:43
6	18.32	19.62	6.62	38:56
7	19.91	21.23	6.21	42:23
8	23.79	25.42	6.43	43:55
9	22.10	23.79	7.11	42:27
10	22.83	24.26	5.90	46:51
Average	21.27	22.76	6.59	41:31

Table 5
The results of the instances with higher fixed cost

Instance	LB (10^5)	UB (10^5)	Gap (%)	CT
1	10.79	11.75	8.13	14:44
2	11.23	11.90	5.63	16:52
3	10.92	11.60	5.86	14:17
4	11.66	12.52	6.89	13:41
5	15.58	16.47	5.44	13:30
6	10.34	10.97	5.73	15:19
7	15.42	16.42	6.07	14:44
8	9.75	10.44	6.62	14:47
9	12.58	13.33	5.57	14:03
10	10.53	11.39	7.51	13:31
Average	11.88	12.68	6.35	14:33

Table 6
The results of the instances with higher holding cost

Instance	LB (10^5)	UB (10^5)	Gap (%)	CT
1	9.96	10.61	6.09	16:43
2	10.21	11.13	8.27	16:49
3	11.08	11.93	7.09	14:55
4	10.92	11.77	7.22	16:04
5	11.86	12.63	6.10	15:16
6	12.03	12.68	5.15	14:55
7	10.87	11.59	6.25	14:20
8	10.18	10.94	6.90	14:22
9	12.40	13.27	6.55	14:05
10	8.69	9.29	6.46	16:21
Average	10.82	11.58	6.61	15:76

Table 7
The results of the instances with lower vehicle capacity

Instance	LB (10^5)	UB (10^5)	Gap (%)	CT
1	13.03	13.88	6.15	14:22
2	12.21	12.98	5.95	15:43
3	13.25	14.09	5.94	14:51
4	12.96	13.85	6.44	16:01
5	9.64	10.30	6.36	15:34
6	8.99	9.56	5.98	15:03
7	8.94	9.50	5.88	15:51
8	12.75	13.65	6.55	14:46
9	10.38	10.99	5.58	14:06
10	12.96	13.70	5.40	14:40
Average	11.51	12.25	6.02	15:06

3) With the increase of the problem size from $N_0 = 100$ to 200, the average gap increases slightly. For example, the average gap increases 0.21% from 6.71% in Table 2 to 6.92% in Table 3.
4) With the increase of the time horizon from $T = 5$ to $T = 10$, the IRPSD may become more flexible to make a tradeoff between the transportation and inventory costs in multiple periods, the average gap thus goes down slightly from 6.71% in Table 2 to 6.59% in Table 4.
5) When the fixed vehicle cost per tour f_t is increased by being generated from the uniform distribution $[800, 1600]$ instead of the uniform distribution $[400, 700]$, the Lagrangian relaxation problem, especially the ROU sub-problem, may provide more useful information for the construction of a good feasible

solution of the IRPSD. As a result, the Lagrangian relaxation approach LR1 may provide a better upper bound for model P', the average gap is thus reduced by 0.36% from 6.71% to 6.35%.

6) The average gap decreases slightly from 6.71% in Table 2 to 6.61% in Table 6. It may be because with higher holding costs, the customers tend to hold fewer inventories and the cost savings due to the inventory reduction can offset a possible increase in the transportation costs.
7) With the decrease of the vehicle capacity, the number of direct deliveries tends to increase, which makes easier to find optimal routes in each period for the IRPSD. As a result, the average gap in Table 7 decreases to 6.02% from 6.71% of the base scenario.

7. Conclusion

The inventory routing problem with split delivery and vehicle fleet size constraint has been studied in this paper. In order to solve large scale problems, this paper has proposed an approximate model whose solution only defines in each period the quantity delivered to each customer, the quantity transported through each directed arc and the number of times that each directed arc is visited by vehicles in the corresponding transportation network. The model was solved by using a Lagrangian relaxation method combined with the surrogate subgradient method. A heuristic was used to construct a feasible solution of the model based on the solution of the Lagrangian relaxed problem. The model's solution, which may be infeasible for the original IRPSD, was then transformed into a feasible one by solving a series of assignment problems. The numerical experiments demonstrated that our proposed approach could obtain high quality solutions with the average relative gap with a lower bound less than 7%, for randomly generated large problems with 200 customers in a reasonable computation time on an ordinary personal computer. The approach is therefore promising for real applications.

References

Adelman, D., 2004. A price-directed approach to stochastic inventory/routing. Operations Research 52 (4), 499–514.

Aghezzaf, E.H., Raa, B., Van Landeghem, H., 2006. Modeling inventory routing problems in supply chains of high consumption products. European Journal of Operational Research 169 (3), 1048–1063.

Anily, S., Bramel, J., 2004. An asymptotic 98.5%-effective lower bound on fixed partition policies for the inventory-routing problem. Discrete Applied Mathematics 145 (1), 22–39.

Belenguer, J.M., Martinez, M.C., Mota, E., 2000. A lower bound for the split delivery vehicle routing problem. Operations Research 48 (5), 801–810.

Campbell, A., Clarke, L., Savelsbergh, M.W.P., 2002. Inventory routing in practice: The vehicle routing problem. In: P. Toth, D. Viego (Eds)., SIAM Monographs on Discrete Mathematics and Applications, 2002. pp. 309–330.

Campbell, A.M., Savelsbergh, M.W.P., 2004. A decomposition approach for the inventory-routing problem. Transportation Science 38 (4), 488–502.

Chandra, P., Fisher, M.L., 1994. Coordination of production and distribution planning. European Journal of Operational Research 72, 503–517.

Dell'Amico, M., Lodi, A., Martello, S., 2001. Efficient algorithms and codes for k-cardinality assignment problems. Discrete Applied Mathematics 110 (1), 25–40.

Dror, M., Ball, M., 1987. Inventory/routing: Reduction from an annual to a short period problem. Naval Research Logistics Quarterly 34 (6), 891–905.

Dror, M., Ball, M., Golden, B., 1985. A computational comparison of algorithms for the inventory routing problem. Annals of Operation Research 4, 3–23.

Dror, M., Trudeau, P., 1989. Savings by split delivery routing. Transportation Science 23, 141–145.

Dror, M., Trudeau, P., 1990. Split delivery routing. Naval Research Logistics 37, 383–402.

Federgruen, A., Prastacos, G., Zipkin, P., 1986. An allocation and distribution model for perishable products. Operations Research 34, 75–82.

Federgruen, A., Zipkin, P., 1984. A combined vehicle routing and inventory allocation problem. Operation Research 32, 1019–1037.

Fumero, F., Vercellis, C., 1999. Synchronized development of production, inventory, and distribution schedules. Transportation Science 33 (3), 330–340.

Ho, S.C., Haugland, D., 2004. A tabu search heuristic for the vehicle routing problem with time windows and split deliveries. Computers & Operations Research 31 (12), 1947–1964.

Jaillet, P., Bard, J.F., Huang, L., Dror, M., 2002. Delivery cost approximations for inventory routing problems in a rolling horizon framework. Transportation Science 36 (3), 292–300.

Kleywegt, A.J., Nori, V.S., Savelsbergh, M.W.P., 2002. The Stochastic inventory routing problem with direct deliveries. Transportation Science 36 (1), 94–118.

Kleywegt, A.J., Nori, V.S., Savelsbergh, M.W.P., 2004. Dynamic programming approximations for a stochastic inventory routing problem. Transportation Science 38 (1), 42–70.

Kumar, A., Schwarz, L., Ward, J., 1995. Risk-pooling along a fixed delivery route using a dynamic inventory-allocation policy. Management Science 41, 344–362.

Lee, C.G., Epelman, M.A., White, C.C., Bozer, Y.A., 2006. A shortest path approach to the multiple-vehicle routing problem with split pick-ups. Transportation Research Part B-Methodological 40 (4), 265–284.

Martello, S., Toth, P., 1987. Linear assignment problems. In: Martello, S., Laporte, G., Minoux, M., Ribeiro, C. (Eds.), Surveys in Combinatorial Optimization, Annals of Discrete Mathematics, vol. 31. North-Holland, Amsterdam, pp. 259–282.

Reiman, M.I., 1999. Heavy traffic analysis of the dynamic stochastic inventory-routing problem. Transportation Science 33 (4), 361–380.

Schwarz, L.B., Ward, J., Zhai, X., 2004. Joint Transportation-and-Inventory Problems in Supply Chains: A Review. Krannert School of Management, Purdue University.

Trudeau, P., Dror, M., 1992. Stochastic inventory routing: route design with stockouts and route failures. Transportation Science 26 (3), 171–184.

Wolsey, L.A., 1998. Integer Programming. Wiley, New York, pp. 38–40.

Yu, Y.G., Chen, H.X., Chu, F., 2005. Large scale inventory routing problem with split delivery: A new model and Lagrangian relaxation approach. In: 2005 IEEE International Conference on Service Operations and Logistics, and Informatics, Beijing, China.

Yu, Y.G., Chu, F., Chen, H.X., 2007. A note on coordination of production and distribution planning. European Journal of Operational Research 177 (1), 626–629.

Zhao, Q.H., Wang, S.Y., Lai, K.K., 2007. A partition approach to the inventory/routing problem. European Journal of Operational Research 177 (2), 786–802.

Zhao, X., Luh, P.B., Wang, J., 1999. Surrogate gradient algorithm for Lagrangian relaxation. Journal of Optimization Theory and Applications 100 (3), 699–712.