OPERATIONS RESEARCH
Vol. 52, No. 3, May–June 2004, pp. 396–408
ISSN 0030-364X | EISSN 1526-5463 | 04 | 5203 | 0396

DOI 10.1287/opre.1030.0096

Warehouse-Retailer Network Design Problem

Chung-Piaw Teo, Jia Shu
Department of Decision Sciences, National University of Singapore, and Singapore-MIT Alliance Program
{bizteocp@nus.edu.sg, tlisj@nus.edu.sg}

In this paper, we study the distribution network design problem integrating transportation and infinite horizon multiechelon inventory cost function. We consider the trade-off between inventory cost, direct shipment cost, and facility location cost in such a system. The problem is to determine how many warehouses to set up, where to locate them, how to serve the retailers using these warehouses, and to determine the optimal inventory policies for the warehouses and retailers. The objective is to minimize the total multiechelon inventory, transportation, and facility location costs. To the best of our knowledge, none of the papers in the area of distribution network design has explicitly addressed the issues of the 2-echelon inventory cost function arising from coordination of replenishment activities between the warehouses and the retailers. We structure this problem as a set-partitioning integer-programming model and solve it using column generation. The pricing subproblem that arises from the column generation algorithm gives rise to a new class of the submodular function minimization problem. We show that this pricing subproblem can be solved in $O(n \log n)$ time, where n is the number of retailers. Computational results show that the moderate size distribution network design problem can be solved efficiently via this approach.

Subject classifications: transportation; models, programming: integer; facilities/equipment planning: location.
Area of review: Transportation.
History: Received March 2002; revision received January 2003; accepted March 2003.

1. Introduction

In today's competitive market, a company's distribution network must meet service goals at the lowest possible cost. In some instances, a company may be able to save millions of dollars in logistics costs and simultaneously improve service levels by redesigning its distribution network. To achieve this, an ideal network must have the optimum number, size, and location of warehouses to support the inventory replenishment activities of its retailers.

Shenzhen ST-Anda Logistics Co. Ltd. (ST Anda), a joint venture of Sembcorp Logistics (previously called Singapore Technologies Logistics) and China Merchant Holdings, is a third-party logistics provider equipped with operating licenses to provide services in warehousing, general haulage, distribution, and IT management throughout China. It is one of the largest transportation companies in southern China and is one of the few logistics providers to operate a nationwide logistics network. ST-Anda is able to deliver goods to its customers in over 270 cities in China within four calendar days. More than 50% of the goods can be delivered within two days.

To help its clients, especially multinational corporations, to expand in China, ST-Anda has to constantly seek ways to improve its service level and its serviceability reach. One major challenge confronting the company these days is: Given the client's demand forecasts (at product level) over the entire country, the location of the production facilities, and the client's retail outlets, ST-Anda would like to add value to its clients by taking over the task of distributing the goods to the retailer outlets. This must be achieved at a competitive price, with a satisfactory service level, and within a defined delivery time window.

To solve the above problem and design the best distribution network, one must consider all relevant costs and service-level constraints. Relevant costs include inbound and outbound transportation, fixed and variable warehouse costs, inventory carrying, and producing or sourcing from different locations. Complex trade-offs make these costs difficult to analyze. For example, as the number of warehouse locations increases, transportation cost will tend to decrease, but inventory cost will tend to increase. Moreover, costs are often dependent on the location and capacity of plants or vendors as well as the location and demand characteristics of customers. In many distribution systems, items are kept at warehouses for intermediate storage and subsequently shipped to retail outlets. Consequently, efficient distribution strategies that reduce total cost must take into account the interactions of the various replenishment activities in the distribution chain. This can be achieved by determining simultaneously (i) the timing and sizes of retailer deliveries, and (ii) replenishment to the warehouses from external suppliers so as to minimize total transportation and inventory costs.

Motivated by these issues, we propose in this paper an integrated model for the optimal distribution network design problem, taking into account the trade-off between transportation cost, warehouse operating cost, and more importantly, warehouse-retailers echelon inventory replenishment cost.

1.1. Inventory Cost

From an inventory modeling viewpoint, the one warehouse multiretailer (OWMR) system with deterministic customer demand has been studied extensively since the breakthrough work of Roundy (1985) (cf., Muckstadt and Roundy 1993). In the OWMR system, customer demand occurs at each retailer at a constant rate. This demand must be met as it occurs over an infinite horizon without shortages or backlogging. Orders placed by retailers generate demands at the warehouse, which acts as the source of supply for the retailers. The warehouse replenishes its inventory from an external supplier. There is a holding cost charged against each unit of inventory per unit time at the retailers and the warehouse and a corresponding set-up cost charged for each order placed at the warehouse and at each retailer. The demand rates, holding cost rates, and set-up costs are stationary and facility dependent. Delivery of orders is assumed to be instantaneous. Note that the cost of operating the warehouse now depends on the ordering patterns of the retailers. The objective is to find the inventory replenishment policies for the warehouse and the retailers that minimize the systemwide inventory cost.

1.2. Inventory and Transportation Cost

Initial work to integrate inventory with transportation cost has focused on numerical experiments with a variety of strategies for the multiperiod inventory-routing problem; see Golden et al. (1984), Dror and Levy (1986), Dror and Ball (1987), and Chandra and Fisher (1994). Subsequent work considers restrictions to other classes of strategies. Gallego and Simchi-Levi (1990) show that direct shipping policies, i.e., policies in which each vehicle visits a single retailer, are within 6% of optimality under certain restricted parameter settings. Herer and Roundy (1997) and Viswanathan and Mathur (1997) show good empirical performance for the so-called power-of-two strategies under which each retailer is replenished at constant intervals, which are power-of-two multiples of a common base planning period. The recent work of Chan et al. (2002) analyzes the effectiveness of a large class of policies, called zero inventory ordering (ZIO) policies, for the single warehouse multiretailer system. In this class, a retailer receives an order when its inventory level is down to zero. This analysis is motivated by the observation that direct shipping, power-of-two policies, etc., are both subsets of the class of ZIO policies.

1.3. Distribution Network Design

Much of the work described above assumed a *given* distribution network structure, where retailer-warehouse assignment has been predetermined. On the other hand, distribution network design problems seek to address the strategic issues of where to site the warehouses and of which warehouse will be used to serve the respective retailers. While many nonquantifiable factors can usually sway the network design decision, a common approach to selecting the optimal network structure is to examine the effect of the network structure on aggregate cost structure comprising inventory-, transportation-, and location-related cost components. Because of the complex nature of the problem, network design problems are usually posed as mixed-integer programming (MIP) models with binary choice variables for fixed charges related to site choices and fixed ordering cost, and continuous variables for the flow of goods. This line of research began to appear in the operations research literature, e.g., in Baumol (1958), which describes a heuristic for a nonlinear warehouse location model. About two decades ago, the structural design of distribution systems using an optimizing approach became technically feasible. Geoffrion and Powers (1995) provide perspectives on algorithmic and associated evolutionary developments in this area, stating that "it is now possible for companies of all sizes to find distribution system designs that are optimal for all practical purposes even, in many cases, when the scope of the design problem is extended to the complete logistics system in the broadest contemporary sense."

Most of the early works on the distribution network design system focus mainly on the finite horizon problem with deterministic demand rate. The work of Chan and Simchi-Levi (1998) is a notable exception. Under mild probabilistic assumption on retailers demand rates and locations, they use an ingenious probabilistic analysis to identify the underlying structure of asymptotically optimal policies for general distribution systems. In particular, they show that there exists an asymptotically optimal policy that satisfies the following properties:

- Each vehicle from an outside vendor ships in full load to a single warehouse, i.e., direct shipment from the outside vendor.
- Each retailer is served by a single warehouse.
- The warehouses act mainly as a cross-dock facility, i.e., no inventory is held at the warehouses.

Note that these qualitative insights on the effective policies apply in the case when the number of retailers tend to infinity.

In the more recent work of Erlebacher and Meller (2000), an analytical model was developed to minimize the sum of fixed operating cost and inventory holding costs incurred by the warehouses, together with the transportation costs between manufacturers and warehouses and between warehouses and retailers. They used heuristic procedures because the model is NP-hard.

In this paper, we extend the MIP methodology to consider an infinite horizon distribution network design problem. We assume that:

- The set of potential warehouse locations is denoted by $\mathscr{W}$.
- The set of retailers is denoted by $\mathscr{R}$. (They are geographically dispersed in the given region.)

- There is an outside vendor that can be used to replenish the warehouses via direct shipment.
- Each retailer can only be served by exactly one warehouse. Furthermore, the warehouse serves the retailers assigned to it via direct shipment.
- Each warehouse w in $\mathcal{W}$ incurs an ordering cost K_w every time it places an order to the outside vendor. The ordering cost is independent of the order quantity. The transportation cost (from outside vendor to the warehouse) incurred is d_w per unit ordered. The inventory holding cost rate in warehouse w is denoted by h_w per unit per year.
- Similarly, each retailer i in $\mathcal{R}$ incurs an ordering cost K_i every time it places an order to the warehouse assigned to serve it. The ordering cost is independent of the order quantity. The transportation cost (from warehouse w to retailer i) incurred is $d_{w,i}$ per unit ordered. The inventory holding cost rate in retailer i is denoted by h_i per unit per year. Each retailer i also faces a constant demand rate λ_i per year.
- The facility cost of operating warehouse w is F_w per year.

Note that the model is flexible enough to incorporate service constraint of the type that the transportation lead time from the warehouse to the retailer cannot be more than, say, two days. In this instance, we merely set $d_{w,i}$ to be a huge cost penalty whenever warehouse w cannot meet the service standard for retailer i.

We would like to determine the optimal number of warehouses to set up, assign the retailers to the warehouses, and determine the optimal inventory replenishment strategies for the warehouses and retailers. The goal is to minimize the total systemwide inventory-, transportation-, and facility-related cost. We call this the *warehouse-retailer distribution network design problem* (WRND).

The main contributions of this paper are as follows:

- We show that the warehouse-retailer distribution network design problem can be modeled approximately (to within 98% accuracy) as a set-partitioning model that can be efficiently solved using the column generation method.
- The pricing subproblem gives rise to an interesting class of the submodular function minimization problem. We show that this problem can be solved in $O(|\mathcal{R}|\log(|\mathcal{R}|))$ time, where $\mathcal{R}$ denotes the set of retailers.
- In the instance when direct shipment cost between the warehouses and retailers can be treated as a constant or can be ignored, our method reduces to a 1.02 approximation algorithm for the warehouse-retailer distribution network design problem. This extends Roundy's (1985) classical result to the distribution network design domain.
- We present extensive computational results to show that our proposed approach to the warehouse-retailer distribution network design problem is very efficient in practice. For instance, for a system with 20 potential warehouse locations and 100 retailers (i.e., 20×2^{100} variables, 100 constraints), our method can be used to churn out the optimal design in under 15 minutes.

2. Review of the One Warehouse Multiretailer (OWMR) System

In the warehouse-retailer distribution network design problem, one of the major challenges is to obtain the cost expression for the systemwide inventory replenishment cost. In fact, characterizing the optimal inventory replenishment policy for the OWMR system is still an unsolved problem to this day. This arises in part because the optimal replenishment policy need not be static and may depend on the inventory levels across all sites in the system. Fortunately, in a seminal work, Roundy (1985) showed that the problem can be approximated to within 98% accuracy (to be made precise later) using a simple convex programming model. Before we present the model for the distribution network design problem, it is thus useful to first review Roundy's results for the OWMR system.

Given a *fixed* warehouse and a set of retailers assigned to it, let the holding cost rate and the fixed ordering cost at the warehouse be denoted by h_0 and K_0. For retailer i, $i = 1, \ldots, N$, the corresponding holding and ordering cost parameters are given by h_i and K_i. The demand rate at retailer i is denoted by λ_i. For ease of exposition, we may assume that $h_i \geqslant h_0$ for every $i = 1, \ldots, N$. In fact, the extension of this problem to allow $h_i - h_0$ negative is not much harder.

A *stationary* inventory control policy for the system can be characterized by an $(N+1)$-tuple, $(T_0, T_1, \ldots, T_N)$, where T_0 is the reorder interval at the warehouse and T_i is that at retailer i, $i = 1, \ldots, N$. By focusing on stationary policies that satisfy an additional constraint (the well-known integer-ratio property), Roundy (1985) derived a convex programming relaxation to the OWMR problem. Furthermore, by a clever rounding argument (see also Teo and Bertsimas 2001 for a different rounding argument), he showed that there exists a power-of-two policy that is close to 98% of the value of the convex programming lower bound. This shows that the convex programming relaxation approximates the optimal solution value to 98% accuracy.

The main results of Roundy (1985) (see also the review article by Muckstadt and Roundy 1993) can be summarized in Proposition 1.

Proposition 1. (i) *The solution to the following convex optimization problem*

$$\min_{T_0, T_i: T_i>0, i=1,\ldots,N} \left(\frac{K_0}{T_0} + \sum_i \frac{K_i}{T_i} + \frac{1}{2}\sum_i \lambda_i h_i T_i + \frac{1}{2}\sum_i \lambda_i h_0 [\max(T_0, T_i) - T_i] \right), \tag{1}$$

is a lower bound on average cost of any feasible inventory control policy (*possibly dynamic*) *and the solution can be rounded off to obtain a feasible integer-ratio policy* (*i.e.,* T_0 *divides* T_i *or vice versa, for every* i) *with a cost within* 98% *of the minimum of* (1).

(ii) *In the solution to* (1), *the retailers can be divided into three groups*: *G, L, and E. For the retailers in G, their reorder interval T_i is given by*

$$T_i = \sqrt{K_i/\tfrac{1}{2}\lambda_i h_i} > T_0.$$

For the retailers in L, their reorder interval is given by

$$T_i = \sqrt{K_i/\tfrac{1}{2}\lambda_i(h_i - h_0)} < T_0.$$

Finally, for the retailers in E, their reorder interval is the same as that at the warehouse and is given by

$$T_i = T_0 = \sqrt{\big[K_0 + \textstyle\sum_{i\in E} K_i\big]\big/\big[\tfrac{1}{2}\sum_{i\in E}\lambda_i h_i + \tfrac{1}{2}\sum_{i\in L}\lambda_i h_0\big]}.$$

Furthermore,

$$\sqrt{K_i/\tfrac{1}{2}\lambda_i(h_i - h_0)} \geqslant T_0 \geqslant \sqrt{K_i/\tfrac{1}{2}\lambda_i h_i} \quad \textit{for all } i \in E.$$

2.1. Remarks

- The results in Proposition 1(ii) follow from applying the Karush-Kuhn-Tucker conditions on (1). We refer the readers to the original article by Roundy (1985) for details.
- Our cost function is slightly different from what Roundy obtained as he uses the notion of *echelon* holding cost (defined to be $H_i = h_i - h_0$ using our notation) in his expression. To recover Roundy's cost function, the readers need to perform an algebraic manipulation to regroup variables into the terms T_i and $\max(T_0, T_i)$ and replace $h_i - h_0$ by the echelon holding cost H_i.
- Let Z^* denote the optimal inventory replenishment cost for the OWMR system. Let Z_B denote the bound obtained by solving the relaxation defined in (1). Roundy (1985) proved that $Z_B \leqslant Z^* \leqslant 1.02 Z_B$. In the rest of this paper, we will use Z_B defined in (1) to approximate the optimum multiechelon inventory cost Z^*, with the understanding that it approximates the actual cost function to 98% accuracy. By solving the optimal distribution network design problem using this cost approximation, the network obtained will be guaranteed to be within 98% optimality of the actual distribution network design problem.

3. Terminology and the Set-Partitioning Model

The distribution network design problem can be viewed as an assignment problem with very complex assignment cost function. Our goal is to obtain the optimal cost configuration by assigning retailers to the warehouses. Given a fixed assignment, the assignment cost is obtained by solving a series of OWMR problems (one for each warehouse and its assigned retailers). In this section, we formulate our decision problem as a set-partitioning model, and we present an approach to solve this model.

Recall that $\mathscr{W}$ is the set of all potential warehouse locations. Let w be a particular warehouse in $\mathscr{W}$ and S be a subset of retailers in $\mathscr{R}$. Let

$$x_{w,S} = \begin{cases} 1 & \text{if } w \text{ is used to serve retailers in } S \text{ and no one else,} \\ 0 & \text{otherwise.} \end{cases}$$

The (WRND) problem reduces to one of finding a minimum cost partition of the set of retailers into $(S_1, \ldots, S_k)$, and the corresponding warehouse assignment $(w_1, \ldots, w_k)$.

To describe the cost components, let $c_{w,S}$ denote the cost of serving retailers in S using the warehouse w. It comprises the following components:

- Systemwide inventory replenishment cost (denoted by $I(w, S)$), approximated by expression (1), with K_w, h_w, and T_w replacing the role of K_0, h_0, and T_0, respectively, in the cost function; i.e.,

$$I(w,S) \equiv \min_{T_w,\, T_i:\, T_i>0,\, i\in S} \left(\frac{K_w}{T_w} + \sum_{i\in S}\frac{K_i}{T_i} + \frac{1}{2}\sum_{i\in S}\lambda_i h_i T_i + \frac{1}{2}\sum_{i\in S}\lambda_i h_w[\max(T_w, T_i) - T_i] \right). \quad (2)$$

- Facility location fixed cost F_w.
- Annual total transportation cost $\sum_{i\in S}(\lambda_i d_w + \lambda_i d_{w,i})$. The first term $\sum_{i\in S}\lambda_i d_w$ refers to the annual transportation cost from outside vendor to warehouse w, whereas $\lambda_i d_{w,i}$ refers to the annual transportation cost incurred for shipment between warehouse w and the retailer i. For the sake of brevity, we let $v_{w,i} \equiv \lambda_i(d_w + d_{w,i})$. Hence, the total transportation cost can be denoted by $\sum_{i\in S} v_{w,i}$.

The warehouse-retailer distribution network design problem can be modeled as a set-partitioning problem in the following way:

$$\text{(WRND)} \quad \min \sum_{w:\, w\in\mathscr{W}} \sum_{S:\, S\subseteq\mathscr{R}} c_{w,S}\, x_{w,S}$$

subject to

$$\sum_{w} \sum_{S:\, S\subseteq\mathscr{R},\, i\in S} x_{w,S} = 1 \quad \forall i \in \mathscr{R},$$

$$x_{w,S} \in \{0, 1\}.$$

Our set-partitioning model looks nice at first because there is no nonlinear component in either its objective function or constraints. However, due to the large number of variables ($O(|\mathscr{W}| \times 2^{|\mathscr{R}|})$) and the fact that each coefficient $c_{w,S}$ can be obtained only by solving a related convex programming problem, it is impossible to solve it using standard MIP methodology to optimality in an efficient manner. We solve instead the corresponding LP relaxation using simplex and the column generation method. To facilitate an efficient implementation, given any set of dual prices, we need a method to quickly find a column with negative reduced cost or to verify that none exists. This is known as the *pricing subproblem*.

4. The Pricing Subproblem

Let $n = |\mathcal{R}|$ and $(u_1, u_2, \ldots, u_n)$ be the dual solution obtained in one of the phases of the column generation algorithm to the linear relaxation of the above set-partitioning model. For each column (w, S), we want to know whether the reduced cost

$$\left(c_{w,S} - \sum_{i \in S} u_i\right) = I(w, S) + F_w + \sum_{i \in S} v_{w,i} - \sum_{i \in S} u_i$$

is nonnegative. Fixing w, this is equivalent to checking whether

$$\min_{S \subseteq \mathcal{R}} \left(I(w, S) + F_w + \sum_{i \in S} v_{w,i} - \sum_{i \in S} u_i \right) < 0.$$

Define a set function f on $2^{\mathcal{R}}$ with

$$f(S) \equiv \left(I(w, S) + \sum_{i \in S} v_{w,i} - \sum_{i \in S} u_i \right)$$
$$= I(w, S) - \sum_{i \in S} (u_i - v_{w,i}).$$

Note that the condition

$$\min_{S \subseteq \mathcal{R}} \left(I(w, S) + F_w + \sum_{i \in S} v_{w,i} - \sum_{i \in S} u_i \right) < 0$$

is equivalent to

$$\min_{S \subseteq \mathcal{R}} f(S) < -F_w.$$

The key in solving the pricing subproblem is to first solve the problem of $\min_{S \subseteq \mathcal{R}} f(S)$.

Let $P_i = u_i - v_{w,i}$. Then,

$$f(S) = \min_{T_w, T_i, i \in S} \frac{K_w}{T_w} + \sum_{i \in S} \frac{K_i}{T_i} + \frac{1}{2} \sum_{i \in S} \lambda_i h_i T_i + \frac{1}{2} \sum_{i \in S} \lambda_i h_w [\max(T_w, T_i) - T_i] - \sum_{i \in S} P_i$$
$$= \min_{T_w, T_i, i \in S} \frac{K_w}{T_w} + \sum_{i \in S} \left(\frac{K_i}{T_i} + \frac{1}{2} \lambda_i h_i T_i \right) + \sum_{i \in S} \frac{1}{2} \lambda_i h_w [\max(T_w, T_i) - T_i] - \sum_{i \in S} P_i.$$

Given a finite set E, a real-valued function $h(\cdot)$ that is defined on the subsets of E is called *submodular* if, for every pair $S, T \subseteq E$, we have that

$$h(S) + h(T) \geqslant h(S \cap T) + h(S \cup T).$$

Theorem 1. *$f(S)$ that arises from the pricing subproblem is a submodular function.*

Proof. Because $f(S) = I(w, S) - \sum_{i \in S} P_i - F_w$, it suffices to show that $I(w, S)$ is submodular.

Let S and T be 2 nonempty subsets of retailers. Let $(T_w^*(S), T_i^*(S) \colon i \in S)$ denote the optimal reorder intervals in $I(w, S)$. Similarly, let $(T_w^*(T), T_i^*(T) \colon i \in T)$ denote the optimal reorder intervals in $I(w, T)$,

$$I(w, S) = \frac{K_w}{T_w^*(S)} + \sum_{i \in S} \frac{K_i}{T_i^*(S)} + \frac{1}{2} \sum_{i \in S} \lambda_i h_i T_i^*(S) + \frac{1}{2} \sum_{i \in S} \lambda_i h_w [\max(T_w^*(S), T_i^*(S)) - T_i^*(S)],$$

and

$$I(w, T) = \frac{K_w}{T_w^*(T)} + \sum_{i \in T} \frac{K_i}{T_i^*(T)} + \frac{1}{2} \sum_{i \in T} \lambda_i h_i T_i^*(T) + \frac{1}{2} \sum_{i \in T} \lambda_i h_w [\max(T_w^*(T), T_i^*(T)) - T_i^*(T)].$$

Without loss of generality, we may assume that

$$T_w^*(S) \geqslant T_w^*(T).$$

Consider a new system with warehouse w serving the retailers in $S \cup T$. The optimal inventory replenishment cost is $I(w, S \cup T)$. Similarly, consider another system with warehouse w serving the retailers in $S \cap T$ (possibly empty). The optimal inventory replenishment cost in this case is $I(w, S \cap T)$. For the new systems, we construct the new reorder intervals in the following way:

$$T_w(S \cup T) = T_w^*(T),$$
$$T_w(S \cap T) = T_w^*(S),$$
$$T_i(S \cup T) = \begin{cases} T_i^*(S) & \text{if } i \in S \backslash T, \\ T_i^*(T) & \text{if } i \in T, \end{cases}$$
$$T_i(S \cap T) = T_i^*(S).$$

Note that $(T_w(S \cup T), T_i(S \cup T), i \in S \cup T)$ gives rise to a feasible solution to the inventory replenishment problem with w serving the set of retailers in $S \cup T$. Similarly, $(T_w(S \cap T), T_i(S \cap T), i \in S \cap T)$ gives rise to a feasible solution to the inventory replenishment problem with w serving the set of retailers in $S \cap T$. Let $I'(w, S \cup T)$, $I'(w, S \cap T)$ denote the corresponding feasible inventory replenishment costs for the two new systems, respectively.

Note that

$$I(w, S \cup T) \leqslant I'(w, S \cup T).$$

Furthermore,

$$I'(w, S \cup T) \equiv \frac{K_w}{T_w(S \cup T)} + \sum_{i \in S \cup T} \frac{K_i}{T_i(S \cup T)} + \frac{1}{2} \sum_{i \in S \cup T} \lambda_i h_i T_i(S \cup T) + \frac{1}{2} \sum_{i \in S \cup T} \lambda_i h_w [\max(T_w(S \cup T), T_i(S \cup T)) - T_i(S \cup T)]$$

$$= \frac{K_w}{T_w^*(T)} + \sum_{i\in S\setminus T} \frac{K_i}{T_i^*(S)} + \frac{1}{2}\sum_{i\in S\setminus T} \lambda_i h_i T_i^*(S)$$
$$+ \frac{1}{2}\sum_{i\in S\setminus T} \lambda_i h_w[\max(T_w^*(T), T_i^*(S)) - T_i^*(S)]$$
$$+ \sum_{i\in T} \frac{K_i}{T_i^*(T)} + \frac{1}{2}\sum_{i\in T} \lambda_i h_i T_i^*(T)$$
$$+ \frac{1}{2}\sum_{i\in T} \lambda_i h_w[\max(T_w^*(T), T_i^*(T)) - T_i^*(T)].$$

Similarly,

$$I(w, S\cap T) \leqslant I'(w, S\cap T),$$

where

$$I'(w, S\cap T)$$
$$\equiv \frac{K_w}{T_w(S\cap T)} + \sum_{i\in S\cap T} \frac{K_i}{T_i(S\cap T)} + \frac{1}{2}\sum_{i\in S\cap T} \lambda_i h_i T_i(S\cap T)$$
$$+ \frac{1}{2}\sum_{i\in S\cap T} \lambda_i h_w[\max(T_w(S\cap T), T_i(S\cap T)) - T_i(S\cap T)]$$
$$= \frac{K_w}{T_w^*(S)} + \sum_{i\in S\cap T} \frac{K_i}{T_i^*(S)} + \frac{1}{2}\sum_{i\in S\cap T} \lambda_i h_i T_i^*(S)$$
$$+ \frac{1}{2}\sum_{i\in S\cap T} \lambda_i h_w[\max(T_w^*(S), T_i^*(S)) - T_i^*(S)].$$

Summing the above expressions, we obtain

$$I(w, S) + I(w, T) - I'(w, S\cup T) - I'(w, S\cap T)$$
$$= \frac{1}{2}\sum_{i\in S\setminus T} \lambda_i h_w[\max(T_w^*(S), T_i^*(S)) - T_i^*(S)]$$
$$- \frac{1}{2}\sum_{i\in S\setminus T} \lambda_i h_w[\max(T_w^*(T), T_i^*(S)) - T_i^*(S)]$$
$$\geqslant 0.$$

The last inequality follows from our assumption that $T_w^*(S) \geqslant T_w^*(T)$, so we get the desired result:

$$I(w, S\cup T) + I(w, S\cap T) \leqslant I'(w, S\cup T) + I'(w, S\cap T) \leqslant I(w, S) + I(w, T). \quad \square$$

The pricing subproblem in the column generation phase is thus a submodular function minimization problem. Minimizing a rational submodular function was shown by Grotschel et al. (1993) to be solvable in polynomial time via the ellipsoid algorithm. More recently, Iwata et al. (2001) and, independently, Schrijver (1999) have developed strongly polynomial combinatorial algorithms for the submodular function minimization problem. Thus, the result we have presented implies that our pricing subproblem is polynomially solvable. In the rest of this section, we present a much faster algorithm to solve the pricing subproblem.

Let S' be the optimal solution to the pricing subproblem. The basic idea for our algorithm hinges on the insight that if we can *guess* the value of the optimal $T_w^*(S')$ (abbreviated by T_w^*) correctly, this alone is enough for us to recover the solution S'. Note that for a fixed T_w^*, the expression in (1) decomposes into a separable function in terms of the T_is. Let L_w, E_w, and G_w be the three groups of retailers obtained in Proposition 1(ii), with w as the warehouse.

Lemma 1. *For all $i\in L_w$, $i\in S'$ if and only if*

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w^* - P_i < 0. \tag{3}$$

Proof. For any $i\in L_w$, the inventory replenishment cost incurred due to the presence of i is given by the expression

$$\frac{K_i}{T_i^*} + \frac{1}{2}\lambda_i h_i T_i^* + \frac{1}{2}\lambda_i h_w (T_w^* - T_i^*),$$

where

$$T_i^* = \sqrt{K_i/\tfrac{1}{2}\lambda_i(h_i - h_w)}.$$

Hence, i will be in the optimal subset S' if and only if

$$\frac{K_i}{T_i^*} + \frac{1}{2}\lambda_i h_i T_i^* + \frac{1}{2}\lambda_i h_w (T_w^* - T_i^*) - P_i < 0,$$

i.e.,

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w^* - P_i < 0. \quad \square$$

Similarly, we can get the following two lemmas.

Lemma 2. *For all $i\in E_w$, $i\in S'$ if and only if*

$$\frac{K_i}{T_w^*} + \frac{1}{2}\lambda_i h_i T_w^* - P_i < 0. \tag{4}$$

Lemma 3. *For all $i\in G_w$, $i\in S'$ if and only if*

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i h_i} - P_i < 0. \tag{5}$$

To summarize, given a fixed T_w^*, we can:

1. partition all the retailers in $\mathcal{R}$ into three groups L_w, E_w, and G_w by comparing the values of T_w^* with $\sqrt{K_i/(1/2)\lambda_i(h_i - h_w)}$ and $\sqrt{K_i/(1/2)\lambda_i h_i}$ (cf., Proposition 1);
2. use the value of P_i and membership in the sets L_w, E_w, and G_w to decide whether to include retailer i in S' or not, using Lemmas 1, 2, and 3; and
3. verify that our original guess of the value T_w^* is correct by checking the conditions in Proposition 1 and whether $T_i^* = T_w^*$ for $i\in E_w\cap S'$ is indeed true.

Unfortunately, it will not be possible to repeat the above procedure for all values of T_w^*. In our algorithm, we first partition the real line into a collection of small intervals. The interval $[a, b]$ in the collection is constructed in the following manner:

• The open interval (a, b) does not contain any of the points $\sqrt{K_i/(1/2)\lambda_i(h_i - h_w)}$ and $\sqrt{K_i/(1/2)\lambda_i h_i}$ for $i = 1, \ldots, |\mathscr{R}|$.

• The open interval (a, b) does not contain the roots to the equations

$$\frac{K_i}{T_w} + \frac{1}{2}\lambda_i h_i T_w - P_i = 0 \tag{6}$$

and

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w - P_i = 0. \tag{7}$$

Any interval that satisfies the above conditions is said to be a *nice* interval.

Note that if our guess is that T_w^* lies in a nice interval $[a, b]$, then we can carry out the steps outlined in the above algorithm (for fixed T_w^* case) without any difficulty because we can check whether the inequalities in Proposition 1 and Lemmas 1, 2, and 3 hold.

More formally, let x_i, y_i ($x_i \leqslant y_i$) denote the real roots to Equation (6) and l_i the solution to Equation (7), if they exist.

Algorithm TS

Step 1. If we partition the real line by $\sqrt{K_i/(1/2)\lambda_i h_i}$, $\sqrt{K_i/(1/2)\lambda_i(h_i - h_w)}$, l_i, x_i, and y_i for all $i \in \mathscr{R}$, we get at most $5n + 1$ intervals along the line. Note that each interval obtained this way is a *nice* interval, i.e., as long as T_w^* falls within that interval, we have enough information to determine for all $i \in \mathscr{R}$, whether $i \in L_w, E_w, G_w$, and whether or not i will be included in the optimal set.

Step 2. Suppose T_w^* falls in a particular nice interval, say $[a, b]$ (choosing the nice intervals from left to right). Obtain the sets L_w, E_w, and G_w, depending on whether $\sqrt{K_i/(1/2)\lambda_i h_i}$, $\sqrt{K_i/(1/2)\lambda_i(h_i - h_w)}$ falls to the left or right of the interval $[a, b]$ (cf., Proposition 1). Note that by construction of the intervals, none of the values in $\{\sqrt{K_i/(1/2)\lambda_i h_i},\ \sqrt{K_i/(1/2)\lambda_i(h_i - h_w)}\colon i \in \mathscr{R}\}$ will fall *in* the interval (a, b). Hence, $L_w \cup E_w \cup G_w = \mathscr{R}$.

Step 2′. Use Lemmas 1, 2, and 3 to determine whether to include i in the optimal subset. Let S' be the optimal retailers' subset selected. Let $G_w \cap S' = G_w(S')$, $E_w \cap S' = E_w(S')$, and $L_w \cap S' = L_w(S')$. Define 11

$$T_w = \sqrt{\big[K_w + \textstyle\sum_{i \in E_w(S')} K_i\big] \big/ \big[\tfrac{1}{2}\sum_{i \in E_w(S')} \lambda_i h_i + \tfrac{1}{2}\sum_{i \in L_w(S')} \lambda_i h_w\big]}.$$

If $T_w \notin [a, b]$, then move to the next interval (i.e., our guess that T_w^* is in $[a, b]$ is wrong). Otherwise, calculate the value of the cost $f(S')$ using the equation

$$\frac{K_w}{T_w} + \sum_{i \in S'} \frac{K_i}{T_i} + \frac{1}{2}\sum_{i \in S'} \lambda_i h_i T_i + \frac{1}{2}\sum_{i \in S'} \lambda_i h_w [\max(T_w, T_i) - T_i] - \sum_{i \in S'} P_i.$$

Step 3. Go to Step 2 until it reaches the last interval. Choose the smallest one among all the $f(S')$ we get. The corresponding set S' and the corresponding value T_w is the optimal selection of retailers to serve in the warehouse and the optimal reorder interval of the warehouse, respectively.

Theorem 2. *Algorithm TS solves the submodular function minimization problem* $\min_S f(S)$.

Proof. Let S^* denote an optimal solution to the problem $\min_S f(S)$ with the maximum cardinality, i.e., among all the optimal solutions, S^* has the largest number of retailers selected. We need only to show that S^* will be constructed in Step 2 at a certain stage of the algorithm. Let T_w^* be the corresponding optimal warehouse reorder interval, and $L_w(S^*)$, $E_w(S^*)$, and $G_w(S^*)$ be the partitioning of retailers in S^* according to Proposition 1. Note that T_w^* must be in one of the intervals considered by the algorithm. We define the interval that contains T_w^* as $[a^*, b^*]$.

When we try for T_w^* to be in the interval $[a^*, b^*]$ in Step 2 of the algorithm, we next obtain the sets E_w, L_w, and G_w. Note that $L_w(S^*) \subseteq L_w$, $E_w(S^*) \subseteq E_w$, and $G_w(S^*) \subseteq G_w$ by construction. Furthermore, all retailers in $L_w(S^*)$ (respectively, $E_w(S^*)$, $G_w(S^*)$) satisfy condition (3) (respectively, (4), (5)); otherwise, S^* cannot be the optimal solution to the problem $\min_S f(S)$. Thus, $S^* \subseteq S'$, where S' is the subset constructed in Step 2 of the algorithm.

If there exists a retailer i in $L_w \backslash L_w(S^*)$ with the property that (3) holds, then $f(S^* \cup \{i\}) \leqslant f(S^*)$, contradicting the assumption that S^* is an optimal solution with the largest cardinality. Hence, every retailer in $L_w \backslash L_w(S^*)$ violates (3). Similarly, we can show that every retailer in $E_w \backslash E_w(S^*)$ and $G_w \backslash G_w(S^*)$ violates (4) and (5), respectively.

Hence, the set S' obtained at the end of Step 2 must be identical to S^*, because all retailers not in S^* will be eliminated from S'. The corresponding $T_w(S')$ obtained will be identical to T_w^* and, hence, will lie in the interval $[a^*, b^*]$. □

We use a small example to illustrate how the algorithm works.

Example.

	u_i	v_i	K_i	λ_i	h_i
1	8.6	1.2	1.5	1.7	2.3
2	9.2	1.4	1.6	1.8	2.2
3	7.6	0.9	1.4	1.6	2.1

$K_w = 9.8$ and $h_w = 1.2$.

• Calculate the value of l_i, T_i, T_i', y_i, and x_i, and sort them along the real line. We obtained the following 16 intervals.

Interval Number	$[a, b]$	Interval
1	$[0, x_1]$	$[0, 0.214903937]$
2	$[x_1, x_2]$	$[0.214903937, 0.217091667]$
3	$[x_2, x_3]$	$[0.217091667, 0.221227105]$
4	$[x_3, T_1]$	$[0.221227105, 0.875935743]$

Continued.

Interval Number	$[a, b]$	Interval
5	$[T_1, T_2]$	[0.875935743, 0.898933149]
6	$[T_2, T_3]$	[0.898933149, 0.912870929]
7	$[T_3, T_1']$	[0.912870929, 1.266600993]
8	$[T_1', T_2']$	[1.266600993, 1.333333333]
9	$[T_2', T_3']$	[1.333333333, 1.394433378]
10	$[T_3', y_1]$	[1.394433378, 3.570262303]
11	$[y_1, y_2]$	[3.570262303, 3.722302273]
12	$[y_2, y_3]$	[3.722302273, 3.766868132]
13	$[y_3, l_3]$	[3.766868132, 4.887516600]
14	$[l_3, l_1]$	[4.887516600, 4.932800141]
15	$[l_1, l_2]$	[4.932800141, 5.000000000]
16	$[l_2, +\infty)$	$[5.000000000, +\infty)$

- Choose one of the intervals, say Interval 6, to show how Steps 2 and 2′ in the algorithm work. Suppose we guess that T_w^* lies in Interval 6. So $G_w = \{3\}$, $E_w = \{1, 2\}$, and $L_w = \varnothing$. Because $\sqrt{2(K_3)(\lambda_3 h_3)} - P_3 = -4.531129326 \leqslant 0$, therefore $G_w(S') = \{3\}$. Because $x_1 = 0.214903937 \leqslant a = 0.898933149$, $y_1 = 3.570262303 \geqslant b = 0.912870929$ and $x_2 = 0.217091667 \leqslant a = 0.898933149$, $y_2 = 3.722302273 \geqslant b = 0.912870929$, therefore $E_w(S') = \{1, 2\}$. Because $L_w = \varnothing$, therefore $L_w(S') = \varnothing$. $T_w = \sqrt{[K_w + K_1 + K_2]/[(1/2)\lambda_1 h_1 + (1/2)\lambda_2 h_2]} = 1.810599878 \notin$ Interval 6. So our guess of T_w lying in this interval is not right. We move to the next interval (Interval 7). If the guess is right, we calculate its corresponding $f(S')$.
- After we finish the test of all the 16 intervals, we compare all the values of $f(S')$ we get in Step 2′, and choose the minimum of them, which is the optimal solution.

Theorem 3. *The computational complexity of the pricing problem is $O(n \log n)$, where $n = |\mathcal{R}|$.*

Proof. For ease of exposition, we assume that the $5n$ points obtained in Step 1 of Algorithm TS are distinct. Step 1 of Algorithm TS requires $O(n \log n)$ comparisons to sort the $5n$ points.

In Steps 2 and 2′, the number of operations can be performed in $O(n)$ operations for each nice interval. Instead of incurring this amount of computational effort for each interval, we show next that the computational effort for each subsequent interval can be performed in $O(1)$ time, by utilizing the information in the previous interval. To be more precise, let L_w, E_w, G_w, and S' be the sets obtained in Steps 2 and 2′ when we guessed that T_w^* lies in an interval, say $[a, b]$. Note that as we move from the interval $[a, b]$ to the next interval (say, $[b, c]$), at most one of the following happens:

- $b = \sqrt{K_i/(1/2)\lambda_i h_i}$ for some i,
- $b = \sqrt{K_i/(1/2)\lambda_i (h_i - h_w)}$ for some i,
- $b = l_i$ for some i,
- $b = x_i$ for some i, or
- $b = y_i$ for some i.

In the first instance, in the new interval $[b, c]$, the new G_w, L_w, and E_w can be updated simply as

$$G_w \leftarrow G_w \backslash \{i\}, \qquad E_w \leftarrow E_w \cup \{i\}, \qquad L_w \leftarrow L_w.$$

The new S' is obtained by simply checking condition (4) for retailer i (instead of condition (5), when considering the interval $[a, b]$), to ensure whether to include or exclude retailer i from the optimal subset. These operations can be performed in $O(1)$ operations.

Similarly, for all the other four instances, the new G_w, L_w, E_w, and S' can be obtained in $O(1)$ operations. Because there are $O(n)$ nice intervals to explore, it takes altogether $O(n)$ operations to execute Steps 2 and 2′ of the algorithm.

In Step 3 we need $O(n)$ comparisons. So the whole problem can be solved within $O(n \log n)$ operations. The pricing subproblem is thus solvable in $O(n \log n)$ time. □

5. Joint Location-Inventory Model

The warehouse-retailer network design problem is clearly NP-hard, because the problem contains the classical facility location problem as a special case (when we ignore the inventory-related cost). In fact, the best known approximation bound for the facility location problem is $O(\log |\mathcal{R}|)$ (for general transportation cost structure). See, for example, Bertsimas and Vohra (1998) for a neat proof of this result. Note that the classical facility location model considers only a trade-off between facility fixed cost and transportation cost. On the other hand, the warehouse-retailer network design problem considers trade-offs between facility fixed cost, transportation cost, and multiechelon inventory replenishment cost. In this section, we present a model to consider the trade-off between facility fixed cost and multiechelon inventory replenishment cost. Note that the complexity of the OWMR inventory replenishment is still open, although the results of Roundy (1985) showed that it is possible to solve the problem to within 98% optimality in polynomial time. It is thus interesting to examine as to whether the same approximation bound can be obtained for the warehouse-retailer network design problem, in the special case that the transportation costs are ignored. This gives rise to the *joint location-inventory* problem, where only the fixed cost and inventory replenishment cost are considered.

Lim et al. (2003) reported an interesting result for this class of problem. They proved that it is possible to attain within 15% of the optimum with only *one* warehouse serving all the retailers. Furthermore, there is a 98% optimal solution that uses only at most *two* warehouses. Unfortunately, even for the case with two warehouses, the issue of retailer-warehouse assignment is still left open, i.e., it is still nontrivial to determine which warehouse should be used to serve a retailer. We complete the puzzle here and answer this question in the rest of this section. The result shows that the joint location-inventory problem can be solved to 98% optimality in polynomial time. This extends the classical result of Roundy (1985) to the network design arena.

Theorem 4 (Lim et al. 1999). *For the joint location-inventory problem, it is possible to attain within* 2% *of the optimum with only two warehouses serving all the retailers.*

Next, we show that when we restrict ourselves to two warehouses, the retailer-warehouse assignment problem can be solved as the dual of a polymatroid intersection problem.

Theorem 5. *If there are only two warehouses, then in the set-partitioning model, the solution obtained by the column generation is the optimal integer solution. The cost solution we get is within* 2% *of the optimum.*

Proof. In this special case, our set-partitioning model (WRND) can be reduced to

$$\text{(IP1)} \quad \min \quad \sum_{S:\, i\in S} c_{w_1,S}\, y_S + \sum_{T:\, i\in T} c_{w_2,T}\, z_T$$

subject to

$$\sum_{S:\, i\in S} y_S + \sum_{T:\, i\in T} z_T = 1 \quad \forall S, T \subseteq \mathcal{R},$$

$$y_S, z_T \in \{0,1\} \quad \forall S, T \subseteq \mathcal{R}.$$

Because the cost function is monotone, the above is also equivalent to

$$\text{(IP2)} \quad \min \quad \sum_{S:\, i\in S} c_{w_1,S}\, y_S + \sum_{T:\, i\in T} c_{w_2,T}\, z_T$$

subject to

$$\sum_{S:\, i\in S} y_S + \sum_{T:\, i\in T} z_T \geqslant 1 \quad \forall S, T \subseteq \mathcal{R},$$

$$y_S, z_T \in \{0,1\} \quad \forall S, T \subseteq \mathcal{R}.$$

Let (LP2) denote the LP relaxation to (IP2), by replacing the constraints

$$y_S, z_T \in \{0,1\}$$

by

$$y_S, z_T \geqslant 0.$$

The dual of (LP2) is now

$$\max \quad x_1 + x_2 + \cdots + x_{|\mathcal{R}|}$$

subject to

$$\sum_{S:\, i\in S} x_i \leqslant c_{w_1,S} \quad \forall S \subseteq \mathcal{R},$$

$$\sum_{T:\, i\in T} x_i \leqslant c_{w_2,T} \quad \forall T \subseteq \mathcal{R},$$

$$x_i \geqslant 0.$$

Because $c_{w_i,S}$ is a nondecreasing submodular function for each fixed w_j, it follows that the dual is just a polymatroid intersection problem which is known to be totally dual integral (cf., Nemhauser and Wolsey 1988), and the (LP2) is known to have integral optimal solution. Therefore, the LP relaxation to (IP2) is integral. □

It follows that to solve the joint location-inventory problems, we need only to enumerate over pairs of two warehouses and solve the above dual problem, and choose the one with minimum cost. Note that the polymatroid intersection problem can also be solved using the column generation framework for the general distribution network design problem.

6. Computational Results

In this section, we summarize our computational experience with the algorithms outlined in the previous section. All the instances were solved on a COMPAQ P3-450 station running the Windows 2000 operating system.

6.1. Submodular Function Minimization

The pricing algorithm is coded in C++. All the computation exclude input times. All cost parameters are randomly generated using MATLAB. The location of the retailers and warehouses are uniformly distributed over $[0,1]\times[0,1]$.

Figure 1 shows a typical instance generated, with 10 warehouses and 50 retailers. The "*" symbols refer to the location of the warehouses in the plane, whereas the "o" symbols refer to the location of the retailers in the plane. We assume that the transportation cost is proportional to the Euclidian distance in the plane. K_i, h_i, λ_i are randomly generated in $(0,100]$ and K_w, h_w are generated uniformly in $(\max_i K_i, 100]$ and $(0, \min_i h_i)$, respectively.

Table 1 presents the relation between the average CPU time needed (over 20 different instances) and the number of retailers in the problem.

Figure 2 is plotted using the data in Table 1. Indeed, for the large retailers set, the submodular function minimization problem can be solved almost instantaneously, using only several seconds (CPU time) to solve a problem involving up to 200 retailers.

6.2. Distribution Network Design

The entire column generation algorithm for the general distribution network design problem is coded in C++, and the linear programming problems were solved using the primal simplex method. To speed up the column generation

Figure 1. Warehouse-retailer locations.

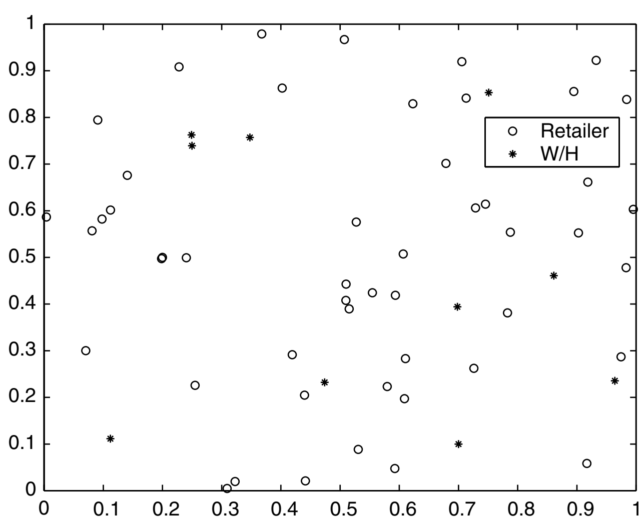


Table 1. CPU time for solving the pricing subproblem.

No. of Retailers	CPU Time (seconds)
5	0.02
10	0.07
25	0.21
50	0.53
100	1.24
150	2.09
200	3.12

algorithm and to avoid adding too many columns to the master problem, we do not solve the pricing subproblem at every iteration. Instead, in each iteration, after solving the master problem and obtaining the dual variables, we update the reduced costs of the columns in a separately maintained column pool and select some columns with negative reduced costs to add to the master problem. At the same time, we remove from the pool those with positive reduced costs. We solve the pricing subproblem only when the column pool is empty, and each time, we add multiple columns (with negative reduced costs) to the column pool.

We show some of the computational results we observed in the rest of this section. Table 2 highlights the results of our computational study. For each of the instances, we first solved the linear programming relaxation of the set-partitioning model via column generation. The initial set of columns includes all singletons. The column labeled "No. of Columns Generated" indicates the total number of columns added during this phase. In all instances generated, the corresponding optimal solutions were integral.

Example. We set the transportation and facility location costs to be zero in this example, and we consider only the

Figure 2. Number of retailers vs. CPU time(s) needed in the pricing subproblem.

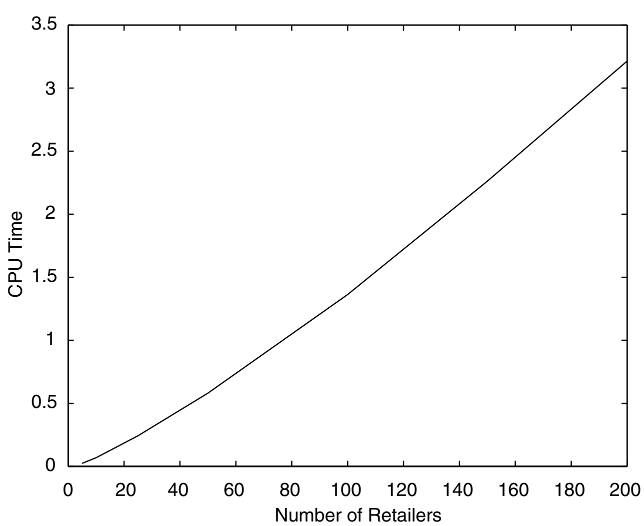


Table 2. Instances of the network design problem with different sizes.

No. of Warehouses	No. of Retailers	No. of Columns Generated	CPU Time (seconds)
2	3	12	0.5
3	10	27	2.3
8	10	44	6.4
8	20	117	15.6
10	50	924	116.3
20	100	6,878	815.8

inventory cost component. We study the following small system with two warehouses and four retailers:

i	K_i	λ_i	h_i
1	2.381910326	120.1545886	291.470977
2	0.009152176	0.419055781	297.1717484
3	0.045295548	2.114537948	291.470799
4	0.000001	60.54503063	291.4723

w	K_w	h_w
w_1	0.191924445	291.470797
w_2	68.44719417	0.27008984

This is the example presented in Lim et al. (1999) to illustrate the fact that it is not optimal to use only one warehouse in the distribution network design problem, even when the overriding concerns are inventory cost (i.e., zero transportation cost). They showed that the distribution network, with retailers 1, 2, and 3 assigned to warehouse 1 and retailer 4 assigned to warehouse 2, is better than any system using only one warehouse. We have tested the problem using the above algorithm and showed that indeed the example constructed by Lim et al. gives rise to the *optimal* distribution network for this instance.

For the rest of this section, we show some sensitivity results concerning the performance of the algorithm when the number of retailers varies with the number of warehouses fixed (at 10 and 20, respectively).

In Tables 3 and 4 and the two corresponding plots (Figures 3 and 4), we show the change in the number of columns generated when the number of retailers increases.

Table 3. Sensitivity analysis based on the 10 warehouses system.

No. of Warehouses	No. of Retailers	No. of Columns Generated	CPU Time (seconds)
10	10	55	8.8
10	20	162	22.2
10	30	423	48.6
10	50	924	116.3
10	70	2,102	202.2
10	100	4,736	478.7

Table 4. Sensitivity analysis based on the 20 warehouses system.

No. of Warehouses	No. of Retailers	No. of Columns Generated	CPU Time (seconds)
20	10	103	15.8
20	20	369	41.3
20	30	712	81.1
20	50	1,367	188.3
20	70	3,145	378.1
20	100	6,878	815.8

Note that the CPU time needed to solve the distribution network design problem increases exponentially with the increase in the number of retailers. The qualitative behaviour of the computational results does not change when the number of warehouses changes from 10 to 20.

Table 5 shows the trade-off between the inventory holding cost and the transportation cost in the 20 warehouses 100 retailers system (where the facility location cost is fixed at zero). We generated random instances of the problem and noted the number of warehouses opened (i.e., assigned to some retailers) in the optimal solution, and the percentage of the total cost attributed to inventory and transportation related activities. From the table, we can see a general trend: The number of warehouses needed in the optimal solution generally depends on the magnitude of the inventory cost with respect to the transportation cost. For a system where inventory cost is of paramount concern (e.g., the spare parts industry), the optimal network design will only use a few warehouses to support the replenishment activities. Note that this is despite the fact that we have not incorporated uncertainty in our demand model (in fact, our demand model is of the simplest type).

7. Extension and Generalization

A common constraint states that a warehouse cannot handle too many retailers (say, not more than k retailers can be served by a single warehouse) due to capacity or other technical limitations. In this section, we briefly describe how our technique can be extended to handle this additional constraint for some fixed k.

Figure 3. Number of retailers vs. CPU time(s).

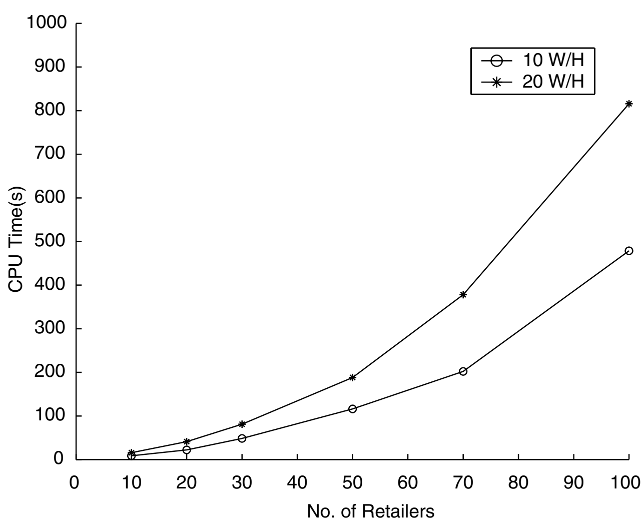
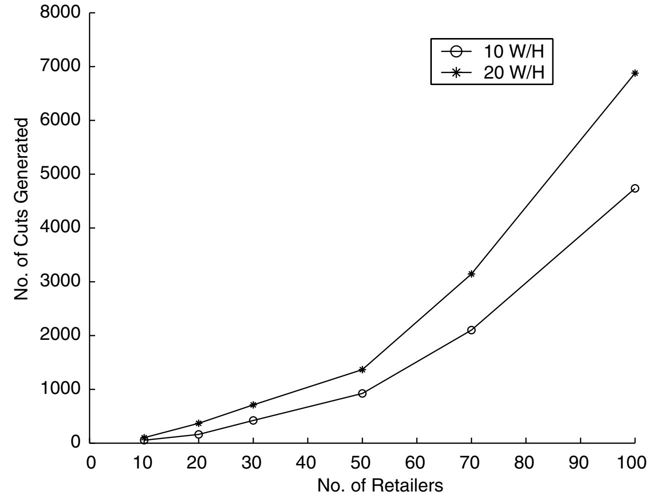


Figure 4. Number of retailers vs. number of cuts.



In this case, the column generation phase reduces to solving a problem of the type

$$\min_{S\subseteq \mathcal{R}} f(S)$$

subject to

$$|S| \leqslant k,$$

where $f(S)$ is the cost of the convex programming relaxation considered before when a fixed warehouse is being used to serve the retailers in S. Algorithm TS, proposed in this paper, cannot be used directly to solve this problem because the optimal solution S' obtained may not satisfy the condition that $|S'| \leqslant k$.

To modify the algorithm to obtain the optimal solution for this problem, we need to ensure that in Steps 2 and 2′ of Algorithm TS, after obtaining G_w, E_w, and L_w for each nice interval, we choose the corresponding optimal subset S' such that $|S'| \leqslant k$. To do so, we need to modify the selection criteria. Lemmas 1, 2, and 3 allow us to choose the retailers that will be profitable for inclusion in the optimal set S'.

Table 5. The effects of costs on the number of warehouses open.

Inventory Cost (%)	Transportation Cost (%)	No. of Warehouses Open
2.1	97.9	14
13.2	86.8	11
31.3	68.7	9
50.7	49.3	7
75.8	24.2	5
90.3	9.7	3
99.3	0.7	2

To make sure that not more than k retailers are selected, we need only to modify the selection criteria to choosing (at most) k most profitable retailers. The trick to doing this is to further partition the intervals into smaller subintervals, so that the relative profitability of the retailers (as a function of T_w) will not change within the smaller subintervals.

For each nice interval, say $[a, b]$, and for each pair of $i, j \in S'$, where S' is obtained from the original Algorithm TS, we solve the following system of equations:

- If $i, j \in L_w(S')$, solve

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w^* - P_i = 2\sqrt{K_j \times \tfrac{1}{2}\lambda_j(h_j - h_w)} + \tfrac{1}{2}\lambda_j h_w T_w^* - P_j.$$

- If $i \in L_w(S')$, $j \in E_w(S')$, solve

$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w^* - P_i = \frac{K_j}{T_w^*} + \frac{1}{2}\lambda_j h_j T_w^* - P_j.$$

- If $i, j \in E_w(S')$, solve

$$\frac{K_i}{T_w^*} + \frac{1}{2}\lambda_i h_i T_w^* - P_i = \frac{K_j}{T_w^*} + \frac{1}{2}\lambda_j h_j T_w^* - P_j.$$

- If $i \in E_w(S')$, $j \in G_w(S')$, solve

$$\frac{K_i}{T_w^*} + \frac{1}{2}\lambda_i h_i T_w^* - P_i = 2\sqrt{K_j \times \tfrac{1}{2}\lambda_j h_j} - P_j.$$

- If $i \in L_w(S')$, $j \in G_w(S')$, solve

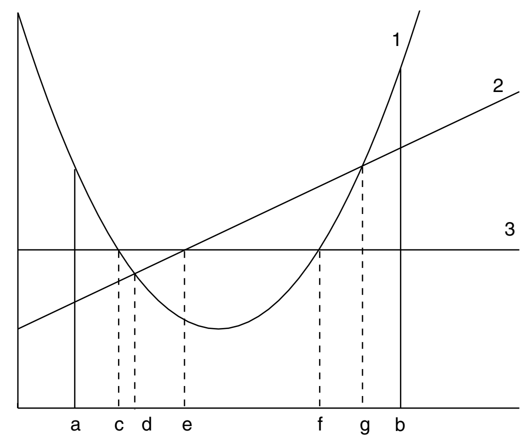
$$2\sqrt{K_i \times \tfrac{1}{2}\lambda_i(h_i - h_w)} + \tfrac{1}{2}\lambda_i h_w T_w^* - P_i = 2\sqrt{K_j \times \tfrac{1}{2}\lambda_j h_j} - P_j.$$

Note that within the interval $[a, b]$, these are the various ways the equations given by $(3, 4, 5)$ for different i, j can intersect. See Figure 5 for an illustration.

We can now subdivide the interval $[a, b]$ into smaller intervals using the solutions to the above system of equations. Within each smaller interval, note that the relative ranking of the profitability of each retailer will not change, as long as T_w^* falls within the smaller interval. This allows us to pick up to k most profitable retailers to serve.

To complete the modification, we only need to modify Step 3 of Algorithm TS to ensure that the T_w constructed from our choice of S' will fall within the smaller subinterval constructed above.

Figure 5. Illustration of the algorithm.



8. Conclusions and Future Research

In this paper, we have outlined a formulation of the warehouse-retailer distribution network system integrating transportation and the infinite horizon multiechelon inventory cost function. We consider the trade-off between inventory cost, direct shipment cost, and facility location cost in such a system. By solving this model, we can determine how many warehouses to open, where to locate them, how to serve the retailers using these warehouses, and the optimal inventory policies for warehouses and retailers to minimize the total multiechelon inventory, transportation, and facility location costs.

We formulate the problem as a mixed-integer set-partitioning problem. Two issues arise in this formulation. First, the number of columns required in the set-partitioning model is exponentially large. We attack this problem using column generation techniques. This leads to the second issue. The pricing problem that must be solved at each iteration of the column generation is a submodular function minimization problem. We showed how this problem can be solved efficiently using Algorithm TS. Computational results are provided for problems ranging in size from 2 to 20 warehouses and from 3 to 100 retailers. The results suggest that the moderate size warehouse-retailer network design problem can be solved in practice in a reasonable amount of time.

We outlined in the last section how the capacitated version of the warehouse-retailer network design problem can be handled by an extension of our technique. This paper can also be extended in many directions. For example, extending the problem to incorporate vehicle-routing cost instead of direct shipment cost would be both challenging and interesting.

Acknowledgments

The authors thank the associate editor and three referees for constructive comments that led to this improved version. In particular, they are grateful to the referees for pointing out gaps in an earlier presentation and computational results.

References

Baumol, W. J., P. Wolfe. 1958. A warehouse-location problem. *Oper. Res.* **6** 252–263.

Bertsimas, D., R. Vohra. 1998. Rounding algorithms for covering problems. *Math. Programming* **80** 63–89.

Chan, L. M. A., D. Simchi-Levi. 1998. Probabilistic analysis and algorithms for three-level distribution systems. *Management Sci.* **44** 1562–1576.

Chan, L. M. A., A. Muriel, Z. J. M. Shen, D. Simchi-Levi, C.-P. Teo. 2002. Effective zero-inventory-ordering policies for the single-warehouse multiretailer problem with piecewise linear cost structures. *Management Sci.* **48** 1446–1460.

Chandra, P., M. L. Fisher. 1994. Coordination of production and distribution planning. *Eur. J. Oper. Res.* **72** 503–517.

Chen, Y. H., M. Goh, J. Lee, J. Ou, M. Sharafali, C.-P. Teo, P.-G. Goh, P. Sivanathan. 2001. ST logistics: Distribution consumer goods in China. *Internat. Trans. Oper. Res.* **8**(2) 203–220.

Dror, M., M. Ball. 1987. Inventory/routing: Reduction from an annual to a short period problem. *Naval Res. Logist. Quart.* **34**(6) 891–905.

Dror, M., L. Levy. 1986. Vehicle routing improvement algorithms: Comparison of a "greedy" and a matching implementation for inventory routing. *Comput. Oper. Res.* **13**(1) 33–45.

Erlebacher, S. J., R. D. Meller. 2000. The interaction of location and inventory in designing distribution systems. *IIE Trans.* **32** 155–166.

Gallego, G., D. Simchi-Levi. 1990. On the effectiveness of direct shipping strategy for the one warehouse multi-retailer R-systems. *Management Sci.* **36** 240–243.

Geoffrion, A. M., R. F. Powers. 1995. 20 years of strategic distribution system design: An evolutionary perspective. *Interfaces* **25**(5) 105–127.

Golden, B., A. Assad, R. Dahl. 1984. Analysis of a large scale vehicle routing problem with an inventory component. *Large Scale Systems* **7**(2–3) 181–190.

Grötschel, M., L. Lovász, A. Schrijver. 1993. *Geometric Algorithms and Combinatorial Optimization*, 2nd corrected ed. Springer-Verlag, Berlin, Germany.

Herer, Y., R. Roundy. 1997. Heuristics for a one warehouse multiretailer distribution problem with performance bounds. *Oper. Res.* **45** 102–115.

Iwata, S., L. Fleischer, S. Fujishige. 2001. A strongly polynomial-time algorithm for minimizing submodular functions. *J. ACM* **48**(4) 761–777.

Lim, W.-S., J. Ou, C.-P. Teo. 2003. Inventory cost effect of consolidating several one-warehouse multiretailer systems. *Oper. Res.* **51**(4) 668–672.

Muckstadt, J. A., R. O. Roundy. 1993. Analysis of multistage production systems. S. C. Graves, A. H. G. Rinooy Kan, P. H. Zipkin, eds. *Logistics of Production and Inventory*. North-Holland, Amsterdam, The Netherlands, 59–131.

Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization*. John Wiley and Sons, New York.

Roundy, R. O. 1985. 98% effective integer-ratio lot-sizing for one warehouse multi-retailer systems. *Management Sci.* **31** 1416–1430.

Schrijver, A. 1999. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *J. Combin. Theory Ser. B* **80** 346–355.

Teo, C.-P., D. Bertsimas. 2001. Multistage lotsizing via randomized rounding. *Oper. Res.* **49**(4) 599–608.

Viswanathan, S., K. Mathur. 1997. Integrating routing and inventory decision in one warehouse multiretailer multiproduct distribution systems. *Management Sci.* **43** 294–312.