TRANSPORTATION SCIENCE
Vol. 40, No. 4, November 2006, pp. 439–454
ISSN 0041-1655 | EISSN 1526-5447 | 06 | 4004 | 0439

informs®

DOI 10.1287/trsc.1050.0140

The Period Vehicle Routing Problem with Service Choice

Peter Francis, Karen Smilowitz
Industrial Engineering and Management Sciences Department, Northwestern University, 2145 Sheridan Road, Technological Institute, Evanston, Illinois 60208-3119
{pete@northwestern.edu, ksmilowitz@northwestern.edu}

Michal Tzur
Industrial Engineering Department, Tel Aviv University, Tel Aviv 69978, Israel, tzur@eng.tau.ac.il

The period vehicle routing problem (PVRP) is a variation of the classic vehicle routing problem in which delivery routes are constructed for a period of time (for example, multiple days). In this paper, we consider a variation of the PVRP in which service frequency is a decision of the model. We refer to this problem as the PVRP with service choice (PVRP-SC). We explore modeling issues that arise when service choice is introduced, and suggest efficient solution methods. Contributions are made both in modeling this new variation of the PVRP and in introducing an exact solution method for the PVRP-SC. In addition, we propose a heuristic variation of the exact method to be used for larger problem instances. Computational tests show that adding service choice can improve system efficiency and customer service. We also present general insights on the impact of node distribution on the value of service choice.

Key words: logistics; vehicle routing; service choice
History: Received: April 2005; revision received: August 2005; accepted: September 2005.

Introduction

In many supply chains, as well as other pick-up and/or delivery operations, customers often require repeated visits over a time horizon. Examples can be found in the delivery of groceries (or other products), the collection of waste, or the distribution of equipment for intermodal operations. In these examples, one must design routes that consider volume, travel time, and visit frequency. Such applications motivated the development of the period vehicle routing problem (PVRP).

The PVRP is a generalization of the classic vehicle routing problem (VRP), in which vehicle routes are constructed for a t-day period (for example, one week). Other units of time may be used; however, this paper employs the above convention. Each day within the period, a fleet of capacitated vehicles performs routes that begin and end at a single depot. Customers are visited a preset number of times over the period, with a schedule that is chosen from a menu of schedule options. Each schedule option represents a set of days on which a node is visited. For example, if over the period of one week a node is to be visited twice, the menu options may be $\{(Mon, Tues); (Mon, Wed); (Wed, Fri)\}$. The objective of the PVRP is to find a set of tours for each vehicle over the period that minimizes total travel time while satisfying operational constraints (vehicle capacity and visit requirements).

The first problem motivating the PVRP was introduced in Beltrami and Bodin (1974) for assigning hoist compactor trucks in municipal waste collection. The PVRP was formally defined in Russell and Igo (1979) as an "assignment routing problem" and first formulated mathematically in Christofides and Beasley (1984). Solution methods in these and subsequent papers have focused on two-stage (construction and improvement) heuristics; see also Tan and Beasley (1984) and Russell and Gribbin (1991). Chao, Golden, and Wasil (1995) review the heuristics proposed in the above papers and compare these methods with a new heuristic developed to overcome issues of poor local optima. Gaudioso and Paletta (1992) consider a PVRP model that minimizes fleet size. More recently, Cordeau, Gendreau, and Laporte (1997) implement a tabu search algorithm for the PVRP. These papers all consider heuristic methods whose quality is unknown because no optimal solutions or lower bounds are provided. However, because these papers share common data sets, results are compared across papers.

In the abovementioned references, each customer is visited with a preset frequency. Each node may be served from a node-specific set of schedule options with a fixed number of visits per week. However, determining schedule choices for a node independent of routing decisions may lead to routing inefficiencies.

Newman, Yano, and Kaminsky (2005) consider a scheduling and routing problem in which the schedule and frequency of visits to customers is a decision variable. The customer's cost of holding inventory is used as an approximation of the willingness to pay for more frequent service. The authors note that the PVRP literature emphasizes routing costs or fleet size and does not consider inventory or other costs associated with the selection of a particular day-of-the-week schedule. However, the model in Newman, Yano, and Kaminsky (2005) assumes that route construction is exogenous; routes are selected from a given set of routes.

In this paper, we consider a variation of the PVRP that accounts for service choice. This work was initially motivated by the delivery operations for interlibrary loan items, described in §3.1. We define a schedule to be the set of days on which a node is visited. Service to a node is characterized by the schedule of visits and the associated visit frequency. In particular, we assume that each customer requires a minimum number of visits per period, but is willing to accept a higher frequency. The customer and/or the system benefits from higher frequencies, which is accounted for in the objective function. The benefit from a higher frequency may represent the customer's savings in holding cost or, more generally, the customer's willingness to pay for more frequent (better) service. Thus, the PVRP with service choice (PVRP-SC) exploits possible efficiencies from combined routing and service decisions. Importantly, due to the service choice, the overall performance of the system may be enhanced because some nodes may receive better service than the minimum required. An implicit constraint in our modeling of the PVRP-SC is to maintain a reasonable set of different routes for each driver to perform.

Like the PVRP-SC, the inventory routing problem (IRP) determines visit frequency and route configuration simultaneously. In the PVRP-SC, the amount delivered to a node is determined by the schedule assigned to the node (all demand accumulated since the last visit), while in the IRP the amount delivered to a node is a decision variable separate from visit frequency. Our modeling of the PVRP-SC ensures that minimum service requirements are met. The service benefit is related to the node and depends on a parameter determined by the node's volume and the chosen service. On the other hand, in the IRP the service-related costs are modeled as holding costs and therefore are associated with each specific unit of item. Typically, no minimum frequencies are imposed.

As with the vast majority of the inventory literature, the IRP literature may be classified into models in continuous time with a constant demand rate over an infinite horizon, and models in discrete time with a finite horizon and varying demand. In continuous time the visit frequency is typically modeled as a continuous decision variable that contributes a nonlinear term to the objective function. A variation of an EOQ-type solution is usually suggested to solve for the replenishment quantity decision variables. Work in this category includes, for example, Anily and Federgruen (1990), Chan, Federgruen, and Simchi-Levi (1998), and the surveys by Federgruen and Simchi-Levi (1995) and by Anily and Bramel (1998). Kleywegt, Nori, and Savelsbergh (2002) provide a classification of recent articles on both deterministic and stochastic demand IRPs. The IRP literature in discrete time is quite sparse, and usually considers special cases of the problem or suggests heuristic solutions. See, for example, Kim and Kim (2000), Gaur and Fisher (2004), and Herer, Tzur, and Setty (2003). Because demand in these models varies from period to period, the solution is typically not repetitive.

In this paper we introduce the PVRP-SC. We consider how service choice impacts the volume accumulated at a node, as well as the stopping cost. This paper explores modeling issues that arise when service choice is introduced and suggests efficient solution methods. We highlight the assumptions that allow for a compact formulation and an exact solution method. Contributions are made both in the formulation of this new variation of the PVRP and in the introduction of an exact solution method for the PVRP-SC. In addition, we propose a heuristic variation of the exact method to be used for larger problem instances.

Section 1 describes the PVRP-SC in detail and presents a mathematical formulation. Section 2 introduces the solution method and §3 presents computational results and general insights. Finally, §4 summarizes the paper and discusses future work.

1. Problem Description

The PVRP-SC is the problem of finding a set of tours for each vehicle for each day over a period that minimizes an objective of total travel cost minus service benefit while satisfying operational constraints (vehicle capacity and visit frequency minima). The example in Figure 1 illustrates how introducing choice in visit frequency can lead to greater control over system efficiency. In the figure, a set of nodes is served from a central depot by two vehicles. Each node is characterized by preset visit frequency (either daily, three-day, or two-day), and vehicles are assigned to nodes as shown. Consider Nodes 1 and 2 served by the second vehicle. Because the nodes have different

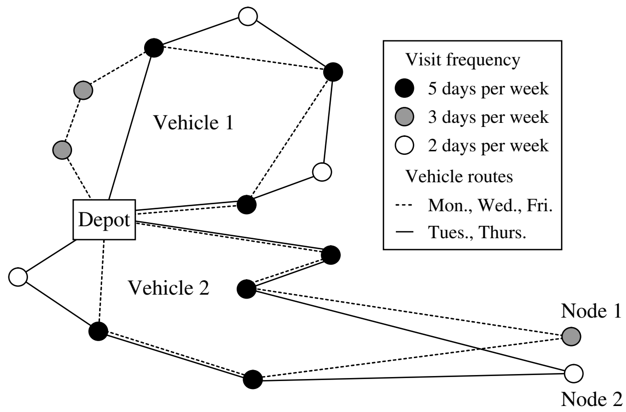

Figure 1 Period Vehicle Routing Problem

preset frequencies, they are visited by different schedules. Vehicle 2 must travel to the region containing Nodes 1 and 2 five days a week, incurring a long travel distance for each trip. If the preset frequencies are treated as lower bounds rather than fixed values, routing efficiency may increase. Node 2 can be visited with Node 1 on a Monday-Wednesday-Friday tour, thereby reducing the travel distance for Vehicle 2.

The increase in efficiency comes with an increase in solution options for an already difficult combinatorial optimization problem. The PVRP-SC is an extension of the VRP, which is shown by Lenstra and Rinnooy Kan (1981) to be NP-hard. Further, the amount delivered to a node must reflect the dependence of accumulation on schedule choice. The stopping cost at a node increases with the amount of material distributed or picked up, which is again a direct function of schedule choice.

The following notation is used to model and formulate the PVRP-SC:

N set of demand nodes; $N = \{1, \ldots, n\}$; node 0 represents the depot
A set of network arcs; $A = \{(i, j): i, j \in N \cup \{0\}\}$
K set of vehicles
C vehicle capacity; (*items per vehicle*)
T set of days $T = \{1, \ldots, t\}$; t represents the length of the period
S set of service schedules; $S = \{1, \ldots, |S|\}$; $s \in S$ is a subset of T

$$a_{sd} = \begin{cases} 1 & \text{if day } d \in T \text{ is in schedule } s \in S \\ 0 & \text{otherwise} \end{cases}$$

t_{ij} travel cost on arc $(i, j) \in A$; (*dollars*)
w_i demand at node $i \in N$; (*items per day*)[1]
f_i minimum visit frequency at node $i \in N$; (*number of days/period*)
γ^s service frequency for schedule $s \in S$; (*number of days*)
τ_i^s stopping cost at node $i \in N$ when served by schedule $s \in S$ (*dollars/stop*)
α^s service benefit for schedule $s \in S$ (*dollars/item*)
β^s demand accumulation adjustment factor for schedule $s \in S$.

The cost of stopping at a node τ_i^s is a function of the demand at node i and the frequency of schedule s. We introduce a service benefit α^s to provide incentive to offer more frequent service to nodes. The formulation can be generalized to relate frequency and service benefit through a volume-related function by defining a parameter α_i^s to be a function of w_i rather than a constant for all nodes. In this way, various types of benefit structures (e.g., quadratic) may be modeled.

Formulating capacity constraints is more complex in the PVRP-SC than in the PVRP. To ensure that capacities are not exceeded, one must know the volume delivered to a node. In existing formulations of the PVRP, it is assumed that the same amount is delivered at each visit to a node, regardless of the spacing of days in a schedule (which can vary even with fixed service levels). In the PVRP-SC, the service frequency determines the demand between visits; the adjustment factor β^s is estimated as the maximum number of days between visits on schedule $s \in S$. This conservative estimate for β^s guarantees feasibility, yet may preclude better solutions.

We introduce the following decision variables:

$$y_{ik}^s = \begin{cases} 1 & \text{if node } i \in N \text{ is visited by vehicle } k \in K \\ & \text{on schedule } s \in S \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ijk}^d = \begin{cases} 1 & \text{if vehicle } k \in K \text{ traverses the arc } (i, j) \in A \\ & \text{on day } d \in T \\ 0 & \text{otherwise.} \end{cases}$$

Because the capacity constraints depend on the specific vehicle and service level at each node, we define y_{ik}^s such that nodes are visited by the same vehicle each time. Allowing nodes to be visited by multiple vehicles would require either a nonlinear capacity constraint or a fifth index on the routing variables for schedule choice. In the motivating example of interlibrary loan delivery, this definition of y_{ik}^s is required.

The following formulation for PVRP-SC is an extension in several dimensions of the VRP formulation in Fisher and Jaikumar (1981):

$$Z^* = \min \sum_{k \in K} \Bigg[\sum_{d \in T} \sum_{(i, j) \in A} t_{ij} x_{ijk}^d + \sum_{s \in S} \sum_{i \in N} (\gamma^s \tau_i^s - w_i \alpha^s) y_{ik}^s \Bigg] \tag{1a}$$

[1] It is assumed that each item consumes the same amount of capacity.

subject to

$$\sum_{s\in S}\sum_{k\in K}\gamma^s y^s_{ik}\ge f_i \quad \forall\, i\in N \tag{1b}$$

$$\sum_{s\in S}\sum_{k\in K} y^s_{ik}\le 1 \quad \forall\, i\in N \tag{1c}$$

$$\sum_{s\in S}\sum_{i\in N}(\beta^s w_i)a_{sd}y^s_{ik}\le C \quad \forall\, k\in K;\ d\in T \tag{1d}$$

$$\sum_{j\in N\cup\{0\}} x^d_{ijk}=\sum_{s\in S}a_{sd}y^s_{ik} \quad \forall\, i\in N;\ k\in K;\ d\in T \tag{1e}$$

$$\sum_{j\in N\cup\{0\}} x^d_{ijk}=\sum_{j\in N\cup\{0\}} x^d_{jik} \quad \forall\, i\in N\cup\{0\};\ k\in K;\ d\in T \tag{1f}$$

$$\sum_{i,\,j\in Q} x^d_{ijk}\le |Q|-1 \quad \forall\, Q\subseteq N;\ k\in K;\ d\in T \tag{1g}$$

$$y^s_{ik}\in\{0,1\} \quad \forall\, i\in N;\ k\in K;\ s\in S \tag{1h}$$

$$x^d_{ijk}\in\{0,1\} \quad \forall\,(i,j)\in A;\ k\in K;\ d\in T. \tag{1i}$$

The objective function (1a) balances travel time and service benefit. The first term represents arc travel times. The second term represents the node stopping costs and a demand-weighted service benefit. Setting $\alpha^s=0$, $\forall\, s\in S$, is equivalent to considering only routing in the objective, and specifying large values of α^s solves the problem of maximizing service benefits first, and minimizing routing costs as a secondary objective.

Constraints (1b) enforce the minimum frequency of visits for each node. Constraints (1c) ensure that one schedule and one vehicle are chosen for each demand node. Constraints (1d) represent vehicle-capacity constraints. The material distributed to or picked up from a node depends on the schedule and node demand, and is allocated to specific days using the parameter a_{sd}. Constraints (1e) link the x and y variables for the demand nodes. Constraints (1f) ensure flow conservation at each node. Constraints (1g) are the subtour elimination constraints and ensure that all tours contain a visit to the depot. Constraints (1h) and (1i) define the binary variables for assignment and routing, respectively.

Formulation (1) can be modified to include other operating constraints that also make the problem more tractable. As mentioned earlier, it may be desirable from an operations perspective to limit the number of different routes performed by each driver to ensure that solutions can be implemented easily. Consider the example in Figure 1. The solution offers three discriminating service levels to nodes (daily service, three times a week, twice a week), while requiring each driver to perform only two different routes. This example can be generalized to any set of schedules that satisfies the property introduced in the following lemma.

Lemma 1. *Assume that the set of schedules S includes $|S|-1$ disjoint schedules (schedules do not share any common days) and a schedule $|S|$ that is the union of all disjoint schedules. Then the set of nodes served on a route on a certain day included in disjoint schedule s' is always visited on the same route each day in schedule s'. This results in at most $|S|-1$ different routes for each vehicle.*

Proof. Let M be a set of nodes visited on a certain day. If there exists no $i\in M$ such that i is assigned to schedule $|S|$, then all nodes $i\in M$ must be assigned to the same schedule by the disjoint assumption. Otherwise, denote by M' the set of nodes that are assigned to schedule $|S|$. Nodes in $M\backslash M'$ are assigned to the same schedule as in the previous case. Now, by definition of schedule $|S|$, all nodes in M' are visited too whenever nodes in $M\backslash M'$ are visited. Thus, the set of nodes visited on this day and all others in the schedule does not change. Given the same set of nodes, the same vehicle assignments and routing decisions are still optimal. Given $|S|-1$ disjoint schedules, there will be at most $|S|-1$ different routes for each vehicle. $\square$

Lemma 1 implies that the number of different routes does not depend on t. With sets of schedules defined as in Lemma 1, the number of routing variables (x^d_{ijk}) can be reduced. Let U be the set of disjoint schedules $\{1,\ldots,|S|-1\}$. Rather than defining x^d_{ijk} over all days, the variables are defined by unique delivery days, x^u_{ijk}; $u\in U$. In the example in Figure 1, $u=1$ corresponds to a Mon-Wed-Fri schedule that is performed three times a week ($\gamma^1=3$) and $u=2$ corresponds to a Tue-Thurs schedule that is performed twice a week ($\gamma^2=2$). Hence, γ^s can also be indexed by u. If the union of the disjoint sets were not included in the schedule options, there would be no need to differentiate between u and s in the formulation. With the union included, the indices on a_{sd} and x^d_{ijk} are changed from d to u rather than introducing new parameters for simplicity of notation. The units remain the same (days) although the dimensions over which the parameters/variables are indexed have changed. The parameter a_{su} is defined as

$$a_{su}=\begin{cases}1 & \text{for } s=u;\ s\in S,\ u\in U\\ 1 & \text{for } s=|S|;\ u\in U\\ 0 & \text{for } s\neq u;\ s\in S\backslash\{|S|\},\ u\in U.\end{cases} \tag{2}$$

The PVRP-SC is now formulated as follows:

$$Z^*=\min\sum_{k\in K}\Bigg[\sum_{u\in U}\sum_{(i,\,j)\in A}\gamma^u t_{ij}x^u_{ijk} + \sum_{s\in S}\sum_{i\in N}(\gamma^s\tau^s_i-w_i\alpha^s)y^s_{ik}\Bigg] \tag{3a}$$

subject to

$$\sum_{s\in S}\sum_{k\in K}\gamma^s y_{ik}^s \geq f_i \quad \forall\, i\in N \tag{3b}$$

$$\sum_{s\in S}\sum_{k\in K} y_{ik}^s \leq 1 \quad \forall\, i\in N \tag{3c}$$

$$\sum_{s\in S}\sum_{i\in N}(\beta^s w_i) a_{su} y_{ik}^s \leq C \quad \forall\, k\in K;\ u\in U \tag{3d}$$

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{s\in S} a_{su} y_{ik}^s \quad \forall\, i\in N;\ k\in K;\ u\in U \tag{3e}$$

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{j\in N\cup\{0\}} x_{jik}^u \quad \forall\, i\in N\cup\{0\};\ k\in K;\ u\in U \tag{3f}$$

$$\sum_{i,\,j\in Q} x_{ijk}^u \leq |Q|-1 \quad \forall\, Q\subseteq N;\ k\in K;\ u\in U \tag{3g}$$

$$y_{ik}^s \in \{0,1\} \quad \forall\, i\in N;\ k\in K;\ s\in S \tag{3h}$$

$$x_{ijk}^u \in \{0,1\} \quad \forall\,(i,j)\in A;\ k\in K;\ u\in U. \tag{3i}$$

Note that the routing component in (3a) is multiplied by γ^u to reflect the full cost of routing throughout the time period. Formulation (3) allows consistent tours for drivers and easier system management. Further, each node that is visited by a disjoint schedule is visited by the same vehicle at the same time each day. This variation is adopted throughout the remainder of the paper.

2. Solution Methods

Section 2.1 provides an overview of the solution algorithm. The first component is a Lagrangian relaxation of (3), in which two subproblems are created by relaxing one constraint of the original problem. Critical to the Lagrangian procedure are the initial lower and upper bounds, and the methods used to update the bounds, described in §2.2 for the lower bounds and §2.3 for the upper bounds. If the lower and upper bounds converge to identical values, the optimal solution is reached and the algorithm terminates. Otherwise, the branch-and-bound component is implemented to close the gap, as described in §2.4. The branch-and-bound procedure uses information from the Lagrangian relaxation to improve performance. With slight variations to this exact algorithm, we design a heuristic algorithm for large problem instances that is described in §2.5. While our algorithm includes computations typically employed in such algorithms, we enhance those computations that are specific to the PVRP-SC and are developed in this research. The details of these developments are described in this section according to the outline above.

2.1. Description of the Lagrangian Relaxation

By relaxing constraints (3e), the assignment decisions (y-variables) are completely separated from the routing decisions (x-variables). The two resulting subproblems are solved independently and the relaxed constraints are incorporated in a Lagrangian fashion. We introduce Lagrange multipliers λ_{ik}^u for $i\in N$, $k\in K$, and $u\in U$, associated with the relaxed constraints (3e), and define $\lambda_{0k}^u = 0$ because constraints (3e) are not defined for $i=0$. We obtain the following Lagrangian function:

$$\begin{aligned} LR(\lambda) &= \sum_{k\in K}\left[\sum_{u\in U}\sum_{(i,\,j)\in A}\gamma^u t_{ij} x_{ijk}^u + \sum_{s\in S}\sum_{i\in N}(\gamma^s\tau_i^s - w_i\alpha^s) y_{ik}^s\right] \\ &\quad + \sum_{u\in U}\sum_{k\in K}\sum_{i\in N}\lambda_{ik}^u\left(\sum_{j\in N\cup\{0\}} x_{ijk}^u - \sum_{s\in S} a_{su} y_{ik}^s\right) \\ &= \sum_{k\in K}\sum_{u\in U}\sum_{(i,\,j)\in A}(\gamma^u t_{ij}+\lambda_{ik}^u) x_{ijk}^u \\ &\quad + \sum_{k\in K}\sum_{s\in S}\sum_{i\in N}(\gamma^s\tau_i^s - w_i\alpha^s) y_{ik}^s \\ &\quad - \sum_{k\in K}\sum_{u\in U}\sum_{i\in N}\sum_{s\in S}\lambda_{ik}^u a_{su} y_{ik}^s. \end{aligned} \tag{4}$$

For a given λ vector, we define

$$Z_\lambda(x,y) = \min_{x,\,y} LR(\lambda)$$
$$\text{subject to} \quad \text{(3b)–(3d), (3f)–(3i).}$$

The solutions to the following two independent subproblems provide the x and y values that minimize the Lagrangian function. The assignment subproblem determines the y values:

$$Z_\lambda(y) = \min \sum_{k\in K}\sum_{s\in S}\sum_{i\in N}(\gamma^s\tau_i^s - w_i\alpha^s) y_{ik}^s - \sum_{k\in K}\sum_{u\in U}\sum_{i\in N}\lambda_{ik}^u\left(\sum_{s\in S} a_{su} y_{ik}^s\right) \tag{5a}$$

subject to

$$\sum_{s\in S}\sum_{k\in K}\gamma^s y_{ik}^s \geq f_i \quad \forall\, i\in N \tag{5b}$$

$$\sum_{s\in S}\sum_{k\in K} y_{ik}^s \leq 1 \quad \forall\, i\in N \tag{5c}$$

$$\sum_{s\in S}\sum_{i\in N}(\beta^s w_i) a_{su} y_{ik}^s \leq C \quad \forall\, k\in K;\ u\in U \tag{5d}$$

$$y_{ik}^s \in \{0,1\} \quad \forall\, i\in N;\ k\in K;\ s\in S. \tag{5e}$$

The routing subproblem determines the x values:

$$Z_\lambda(x) = \min \sum_{k\in K}\sum_{u\in U}\sum_{(i,\,j)\in A}(\gamma^u t_{ij}+\lambda_{ik}^u) x_{ijk}^u \tag{6a}$$

subject to

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{j\in N\cup\{0\}} x_{jik}^u \quad \forall\, i\in N\cup\{0\};\ k\in K;\ u\in U \tag{6b}$$

$$\sum_{i,\,j\in Q} x_{ijk}^u \leq |Q|-1 \quad \forall\, Q\subseteq N;\ k\in K;\ u\in U \tag{6c}$$

$$x_{ijk}^u \in \{0,1\} \quad \forall\,(i,j)\in A;\ k\in K;\ u\in U. \tag{6d}$$

The routing subproblem decomposes by (k, u) pairs for each vehicle/delivery day combination, resulting in $|K| \cdot |U|$ instances of the following subproblem, in which k and u are given:

$$Z_\lambda(x_{(k,u)}) = \min \sum_{(i,j)\in A} (\gamma^u t_{ij} + \lambda_{ik}^u) x_{ijk}^u \tag{7a}$$

subject to

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{j\in N\cup\{0\}} x_{jik}^u \quad \forall\, i \in N \cup \{0\} \tag{7b}$$

$$\sum_{i,j\in Q} x_{ijk}^u \le |Q| - 1 \quad \forall\, Q \subseteq N \tag{7c}$$

$$x_{ijk}^u \in \{0, 1\} \quad \forall\, (i, j) \in A. \tag{7d}$$

For each $\lambda \in \Re^{n+|K|+|U|}$, it follows that

$$Z_\lambda(x, y) = Z_\lambda(x) + Z_\lambda(y) = \sum_{(k,u)} Z_\lambda(x_{(k,u)}) + Z_\lambda(y) \quad \text{and}$$

$$Z_\lambda(x, y) \le Z^*.$$

Each resulting subproblem is a difficult problem by itself. In the assignment subproblem, a schedule and a vehicle must be assigned to each node, such that frequency and capacity constraints are satisfied. The frequency constraints are easy to satisfy, and the best feasible schedule for each node may be chosen according to the node's coefficient. While the capacity constraint is a "knapsack" type of constraint that is theoretically difficult, the assignment problems can be solved to optimality with CPLEX in seconds.

The (k, u) routing subproblem is known in the literature as a version of the traveling salesman problem with profits (TSP with profits), in which a profit is associated with each node and it is not necessary to visit all nodes. The objective is to find a route with maximum profit and minimum travel cost. In some versions, one objective is optimized subject to constraints on the value of the other objective; in other versions, both objectives are weighted, as in our case. The problem can also be modeled as the orienteering problem, or the prize collecting TSP; see Feillet, Dejax, and Gendreau (2003) for a survey and classification of the TSP with profits. Feillet, Dejax, and Gendreau (2003) also prove formally that the TSP with profits is NP-hard. Because the problem is solved for every (k, u) pair at each Lagrangian iteration, investing the time to solve (7) optimally is not realistic. Alternatively, lower and upper bounds may be embedded within the Lagrangian iterations; see §§2.2 and 2.3, respectively.

After each iteration of the Lagrangian algorithm, the Lagrange multipliers are updated according to a standard subgradient optimization procedure; see Fisher (1981) and Fisher (1985).

2.2. Lower Bounds

The subgradient optimization procedure for updating the Lagrange multipliers requires an initial lower bound for the problem. One simple lower bound is represented by the following formula:

$$\sum_{i\in N} f_i\Big(\min_{j\in J} t_{ij}\Big) + \sum_{i\in N}\big(\gamma^{\bar{s}(i)}\tau_i^{\bar{s}(i)} - w_i\alpha^{\hat{s}}\big)$$

where $\bar{s}(i)$ is the schedule that satisfies $\gamma^{\bar{s}(i)}\tau_i^{\bar{s}(i)} = \min_{s\in S}\{\gamma^s\tau_i^s\}$ for node i, and $\hat{s}$ is the schedule that satisfies $\hat{s} = \max_{s\in S}\{\alpha^s\}$.

An alternative lower bound may be found by computing the routing term based on a minimum spanning tree (MST) method, which constructs an MST through all the nodes. The nodes are assumed to be visited at the (same) lowest possible frequency, with stopping costs and service benefit determined as above. Let C_{MST} be the cost of the MST. This lower bound is given by

$$\Big(\min_{i\in N} f_i\Big) C_{MST} + \sum_{i\in N}\big(\gamma^{\bar{s}(i)}\tau_i^{\bar{s}(i)} - w_i\alpha^{\hat{s}}\big).$$

The lower bound is updated at each iteration of the Lagrangian algorithm by solving the subproblems for the given Lagrange multiplier values. Ideally, optimal solutions would be found for each subproblem; however, this is not realistic for the routing subproblem. Therefore, we set the iteration's lower bound as the sum of the optimal solution to the assignment problem and the lower bounds on the routing subproblems.

The difficulty in solving the routing subproblem is due to the exponential number of subtour elimination constraints (7c). Typically, only some of those constraints are needed; however, the specific constraints required are not known in advance. Therefore, the following approach is adopted. Initially, all subtour elimination constraints are relaxed. If subtours exist in the resulting solution, the associated subtour elimination constraints are added to the formulation, and the problem is re-solved (with CPLEX). This is repeated until either a feasible solution is obtained or some stopping rule applies. In the first (second) case the solution is optimal (a lower bound) for the routing subproblem and in both cases the objective value is a lower bound on the routing component of the original objective value. We use a predetermined time budget as a stopping rule for the routing subproblem. See §2.5 for a discussion on this choice. When a time budget is used, the following cut is added for every (k, u) routing subproblem (7) to ensure that the depot is included in the routes:

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u \le \sum_{j\in N} x_{0jk}^u \quad \forall\, i \in N. \tag{7e}$$

Let $LB_\lambda(x)$ be a lower bound on $Z_\lambda(x)$ and let $LB_\lambda(x_{(k,u)})$ be a lower bound on $Z_\lambda(x_{(k,u)})$ for a given λ. We obtain

$$Z_\lambda(y)+\sum_{(k,u)} LB_\lambda(x_{(k,u)}) = Z_\lambda(y)+LB_\lambda(x)\le Z_\lambda(y)+Z_\lambda(x) = Z_\lambda(x,y)\le Z^*.$$

2.3. Upper Bounds

Upper bounds are used in various phases of the algorithm. At the end of the algorithm, the best feasible upper bound represents the suggested solution, whether it is a provably optimal solution or a heuristic solution. The capacity constraint makes the feasibility problem hard, particularly when searching for an initial upper bound, and especially if the capacity constraint is tight.

Recall that the objective function (3a) consists of total time (travel plus stopping) to perform all routes during the period

$$\sum_{k\in K}\left[\sum_{u\in U}\sum_{(i,j)\in A}\gamma^u t_{ij}x^u_{ijk}+\sum_{s\in S}\sum_{i\in N}\gamma^s\tau^s_i y^s_{ik}\right]$$

and demand-weighted service factor for all nodes

$$\sum_{k\in K}\left[\sum_{s\in S}\sum_{i\in N}(-w_i\alpha^s)y^s_{ik}\right].$$

Each component can be bounded from above. An upper bound on the first component is based on the assessment that, in practice, each vehicle has a limited time Γ to perform its tour, even if not explicitly expressed by the constraints. (See §4 for a discussion of the implications of such a constraint.) An upper bound on the total time traveled during the period by all vehicles is given by $|K|\cdot t\cdot\Gamma$ (the parameter Γ may be chosen somewhat loosely if its value is not known explicitly). The second component is bounded by $-\sum_{i\in N} w_i\alpha^{s'}$ where s' is the schedule that satisfies: $\alpha^{s'}=\min_{s\in S}\{\alpha^s\}$. Therefore, the upper bound on the objective function is given by: $|K|\cdot t\cdot\Gamma-\sum_{i\in N} w_i\alpha^{s'}$.

The initial upper bound serves in the subgradient optimization step that updates the initial Lagrange multipliers. In each subsequent iteration of the LR algorithm, we attempt to improve the upper bound. Given the assignment decisions from solving (5), we search for feasible vehicle routings for these assignments using a randomized saving heuristic for the VRP problem from Clarke and Wright (1964), described in the appendix. Because the assignment decisions are feasible, the heuristic solution is guaranteed to be feasible as well. This solution value may update the best known upper bound. This method can also be used to find an initial upper bound if Γ is not known.

Upper bounds may also be obtained from solutions to the routing subproblem. Two feasibility issues must be addressed with solutions to (7). First, the routing solutions must not contain subtours. Second, the assignment and routing decisions must satisfy the relaxed constraints (3e). Therefore, in each iteration in which the solutions to (7) do not contain any subtours, we fix the assignment variables using linking constraints (3e). We then check the capacity and frequency constraints. If the solution is feasible, it may update the upper bound.

2.4. Closing the Gap: The Branch-and-Bound Procedure

At the conclusion of the Lagrangian relaxation phase, the solution that corresponds to the best known upper bound is a candidate solution to the problem. If the upper bound does not equal the lower bound, the remaining gap is closed with branch and bound. Typically, branch and bound is not a preferred method for hard combinatorial problems like the PVRP-SC because significant computational effort and time are required. While this is still a valid concern, there are two factors that facilitate the branch-and-bound procedure. First, the upper bound from the LR phase is likely to be relatively close to the optimal solution, certainly much closer than the initial upper bound. This claim is supported with empirical evidence in §3. As a result, significant parts of the branch-and-bound tree are likely to be truncated. Second, solutions to the subproblems from the Lagrangian phase eliminate the need for repetitive computation and provide useful information in guiding the branch-and-bound algorithm, as explained below.

Due to the large number of routing variables, the branch-and-bound procedure enumerates the assignment variables only. The routing variables are determined subsequently through the solution to the routing subproblem, either optimally or via a lower bound. A key observation here is that, given a partition of nodes into routes, the vehicle index assigned to each route is arbitrary. Therefore, in branching over the assignment decisions, an aggregate decision variable is introduced, $z^s_i=\sum_{k\in K} y^s_{ik}$, which equals 1 if node i is assigned to schedule s and 0 otherwise. This limits the number of variables over which branching is performed.

The branching order is an important decision. When setting the value of a variable to either zero or one, better feasible solutions are likely if the variable is set to its "correct" value, that is, its value in the optimal solution. This is of particular importance in a depth-first traversal of the branch-and-bound tree, where variables considered early in the branching order retain their values for a long time. Hence, it is important to branch first on those variables with values that can be speculated with greater confidence.

In particular, the frequency with which the value of each y-variable is 1 in a feasible solution to the whole problem is recorded in the Lagrangian phase. From the frequency of the y-variables we derive the frequency of the z-variables to determine the branching order. Recall that the subproblems are solved in each iteration, each time for a different set of Lagrange multipliers; therefore, a large number of y-variable solutions are used in this statistic.

At each node, an attempt is made to find a good feasible solution using the randomized savings heuristic described in the appendix. A lower bound for each node is obtained by solving two subproblems similar to subproblems (5) and (7) used in the Lagrangian, except that additional constraints are imposed by the fixed z-variables. We introduce additional notation for these subproblems. Let L_1 be the set of all pairs of indices (i, s) that have been fixed at the current node by assignments of $z_i^s = 1$ (including those assignments at preceding nodes in the depth-first traversal of the branch-and-bound tree). Similarly, let L_0 be the set of pairs of indices (i, s) fixed by assignments of $z_i^s = 0$.

The y-subproblem at a given node of the tree is

$$\min \sum_{s\in S}\sum_{k\in K}\sum_{i\in N}(\gamma^s\tau_i^s - w_i\alpha^s)y_{ik}^s \tag{8a}$$

subject to

$$\sum_{s\in S}\sum_{k\in K}\gamma^s y_{ik}^s \geq f_i \quad \forall\, i\in N \tag{8b}$$

$$\sum_{s\in S}\sum_{k\in K} y_{ik}^s \leq 1 \quad \forall\, i\in N \tag{8c}$$

$$\sum_{s\in S}\sum_{i\in N}(\beta^s w_i)a_{su}y_{ik}^s \leq C \quad \forall\, k\in K;\ u\in U \tag{8d}$$

$$\sum_{k\in K} y_{ik}^s = z_i^s \quad \forall\,(i, s)\in L_0\cup L_1 \tag{8e}$$

$$y_{ik}^s \in \{0, 1\} \quad \forall\, i\in N;\ k\in K;\ s\in S. \tag{8f}$$

Constraints (8b), (8c), (8d), and (8f) are identical to those in (5). Constraints (8e) are added for all (i, s) combinations whose z-value is fixed.

The x-subproblem at a given node of the tree is

$$\min \sum_{k\in K}\sum_{u\in U}\sum_{(i,j)\in A}\gamma^u t_{ij}x_{ijk}^u \tag{9a}$$

subject to

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{j\in N\cup\{0\}} x_{jik}^u \quad \forall\, i\in N\cup\{0\};\ k\in K;\ u\in U \tag{9b}$$

$$\sum_{i, j\in Q} x_{ijk}^u \leq |Q| - 1 \quad \forall\, Q\subseteq N;\ k\in K;\ u\in U \tag{9c}$$

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u \geq a_{su}z_i^s \quad \forall\,(i, s)\in L_1,\ u\in U \tag{9d}$$

$$x_{ijk}^u \in \{0, 1\} \quad \forall\,(i, j)\in A;\ k\in K;\ u\in U. \tag{9e}$$

Constraints (9b) and (9c) are the flow-balance and subtour elimination constraints, respectively. Constraints (9d) ensure that at least one vehicle is routed to node i, if i has a fixed schedule choice s, for every unique delivery day u such that $a_{su} = 1$. This constraint will never result in choosing more than one vehicle because we are minimizing (9a). Note that this formulation decomposes by u, but not by k.

The following lemmas show that the sum of the objective values of subproblems (8) and (9) constitutes a lower bound at the node.

Lemma 2. *Given the fixed z values at a given node in the branch-and-bound tree, the solution to* (9a)–(9e) *is a lower bound on the routing cost*

$$\sum_{k\in K}\sum_{u\in U}\sum_{(i,j)\in A}\gamma^u t_{ij}x_{ijk}^u$$

corresponding to the optimal solution of (3) *with additional constraints on the x and y values imposed by the fixed z values.*

Proof. Consider formulation (3) with additional constraints fixing x and y to match the fixed z values. Constraints equivalent to (3b)–(3d) and (3h) do not appear in (9a)–(9e); and constraints (9b), (9c), and (9e) are identical to (3f), (3g), and (3i), respectively. Therefore, it remains to show that constraints (9d) are relaxations of (3e). Recall that constraints (3e) are

$$\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{s\in S} a_{su}y_{ik}^s \quad \forall\, i\in N;\ k\in K;\ u\in U.$$

Weaker constraints are derived by summing over all vehicles:

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{s\in S} a_{su}\sum_{k\in K} y_{ik}^s \quad \forall\, i\in N;\ u\in U.$$

Using the definition of $z_i^s = \sum_{k\in K} y_{ik}^s$, we have

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{s\in S} a_{su}z_i^s \quad \forall\, i\in N;\ u\in U.$$

Equivalently, we can exclude from the summation on the right-hand side any schedule s for which $z_i^s \notin L_1$, because minimizing (3) will set to 1 only those x values that are forced to 1 by (3e). We rewrite the constraints as follows:

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u = \sum_{s\in S:(i,s)\in L_1} a_{su}z_i^s \quad \forall\, i\in N;\ u\in U.$$

Because only one schedule can be chosen for a node, we have

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u = a_{su}z_i^s \quad \forall\,(i, s)\in L_1;\ u\in U.$$

Constraints (9d) are relaxations of the above, which have been shown to be a relaxation of (3e). Hence, all constraints of (3) are either equivalent or relaxed in (9a)–(9e), and the objective function is a lower bound on the optimal routing cost. □

Given the fixed z values, it is possible to strengthen the lower bound obtained from (9a)–(9e) with additional cuts. Let L_0^* be a subset of L_0 containing (i, s) pairs such that $z_i^s = 0$ for some $s \in S\backslash\{|S|\}$ and visiting on schedule $|S|$ has been ruled out for these nodes. Thus, node i is excluded from the corresponding unique delivery days. Note by definition that $(i, |S|)$ can never be in L_0^*. Formally, we have

$$L_0^* = \{(i, s) \in L_0\colon s \in S\backslash\{|S|\},\ i\colon \{(i, |S|) \in L_0\}\}.$$

Constraints similar to (9d) set to 0 specific routing variables corresponding to unique delivery days that have been ruled out by fixed z values as follows:

$$\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u \le a_{su} z_i^s \qquad \forall\,(i, s) \in L_0^*,\ u \in \{u\colon (i, s) \in L_0^*,\ a_{su} = 1\}. \tag{9f}$$

These cuts are weaker than constraints (3e). This can be demonstrated in a manner similar to the discussion of constraints (9d) in the proof of Lemma 2.

The following constraints ensure that all nodes are included in some route:

$$\sum_{u\in U}\sum_{k\in K}\sum_{j\in N\cup\{0\}} x_{ijk}^u \ge 1 \quad \forall\, i \in N. \tag{9g}$$

Again, (9g) is weaker than (3e) because it is equivalent to the constraints obtained by summing (3e) over all vehicles k and delivery days u. With (9g) added, the problem can no longer be decomposed by unique delivery day $u \in U$.

Constraints can be added to guide the routing subproblem toward capacity-feasible routes:

$$\sum_{i\in N}\sum_{j\in N\cup\{0\}} \hat{\beta}_i^u w_i x_{ijk}^u - C\sum_{j\in N} x_{0jk}^u \le 0 \quad \forall\, k \in K,\ u \in U, \tag{9h}$$

where $\hat{\beta}_i^u$ is a demand accumulation factor used to adjust the demand accumulated at each served node according to the chosen level of service. We define

$$\hat{\beta}_i^u = \begin{cases} \beta^s & \text{if } a_{su} = 1, \text{ and either } z_i^s = 1,\ s \in S \\ & \text{or } z_i^{|S|} = 0,\ s \in S\backslash\{|S|\} \\ \beta^{|S|} & \text{otherwise.} \end{cases}$$

If a particular node i has been assigned to a specific schedule (i.e., $z_i^s = 1$), then the demand accumulation adjustment β^s corresponding to the schedule s is used. Note from (2) that for a given delivery day u, only one s can satisfy $a_{su} = 1$, $s \in S\backslash\{|S|\}$. Otherwise, the least possible accumulation, corresponding to $|S|$, is used. Constraints (9h) become stronger with greater depth of traversal down the tree, ensuring capacity-feasible routes at the leaves. Constraints (9h) ensure that the depot is included in all routes, even if all subtour elimination constraints are not added.

Lemma 3. *Given the fixed z values at a given node in the branch-and-bound tree, the solution to* (9a)–(9h) *is a lower bound on the routing cost*

$$\sum_{k\in K}\sum_{u\in U}\sum_{(i,j)\in A} \gamma^u t_{ij} x_{ijk}^u$$

corresponding to the optimal solution of (3) *with additional constraints on the x and y values imposed by the fixed z values.*

Proof. The constraints (9f)–(9h) have been shown to be either equivalent to or weaker than constraints (3e). Hence, the additionally constrained problem is still a lower bound. □

If the sum of the objective values of (8) and (9) is larger than the best known upper bound, the entire subtree emanating from that node is truncated. As in the Lagrangian phase, the assignment subproblem is solved optimally, while the routing subproblem may terminate with a lower bound rather than an optimal solution. Note, however, that further down the branch-and-bound tree, more z-variables become fixed, and both subproblems become more restricted.

When reaching a node that is a leaf of the tree, the time limit is removed to find the optimal solution of the routing subproblem. This is necessary to guarantee the optimality of the resulting solution at the end of branch and bound. Fortunately, at the leaf of the tree all z-variables are fixed; therefore, the routing subproblem with the added set of aggregated constraints includes many cuts. Hence, fewer additional cuts must be added after initially solving the relaxed subproblem.

2.5. A Related Heuristic Approach

For small and medium problem instances, the combination of the Lagrangian and branch-and-bound algorithms yields an optimal solution in a reasonable amount of time, as shown in §3. However, the PVRP-SC is a complex combinatorial problem that is a generalization of several other NP-hard problems such as the TSP and the VRP. It is not realistic to expect that arbitrarily large instances may be solved to optimality. In this section, the exact approach is modified to provide good heuristic solutions that apply most of the useful elements described in the previous subsections.

As expected, the routing subproblems are the major bottlenecks in the solution method. To reduce solution times, routing subproblems in the Lagrangian phase are not necessarily solved to optimality. The magnitude of deviation from optimality can be controlled through the "time budget" stopping rule discussed in §2.2. With a larger budget, more cuts are likely to be found that could improve the solution. When performing the upper-bound heuristic to find a feasible solution, it may be desirable to limit the time spent on each iteration. This must be balanced with the resulting reduction in solution quality. Similar considerations arise in the branch-and-bound phase, in which both the relaxation and the heuristic are performed. The time spent at the leaves could be restricted too. The total solution time may be controlled by limiting the use of the upper-bound heuristic to a subset of Lagrangian iterations and/or branch-and-bound nodes.

The extent to which each of the above control parameters may be used is an algorithmic design issue. These choices depend on the problem instance size and the amount of time available to search for a solution. Roughly speaking, the more time spent on finding good feasible solutions, the better the quality of the solution will be. However, it is also important to develop a good understanding of where it is most useful to spend the available time. In §3.1, we illustrate the effect of controlling the amount of time per iteration on solution quality.

Finally, we may choose to terminate the branch-and-bound algorithm with a solution within some δ percentage of the optimal value. In this case, a node is fathomed if the lower bound at the node is greater than $(1-\delta)$ of the upper bound.

3. Numerical Study

In this section, we examine the impact of service choice on period vehicle routing problems and the effectiveness and robustness of the solution method for the PVRP-SC. Two data sets are included in this study. The first set, described in §3.1, is drawn from the application that motivated the development of the PVRP-SC and the second set, described in §3.2, is taken from the PVRP literature. Section 3.3 presents insights on the impact of service choice on operations.

The solution algorithm is implemented using C with the CPLEX Callable Library Interface and the CPLEX 8.1 solver, running on a Sun Fire v250 1.28-GHz UltraSPARC IIIi computer with two processors. The subproblems are modeled as described in §2. For larger problem instances, constraints (9g) are not included due to computational limitations.

3.1. Motivating Problem

The North Suburban Library System (NSLS) is a state-funded library system that delivers interlibrary loan items (books, video cassettes, etc.) to its member libraries in the suburbs north of Chicago. The 50 public libraries served by NSLS account for 89% of the total demand. These libraries, along with information

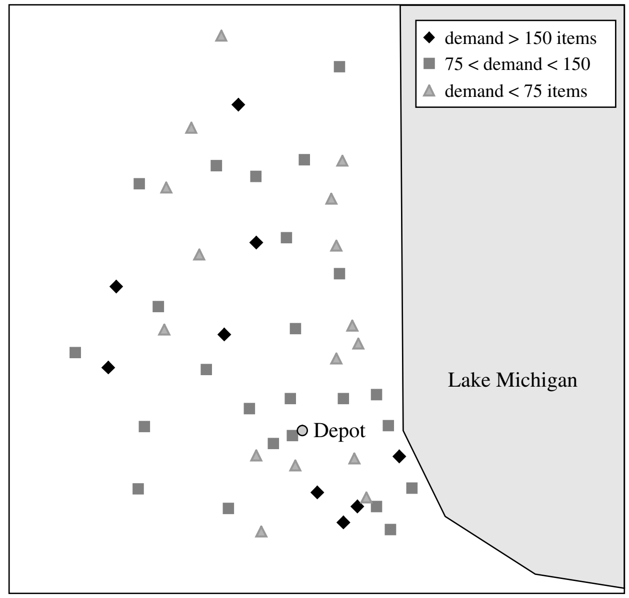

Figure 2 **Geographic Distribution of Libraries in NSLS Service**

on their daily volume of loan items, are shown in Figure 2. Four vans operate from a sorting facility (the depot) and visit the libraries to pick up outgoing items and deliver incoming items. At the end of the work day, the items that have been picked up are sorted at the depot for delivery to their destination libraries when they are next visited. NSLS would like to provide the greatest possible visit frequency to its members. However, due to budgetary restrictions and rising demand for service from libraries, NSLS may reduce frequencies and design routes in a way that takes into account routing efficiency as well as the demand of the libraries.

3.1.1. Solution Method Performance. We examine the sensitivity of the solution method to the structure of problem instances. We construct subsets of nodes randomly drawn from the 50 public libraries, ranging in size from 12 to 44 nodes. Each subset is tested with three and four vehicles, and capacities at an additional 10%, 20%, 30%, and 40% above the base capacity, which is adjusted by the number of nodes.

Figure 3 depicts the three options for implementation of the solution method: terminating after Lagrangian relaxation (LR); continuing with exact branch and bound (B&B); and continuing with heuristic branch and bound. Lagrangian relaxation finishes with a gap, G_1, between the feasible solution, Z^{LR}, and the final lower bound, LB^{LR}. In the exact variation, branch and bound runs until an optimal solution is found; in the heuristic variation, branch and bound runs until a solution within δ% of the optimal is found. While it is preferable to use one of these two options, this may not be realistic for large instances. We use small- to medium-size instances that can be

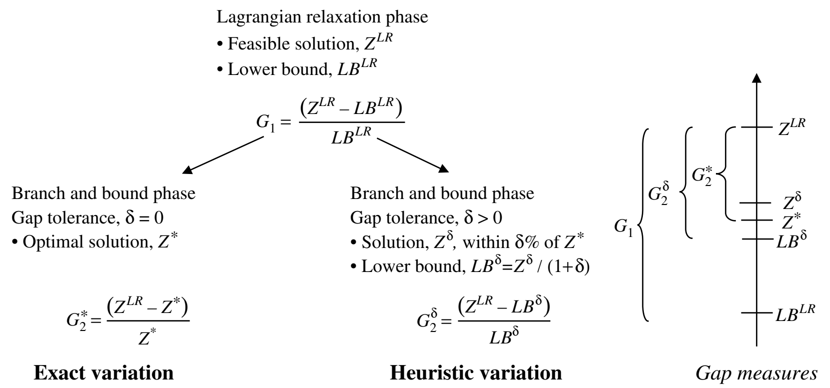

Figure 3 Solution Approach

solved to optimality (or δ-optimality) to estimate the gap that can be expected from the Lagrangian solution for larger instances. If the exact solution is available, the quality of Z^{LR} is assessed relative to Z^* with G_2^*. If a solution with δ-optimality is available, the quality of Z^{LR} is assessed relative to a lower bound on the optimal solution, $LB^\delta = Z^\delta/(1+\delta)$, with a gap G_2^δ.

The performance metrics G_1 and G_2^* are used to analyze the Lagrangian solutions for the NSLS test cases, shown in Table 1. Each test case is represented by one cell. The top value in the cell is G_1 and the bottom value is G_2^*. Cells without G_2^* values indicate that branch and bound did not terminate within the eight-hour limit.

Three clear trends emerge when evaluating the quality of Z^{LR} with G_1. First, the average gap usually increases with the number of nodes; with both three and four vehicles, the value of G_1 for 44 nodes is two to three times the value for 12 nodes. With more nodes, the routing subproblems (7) are more difficult to solve within the time budget, which results in weaker lower bounds. If the solutions to (7) contain subtours, corresponding upper bounds cannot be obtained, and feasible solutions do not improve. Second, the average gap decreases with additional capacity. More capacity increases the likelihood of finding feasible solutions to update upper bounds. Finally, the average gap decreases with the number of vehicles. The number of vehicles impacts the flexibility of the system to service the nodes (the formulation of the PVRP-SC does not include a fixed vehicle cost).

These trends are less clear when evaluating the quality of Z^{LR} relative to the optimal solution with the values of G_2^*. This is to be expected because the above trends impact both upper and lower bounds in the LR phase. In general, the LR solution is within 10% of the optimal solution with relatively large variations among the instances. The number of nodes impacts which test cases can be solved to optimality. The exact branch-and-bound method cannot be completed

Table 1 NSLS Data Set Results: Exact Solution Method

	Additional capacity (%)				
$\lvert N\rvert$	10	20	30	40	Averages (%)
(a) Problem instances with four vehicles					
12	13.8	13.8	8.1	9.6	11.3
	3.6	4.4	4.8	7.3	5.0
16	15.7	15.6	11.8	10.7	13.5
	2.3	14.8	11.8	1.9	7.7
20	22.2	22.1	16.0	14.7	18.7
	3.6	6.5	3.2	14.7	7.0
28	29.3	28.5	19.9	15.9	23.4
	22.7	2.8	5.4	8.8	9.9
36	26.4	23.9	16.4	16.3	20.7
	7.0	20.4	5.9	11.4	11.2
40	34.6	32.7	23.0	15.7	26.5
	16.4	5.9	7.8	3.6	8.4
44	36.4	36.1	30.8	23.2	31.6
	21.2	—	11.8	5.3	12.8
Averages	25.5	24.7	18.0	15.2	20.8
	10.9	9.2	7.3	7.6	8.7
(b) Problem instances with three vehicles					
12	17.9	16.9	11.2	10.7	14.2
	9.1	2.6	10.2	7.3	7.3
16	19.3	18.0	12.5	10.7	15.1
	15.7	4.9	4.8	10.1	8.9
20	26.7	25.0	16.3	19.5	21.9
	5.9	9.4	14.7	2.7	8.2
28	35.1	30.5	21.8	24.4	27.9
	27.9	14.3	5.3	4.0	12.9
36	31.4	31.4	20.1	17.9	25.2
	10.1	24.0	2.3	9.8	11.6
40	45.6	44.2	26.6	19.4	33.9
	—	—	15.2	18.4	16.8
44	39.3	39.2	32.4	29.8	35.2
	—	—	—	4.7	4.7
Averages	30.8	29.3	20.1	18.9	24.8
	13.8	11.0	8.8	8.2	10.4

Note. Test cases without values for G_2^* could not be solved with the exact branch and bound.
Legend. $G_1 = (Z^{LR} - LB^{LR})/LB^{LR}$. $G_2^* = (Z^{LR} - Z^*)/Z^*$.

Table 2 NSLS Data Set Results: Heuristic Solution Method

	Additional capacity (%)			
$\|N\|$	10	20	30	40
(a) Problem instances with four vehicles				
44		7.2		
(b) Problem instances with three vehicles				
40	22.0	10.6		
44	15.9	14.4	16.0	

Legend. $G_2^\delta = (Z_{LR} - LB^\delta)/LB^\delta$. $LB^\delta = Z^\delta/(1+\delta)$.

within eight hours for some cases over 36 nodes. For these cases, Table 2 presents the values of G_2^δ with $\delta = 0.02$. The other cases are not shown, although all cases are solved with the heuristic method as well. By definition, the difference between Z^* and Z^δ is bounded by 2%, and the average difference observed over all test cases is 1.8%. The average difference between G_2^* and G_2^δ for the cases not shown is 0.2%, and never more than 1%. Table 2 indicates that for larger problem instances, the Lagrangian solutions may be as much as 22% from optimality. As discussed next, the advantage of the LR phase is its speed.

Figure 4 presents the solution times for the LR phase and the exact and heuristic B&B phases. Each data point represents the average of all test cases of that size. Solution times are also presented for the B&B phase without running the LR phase first. In this case, we begin with the initial upper bound, and branch in increasing order of variable indices. Results could not be obtained with B&B within the time limit for test cases of 28 or more nodes without the LR phase. With the two-phase approach, the LR phase significantly reduces the gap before performing B&B. Further, the branching scheme in B&B benefits from results of the LR phase. These factors explain the dramatic differences in solution times.

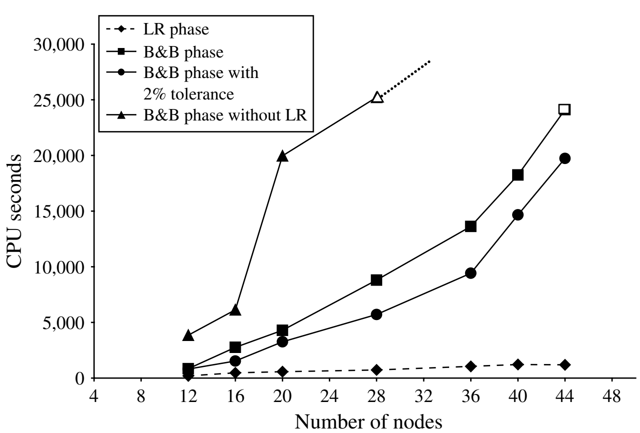

Figure 4 Comparison of Time Spent (Average Over Multiple Test Cases)

As expected, using the heuristic version of the B&B phase requires less computational time than the exact version and, by definition, the corresponding difference in the solution is bounded by 2%. This variation is capable of solving test cases with more nodes than the exact approach.

Recall that solution times for the routing subproblems in the LR and B&B phases can be controlled by a time budget. We impose a maximum limit of 300 CPU seconds for each subproblem in both phases. This limit is determined empirically to allow a majority of the subproblem iterations to run to completion, while terminating more difficult subproblem iterations at the end of the time limit. As a result, the LR and B&B phases finish in a reasonable amount of total time. With a budget of 300 CPU seconds, for the larger test cases (≥40 node), 20% of the routing subproblems are truncated by the time budget before all necessary subtour elimination cuts are added and the upper-bound heuristic cannot be applied at those iterations. With a 50 CPU-second time budget, 76% of the routing subproblems for these test cases are truncated. As a result, the average remaining gaps after LR rise by about 14%. With a 500 CPU-second time budget, 20% of the routing subproblems are truncated, which is the same fraction obtained with 300 CPU-seconds. However, the average solution time rises by a minimum of 36% for all test cases.

3.1.2. Impact of Service Choice. Currently, NSLS serves the 50 public libraries daily, including libraries in remote regions with low demand, as shown in Figure 2. Libraries are allocated to routes in an ad hoc manner as demands increase and driver hours change. The sequencing of visits within a route is left to the discretion of the driver. We compare the following three delivery options relative to the current operations:

1. Maintain daily visits to all libraries and optimize routes,
2. Reduce visit frequency based on demand volumes and optimize routes, and
3. Determine visit frequency based on demand, routing efficiency, and service benefit, and optimize routes.

Option 1 is modeled as a VRP, whose solution is repeated every day of the week. For Options 2 and 3, libraries may be served by one of the following schedules: $\{(Mon, Wed, Fri); (Tue, Thr); daily\}$. This set satisfies the condition of Lemma 1. The visit frequency assigned to each library for Option 2 (and the minimum frequency for Option 3) is based on daily demand, to ensure that nodes with larger demands receive more frequent service. Option 2 is modeled as a PVRP; Option 3 is modeled as a PVRP-SC. The service benefit values, α^s, $s \in S$, provide incentive

Table 3 Solution Results for NSLS Delivery Options

Option	Routing cost	Service benefit	Total objective
0. Current operations	6,315	−1,642	4,673
1. Optimized daily service	5,849	−1,642	4,208
2. Reduced service PVRP	5,751	−1,243	4,507
3. PVRP-SC*	5,425	−1,492	3,933

*Solution obtained with $\delta = 0.02$; solution is within 2% of the optimal solution.

to increase frequency while balancing routing costs: $\alpha^1 = 0.15$, $\alpha^2 = 0.1$, and $\alpha^3 = 0.2$.

Table 3 displays the routing costs, service benefit, and total costs for the three options and the current operations. Comparing Option 1 to the current operations, there is an initial 7% reduction in routing costs from optimizing route configurations. Routing costs decrease further as the daily visitation requirements are relaxed and more flexibility is introduced. However, comparing Options 1 and 2, routing costs decrease by an additional 2%, but there is a 24% loss in service benefit, leading to a higher total objective. This shows that the traditional PVRP may not provide sufficient flexibility to compensate for the reduced service benefit. This is remedied with the PVRP-SC. Allowing flexibility in visit frequency leads to a 16% improvement when comparing Option 3 with the current operations with respect to total costs, and a 7% improvement over Option 1. With the PVRP-SC, the loss in service benefit is smaller, and is offset by lower routing costs. The routing cost of the PVRP-SC solution is 14% less than that of the base case and 7% less than that of the VRP solution. For a budget-constrained agency, this improvement is quite significant.

3.2. Test Cases from the Literature

We apply service choice to a 50-node PVRP test case from Christofides and Beasley (1984). Test case 50b, unlike other 50-node cases in that reference, includes demand-dependent preset frequencies $f_i = 1$ for $w_i \leq 10$ cwt (500 kgs); $f_i = 2$ for $10 < w_i \leq 25$; and $f_i = 3$ for $w_i > 25$. The period is three days, and there are three vehicles each with capacity of 8,000 kgs. Figure 5 shows the geographic distribution of the nodes. Note that nodes appear to be uniformly distributed in a square.

Three scenarios are derived from 50b to test the impact of service choice. Clearly introducing service choice can improve results, yet the magnitude of the savings is less clear. Case I is the standard PVRP; Case II is the PVRP-SC when service benefit is not considered in the objective, but changes in visit frequency are allowed to increase routing efficiency; and Case III is the PVRP-SC with service benefit considered. The preset frequencies are treated as minimum

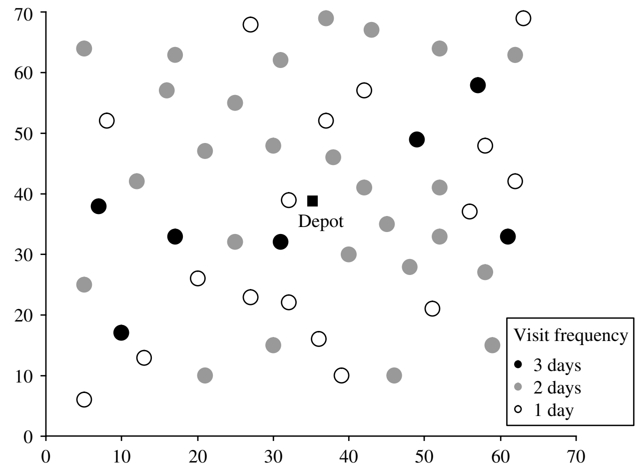

Figure 5 Geographic Distribution of Libraries in Christofides and Beasley (1984) 50b

frequencies in the PVRP-SC. Case II shows the impact of service choice on routing efficiency, and Case III shows the impact on a balance of routing efficiency and service level.

Schedules are limited to $\{(Mon, Wed); (Tue); (Mon, Tue, Wed)\}$. Note that this represents a smaller set of schedules than allowed in 50b, which is originally solved over five days. The demand accumulation adjustments are: $\beta^1 = 2$, $\beta^2 = 3$, and $\beta^3 = 1$. The service frequencies are: $\gamma^1 = 2$, $\gamma^2 = 1$, and $\gamma^3 = 3$. The node stopping costs are set to 0. The service-level benefit is set to 0 for all schedules in Cases I and II. In Case III, $\alpha^1 = 0.2$, $\alpha^2 = 0.1$, and $\alpha^3 = 0.3$.

The value of the objective function for Case I is an upper bound on the objective function for Case II because the visit constraints are relaxed in Case II. Similarly, the value of the routing component of the optimal solution for Case II is a lower bound for the optimal solution for Case III. These bounds can be used in the solution method to improve solution times. The results in Table 4 are obtained with the heuristic variation with $\delta = 0.02$. The service benefit, although not part of the objective function for Cases I and II, is added back in Table 4 for comparison.

The results show that in Case II nodes are served with higher frequency, and this flexibility leads to a slight improvement in routing cost. A much larger change in the total objective value is observed in

Table 4 Test Case Results for 50b

Instance	Routing cost	Service benefit	Total objective
Case I	998.8	−2,765	−1,766
Case II	972.7	−2,803	−1,830
Case III*	1,326.9	−4,065	−2,738

*Stopped with less than 640 nodes remaining (out of 2^{151}) due to computational constraints.

Case III when the service term is given a positive weight. Compared to Case I, the total objective improves by 55%, with a corresponding rise of 32% in the routing cost.

3.3. Discussion

An analysis of the results from the previous sections leads to general insights regarding the benefits of service choice in the PVRP-SC. As mentioned in the introduction, adding service choice can improve overall system performance due to increased routing efficiency and/or raised service levels for some customers. The test cases developed from NSLS and Christofides and Beasley (1984) clearly show significant improvement in overall system performance when service benefit is included. The contribution of routing cost savings to this overall improvement is less clear. With NSLS, improvements in both routing and service are observed; however, with 50b, the improvement comes strictly from service, and routing costs increase. To explore this issue further, we use the simple example in Figure 6.

In the example, the depot and three customer nodes are located at one-mile intervals along a road segment. We ignore the service benefit and stopping costs, and focus on the routing cost savings from moving Node B (with one-day minimum service) to a two-day schedule. The solution in the figure shows how the relative position of the nodes affects routing savings due to service choice. In Figure 6(a), Node C, located furthest from the depot, has a three-day visit requirement. The total roundtrip distance of six miles must be covered on all days. Hence, we are indifferent between serving Node B on Tuesday or Monday/Wednesday. In this case, introducing service choice does not impact routing cost. Alternatively, in Figure 6(b), Node A, located closest to the depot, has a three-day visit requirement. In this case, adding service choice can reduce routing costs. We choose to serve Node B on a Monday/Wednesday route because this route passes by Node B on the way to Node C.

This example shows that the geographic distribution of the highest minimum frequency nodes impacts the extent to which service choice can reduce the travel distances between remote nodes and the depot. In particular, the potential for savings is greater when high-frequency nodes are closer to the depot than lower-frequency nodes. These observations can be extended to two-dimensional cases with many nodes. If there are a large number of randomly scattered nodes of different minimum frequencies, it is more likely that some highest minimum-frequency nodes will be located far from the depot. Thus, service choice may not reduce routing costs significantly, as with the Christofides and Beasley (1984) nodes. However, if nodes are distributed according to some pattern (such as highest minimum frequencies closer to the depot and lower minimum frequencies spread out), then service choice may reduce routing costs significantly. The NSLS nodes exhibit a mixture of these patterns. The highest minimum-frequency nodes are scattered randomly in the west region, clustered in the southeast region close to the depot (and nodes of other minimum frequencies are further), and nonexistent in the northeast region. Savings are achieved in the last two regions.

With the service benefit included, frequency flexibility improves the solutions in the examples in Figure 6(a) and (b), as higher benefits can be obtained by moving Node B to the two-day service level or by moving all nodes to the three-day service level. If there are significant stopping costs, then the term $(\gamma^s \tau_i^s - w_i \alpha_i^s)$ combined with routing costs determines if moves to greater service levels are desirable.

Because the PVRP-SC relaxes the frequency constraint of the PVRP, the cost of the PVRP-SC solution will always be lower than or equal to the cost of the PVRP solution. We can make the following observations about the cost differences.

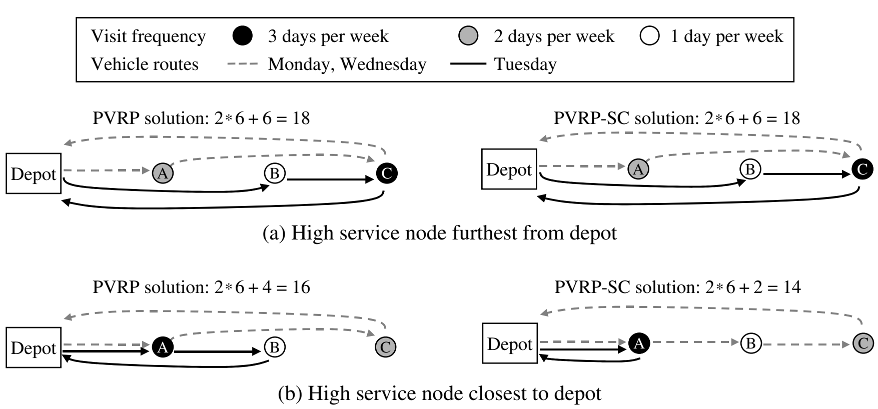

Figure 6 **Impact of Node Distribution on Routing Efficiency Due to Service Choice**

- If stopping costs and service benefits are ignored in the objective function, then the routing cost of the PVRP-SC solution will always be lower than or equal to the routing cost of the PVRP solution. The magnitude of the savings for any given problem instance depends on the geographic dispersion of nodes (in particular, nodes at the highest minimum frequency).
- If service benefit is considered, the routing cost may rise in the PVRP-SC if the increase is offset by an increase in service benefit. It is also possible for both routing costs and service benefit to improve.
- The introduction of stopping costs limits the routing cost savings. With higher stopping costs, increased frequencies above the minima may not be desirable.

4. Conclusions and Future Research

This paper introduces the period vehicle routing problem with service choice (PVRP-SC), which allows for service choice and accounts for the benefit of improved service. We model and formulate the PVRP-SC and develop an exact solution method with heuristic variations. The exact method presented here can serve as a benchmark for future heuristic methods. Computational tests on the motivating problem show improvements in routing efficiencies and in the overall performance of the system as a result of introducing service choice.

As this is the first presentation of the PVRP-SC, we envision several directions for future research, including generalizations of assumptions required for the formulation and exact solution method for instances of reasonable size. Heuristic solution methods may be used to solve problems resulting from the relaxation of these assumptions. Because restrictions such as disjoint schedules and conservative estimates of demand accumulation reduce the size of the feasible region, it may be possible to achieve better solution values with heuristics that relax these restrictions. For example, β^s is set according to the maximum number of days between visits to ensure feasibility. For a Tuesday-Thursday schedule, $\beta^s = 3$ because the maximum accumulation of three days occurs between Thursday and Tuesday, while the accumulation from Tuesday to Thursday is only two days. In the case of disjoint schedules and their union, and given the current definition of y^s_{ik}, this conservative approximation of β^s is justified because a vehicle performs the same route each day within a disjoint schedule; hence, no solutions are precluded. However, in a more general case with overlapping schedules, nodes may be visited by different vehicles. An accumulation parameter defined by schedule and day in this case, β^{sd}, would represent accumulation precisely and allow for a richer set of solutions.

In some practical applications of the PVRP-SC, it may be desirable to set route length constraints, for example, for driver shifts. Route length is determined by two factors—travel time, which is a function of the x variables, and stopping cost, which is a function of the y variables. A route-length constraint adds a second constraint linking the x and y variables, and can be written as

$$\sum_{(i,j)\in A} t_{ij} x^u_{ijk} + \sum_{s\in S}\sum_{i\in N} a_{su}\tau^s_i y^s_{ik} \le \Gamma \quad k \in K;\ u \in U.$$

To apply Lagrangian relaxation to decompose the problem, both linking constraints would need to be relaxed.

Acknowledgments

The authors would like to thank Jean-François Cordeau for his help in translating the data sets. This research has been supported by Grant DMI-0348622 from the National Science Foundation, the Alfred P. Sloan Foundation, and the Transportation Center of Northwestern University.

Appendix

Because the routing subproblem is time consuming to solve to optimality, a heuristic is used to solve the routing subproblem. A similar heuristic, proposed by Daskin (2004), is based on a randomization of the Clarke and Wright (1964) algorithm and modified to solve the TSP. Briefly, the Clarke-Wright algorithm has the following steps:

Step 1. Begin by putting each node into its own tour.

Step 2. Compute a savings list $S = \{s_{ij}, \forall i, j \in N\}$ where the savings obtained by merging the routes between nodes i and j is $s_{ij} = t_{i0} + t_{j0} - t_{ij}$. Sort the savings list in descending order.

Step 3. Select the best saving s_{ij} from the list. Merge the routes, removing from the solution arcs $(i, 0)$ and $(0, j)$ and adding arc (i, j). Update the savings list, $S \leftarrow S\backslash\{s_{ij}\}$.

Step 4. Repeat Step 3 until all nodes are contained within one route.

To escape the greedy nature of the heuristic, randomization is introduced in Step 3. Instead of picking the best possible saving, we pick randomly from among the top ϕ savings in the list. The rest of the algorithm proceeds unchanged. The whole algorithm is repeated a number of times and the best set of routes obtained is used. A similar heuristic has been used to solve the IRP with satellite facilities by Bard et al. (1998).

In our implementation of randomized Clarke-Wright, the algorithm is repeated for 300 iterations (based on the total amount of time allocated to find a heuristic VRP solution) and we use $\phi = 5$. At the end of each iteration, we also perform 2-opt and Or-opt on each route obtained.

References

Anily, S., J. Bramel. 1998. Vehicle routing and the supply chain. R. G. S. Tayur, M. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Norwell, MA.

Anily, S., A. Federgruen. 1990. One warehouse multiple retailer systems with vehicle routing costs. *Management Sci.* **36**(1) 92–114.

Bard, J., L. Huang, P. Jaillet, M. Dror. 1998. A decomposition approach to the inventory routing problem with satellite facilities. *Transportation Sci.* **32** 189–203.

Beltrami, E., L. Bodin. 1974. Networks and vehicle routing for municipal waste collection. *Networks* **4** 65–94.

Chan, L., A. Federgruen, D. Simchi-Levi. 1998. Probabilistic analysis and practical algorithms for inventory-routing models. *Oper. Res.* **46**(1) 96–106.

Chao, I.-M., B. Golden, E. Wasil. 1995. An improved heuristic for the period vehicle routing problem. *Networks* **26** 25–44.

Christofides, N., J. Beasley. 1984. The period routing problem. *Networks* **14** 237–256.

Clarke, G., J. Wright. 1964. Scheduling of vehicles from a central depot to a number of delivery points. *Oper. Res.* **12** 568–581.

Cordeau, J.-F., M. Gendreau, G. Laporte. 1997. A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks* **30** 105–119.

Daskin, M. S. 2004. Heuristics for the vehicle routing problem. Personal communication (February).

Federgruen, A., D. Simchi-Levi. 1995. Analysis of vehicle routing and inventory-routing problems. *Handbooks in Operations Research and Management Science*, Vol. 8. Elsevier Science B.V., Amsterdam, The Netherlands, 297–373.

Feillet, D., P. Dejax, M. Gendreau. 2003. Traveling salesman problems with profits. *Transportation Sci.* **39**(2) 188–205.

Fisher, M. L. 1981. The Lagrangian relaxation method for solving integer programming problems. *Management Sci.* **27**(1) 1–18.

Fisher, M. L. 1985. An application oriented guide to Lagrangian relaxation. *Interfaces* **15**(2) 2–21.

Fisher, M. L., R. Jaikumar. 1981. A generalized assignment heuristic for vehicle routing problems. *Networks* **11** 109–124.

Gaudioso, M., G. Paletta. 1992. A heuristic for the period vehicle routing problem. *Transportation Sci.* **26**(2) 86–92.

Gaur, V., M. Fisher. 2004. A periodic inventory routing problem at a supermarket chain. *Oper. Res.* **52**(6) 813–822.

Herer, Y., M. Tzur, O. Setty. 2003. An inventory routing problem with dynamic deterministic demand. Working paper, Industrial Engineering Department, Tel Aviv University, Israel.

Kim, J.-U., Y.-D. Kim. 2000. A Lagrangian relaxation approach to multi-period inventory/distribution planning. *J. Oper. Res. Soc.* **51** 364–370.

Kleywegt, A., V. Nori, M. Savelsbergh. 2002. The stochastic inventory routing problem with direct deliveries. *Transportation Sci.* **36**(1) 94–118.

Lenstra, J. K., A. Rinnooy Kan. 1981. Complexity of vehicle routing and scheduling problems. *Networks* **11** 221–227.

Newman, A. M., C. A. Yano, P. M. Kaminsky. 2005. Third party logistics planning with routing and inventory costs. J. Geunes, P. M. Pardalos, eds. *Supply Chain Optimization*. Springer-Verlag, New York, 87–122.

Russell, R., D. Gribbin. 1991. A multiphase approach to the period routing problem. *Networks* **21** 747–765.

Russell, R., W. Igo. 1979. An assignment routing problem. *Networks* **9** 1–17.

Tan, C., J. Beasley. 1984. A heuristic algorithm for the period vehicle routing problem. *Omega* **12**(5) 497–504.