TWO-ECHELON DISTRIBUTION SYSTEMS WITH VEHICLE ROUTING COSTS AND CENTRAL INVENTORIES

S. ANILY

Tel-Aviv University, Ramat-Aviv, Israel

A. FEDERGRUEN

Columbia University, New York, New York

(Received December 1990; revisions received December 1991, January 1992; accepted February 1992)

We consider distribution systems with a single depot and many retailers each of which faces external demands for a single item that occurs at a specific deterministic demand rate. All stock enters the systems through the depot where it can be stored and then picked up and distributed to the retailers by a fleet of vehicles, combining deliveries into efficient routes. We extend earlier methods for obtaining low complexity lower bounds and heuristics for systems without central stock. We show under mild probabilistic assumptions that the generated solutions and bounds come asymptotically within a few percentage points of optimality (within the considered class of strategies). A numerical study exhibits the performance of these heuristics and bounds for problems of moderate size.

We consider distribution systems with a single depot and many retailers with external demands for a single item that occur at a specific constant (but retailer-dependent) deterministic rate. The depot places orders with an outside supplier. Goods are distributed from the depot to the retailers by a fleet of identical vehicles, combining deliveries into efficient routes.

In an earlier paper, Anily and Federgruen (1990a) analyze a model where the depot serves as a mere coordinator of the replenishment process or alternatively as a transshipment point in which no inventory can be kept. In such systems, one has to determine replenishment policies for all retailers, as well as matching efficient routing patterns. In this paper, we extend the analysis to the case where central inventories may be kept in the warehouse. As a consequence, the above problems are compounded by that of determining a replenishment strategy for the warehouse, optimally coordinated with that of each retailer and synchronized with the transportation schedules.

We assume that at each outlet, customer demands occur at a constant, deterministic but *outlet specific* rate. These demand rates are assumed to be rational, so that after appropriate scaling they are even integers. An outlet may thus be viewed as the aggregate of an integer number of *demand points*, each of which faces a demand rate of two. Inventory carrying costs are incurred at a constant rate per unit of time, and per unit stored. (This rate is identical for all retailers, but is different at the warehouse.) The transportation costs include a fixed (leasing or renting) cost per route driven by one of the vehicles and variable costs proportional to the total (Euclidean) distance on all routes (but no unloading costs). As in most standard inventory models we assume that the cost of an order from the outside supplier is fixed-plus-linear.

The objective is to minimize the system-wide long-run inventory, transportation and order costs. We refer the reader to Anily (1987), and Anily and Federgruen (1990a, b) for a review of the literature on the vehicle routing problem (VRP) and models that integrate inventory allocation and vehicle routing problems in one-warehouse, multiretailer systems. More recent work includes Dror and Ball (1987), Chien, Balakrishnan and Wong (1989), and Gallego and Simchi-Levi (1990).

The classical single-warehouse multiretailer model assumes that each retailer is served on an *individual* basis, rather than deliveries being combined into efficient vehicle routes. Even with individual and uncoordinated deliveries, the structure of optimal policies can be very complex (Roundy 1985) making them unattractive even if their computation were tractable. The combinatorial nature of the routing costs compounds this complexity. We thus restrict ourselves, as

Subject classifications: Inventory/production: approximations and heuristics; multi-item, multi-echelon, multistage. Transportation: vehicle routing.
Area of review: DISTRIBUTION, TRANSPORTATION AND LOGISTICS (SPECIAL ISSUE ON STOCHASTIC AND DYNAMIC MODELS IN TRANSPORTATION).

0030-364X/93/4101-0037 $01.25

in Anily and Federgruen (1990a), to the class of replenishment strategies Φ with the following properties. A replenishment strategy specifies a collection of *regions* (subsets of outlets) covering all outlets: If an outlet belongs to several regions a specific fraction of its sales/operations is assigned to each of these regions. Each time one of the outlets in a given region receives a delivery, this delivery is made by a vehicle who visits (in an efficient sequence or *route*) all outlets in the region and none outside the region. We use the terms regions and routes interchangeably.

Note that a large amount of flexibility is preserved within the class Φ by allowing retailers to be assigned to several regions, i.e., by allowing regions to overlap. On the other hand, under a strategy in Φ, all regions are controlled independently of each other. Thus, if an outlet belongs to two regions, it is treated as two separate suboutlets each responsible for a specific fraction of the sales; it is therefore possible that a delivery is made to one suboutlet at an epoch at which the other suboutlet continues to have stock. However, our proposed heuristics generate regions in which only a *few* retailers are split among different (usually two) routes. See Dror and Trudeau (1990) for related work on split delivery routing.

Also, note that under strategies in Φ, outlets assigned to different regions are never served in a common route even though in an optimal strategy any given outlet may be served in varying rather than constant combinations of other outlets. This is illustrated by the example in Hall (1991). For further discussion regarding the merits of our restriction approach and a review of other joint replenishment problems for which a similar restriction has been employed, see Anily and Federgruen (1991c).

In Section 1 we present some notation and preliminaries. In Section 2 we develop lower bounds for *uncapacitated systems*, where only a bound on the sales volume per region prevails, as well as *capacitated systems*. In capacitated systems, we allow for additional upper bounds on the frequency with which the routes may be driven, possibly in combination with capacity bounds for the vehicles and/or bounds on the sales volumes per region. See Anily and Federgruen (1990a) for a discussion of these three types of regional constraints. In Section 3, we develop heuristic solution methods as well as upper bounds for the minimum cost and discuss their asymptotic complexity, optimality and accuracy gaps, respectively. Section 4 complements these with a numerical study conducted to gauge the performance of the heuristics and bounds for problems of moderate size. Section 5 summarizes our conclusions.

We show that the lower and upper bounds on the minimum, long-run average system-wide costs (among all strategies in Φ) as well as a heuristic solution may be computed in $O(N \log N)$ time only, with N the total number of demand points. Moreover, considering a general stochastic sequence of locations with the system's demand points located in the first N points of this sequence, we show that the lower and upper bounds come, almost surely, within 6% (and for uncapacitated systems within 2%) of each other for sufficiently large N. In addition, we briefly explain how our results can be extended to cases where backlogging is allowed if all retailers face identical demand rates. The same restriction is necessary in systems without central stock (Anily and Federgruen 1990a).

These results are all the more remarkable because the performance of Roundy's heuristics for the classical one-warehouse, multiretailer models with *individual, uncoordinated* deliveries, and hence, separable delivery costs, may deteriorate significantly when adapted to incorporate restrictions on delivery sizes; moreover, no (simple) modification of these heuristics with comparable worst-case performance seems to exist. The discrepancy between the positive results in our capacitated models and the apparent lack thereof in capacitated versions of the classical model is due to the restriction to the class Φ and different assumptions regarding the *structure* of the *delivery costs.*

1. NOTATION AND PRELIMINARIES

We use the same notation and assumptions as in Anily and Federgruen (1990a). Here we confine ourselves to the additionally required notation. Let:

h_0 = the inventory holding cost per unit of time, per unit stored at the warehouse;

K_0 = the fixed cost per order placed by the warehouse;

h = $h^+ - h_0$ denotes the *echelon holding cost* rate.

We assume that $h > 0$. Since the holding cost rate usually increases with the (cumulative) value added, this assumption is almost always satisfied. Also let,

$\theta_l = \mathrm{TSP}(X_l^0)$ = the length of an optimal traveling salesman tour through X_l and the depot.

The problem of finding an optimal strategy in Φ, with minimum cost $V^*(X)$ can be stated as a special case of the general problem of partitioning the set X of demand points into L regions with capacities $\{M_1^*, \ldots, M_L^*\}$ and a cost $U(\chi)$ assigned to each partition χ.

Problem P

$$\mathrm{Min}\{U(\chi)\colon \chi = \{X_1, \ldots, X_L\} \text{ and } m_l = |X_l| \leqslant M_l^*,\; l = 1, \ldots, L\}.$$

(The numbers M_l^* depend on vehicle capacity, regional sales volume and frequency constraints. See Anily and Federgruen 1990a.)

It is easy to see that **P** is NP-complete even in the simplest case where all cost components, except for the routing cost, are zero as the problem is then reduced to the well known vehicle routing problem. Even the latter cannot be solved to optimality for all but the smallest size problems.

Instead we concentrate on *heuristic* solution methods. For a given heuristic **H,** applied to the set X, let $V^H(X)$ denote the cost of the generated solution and define the *relative error* $e^H(X) = (V^H(X) - V^*(X))/V^*(X)$. If $X_{(N)}$ denotes the first N points of a randomly generated sequence $\{x_1, x_2, \ldots\}$, we call **H** asymptotically ϵ-optimal if $\lim_{N\to\infty} e^H(X_{(N)}) \leqslant \epsilon$, almost surely.

An important step in the design and evaluation of our heuristics is the derivation of a lower bound. The latter is obtained by replacing the cost function $U(\chi)$ by a lower bound cost structure such that the resulting partitioning problem is easy to solve. We refer the reader to Section 1 in Anily and Federgruen (1990a) for the definitions of consecutive and monotone partitions and extremal partitioning problems. If a partitioning problem is extremal, an optimal partition is obtained by an exceedingly simple, linear time procedure (the Extremal Partitioning Algorithm in Anily and Federgruen 1991a).

For a given partition $\{X_1, \ldots, X_L\}$ of X, the remaining problem reduces to identifying an optimal inventory replenishment strategy in a classical one-warehouse L-retailer system in which each set X_l plays the role of a single "super retailer," with demand rate $2|X_l|$ and a fixed procurement cost given by $c + \mathrm{TSP}(X_l^0)$ $(l = 1, \ldots, L)$. No method is known for computing an optimal strategy, even in the uncapacitated version of this problem, but Roundy has shown that for the latter a *close-to-optimal simple* strategy may be found of the following *power-of-two* structure: the warehouse (region l) replenishes its inventory every T_0 (T_l) time units when its inventory reaches zero $(l = 1, \ldots, L)$; also, $(T_0, T_1, \ldots, T_l)$ are power-of-two multiples of a base planning period T^B. A power-of-two policy exists whose cost comes within 6% or 2% of the optimum cost depending upon whether the base planning period is fixed or variable, respectively.

2. LOWER BOUNDS

In this section, we derive lower bounds for uncapacitated and capacitated models. In uncapacitated models, we assume that only upper bounds on the regions' sales volumes are imposed, i.e., no frequency constraints apply ($b = f^* = \infty$), whereas in capacitated models frequency constraints may be imposed as well. For the sake of notational convenience we assume that the sales volume bound is identical for all regions which implies a uniform upper bound (say M^{**}) on the number of demand points included in a region. (Extension to uncapacitated systems with nonidentical bounds is straightforward, given the general treatment in Anily and Federgruen (1990b, 1991a); for capacitated models, the upper bounds may be general power-of-two integers.)

In capacitated systems and under policies that employ constant replenishment intervals, the frequency and capacity constraints translate into upper and/or lower bounds for these intervals $\{T_l;\ l = 1, \ldots, L\}$ of the following form: $\lambda_l \leqslant T_l \leqslant \upsilon_l/|X_l|$, $\lambda_l \geqslant 0$, $\upsilon_l \leqslant \infty$, $l = 1, \ldots, L$; υ_l represents half the capacity of the vehicle assigned to route l and λ_l^{-1}, the maximum frequency with which this route may be driven. (For the sake of notational convenience we only consider cases where $\lambda_l = \lambda$ and $\upsilon_l = \upsilon$, $l = 1, \ldots, L$; our results may, however, be extended to cases where all λ_l (υ_l) are powers of two times some base value.) In uncapacitated models, $\lambda = 0$ and $\upsilon = \infty$ so that $\upsilon/\lambda = \infty$. We also need the following parameter restrictions:

i. υ/λ is either integer or $+\infty$;
ii. if $M^{**} < \upsilon/\lambda < \infty$, υ/λ is a power-of-two times M^{**};
iii. the maximum number of demand points that can be assigned to a single region is $M^* \overset{def}{=} \min\{M^{**}, \upsilon/\lambda\} < \infty$.

In view of ii, we always have that $\upsilon/M^* = \infty$ or υ/M^* is a power-of-two multiple of λ.

In view of iii we assume in all our models that $M_l^* < \infty$ for $l = 1, \ldots, L$, i.e., the upper bounds on the number of demand points per region are all finite.

For a given partition $\chi = \{X_1, \ldots, X_L\}$ of X not sales regions and a given power-of-two policy $T = (T_0, T_1, \ldots, T_L)$ we denote the corresponding *average cost* by

$$C_\chi(T) = K_0/T_0 + \sum_{l=1}^{L} D_{T_0}(T_l, \theta_l, m_l).$$

Here $D_{T_0}(T_l, \theta_l, m_l)$ represents the average cost per unit time of replenishing the lth sales region, which depends on T_0, T_l, $m_l = |X_l|$ and $\theta_l = \text{TSP}(X_l^0)$ ($l = 1, \ldots, \text{L}$); D_{T_0} includes the transportation costs which are incurred for X_l as well as the carrying costs for the inventories at the region's demand points and the part of the warehouse inventory which is destined to be shipped to X_l. According to Roundy

$$D_{T_0}(T_l, \theta_l, m_l)$$
$$= (\theta_l + c)/T_l + m_l(hT_l + h_0 \max(T_0, T_l)). \qquad (1)$$

Lemma 1 states that $\inf\{C_\chi(T): T_0 > 0 \text{ and } \lambda \leqslant T_l \leqslant v/|X_l|, l = 1, \ldots, L\}$ provides a lower bound for the minimum long-run average costs over *all* feasible policies that employ the regions in χ. For uncapacitated systems, this lower bound follows directly from Roundy.

To prove Lemma 1 for capacitated systems we need the following definitions. Let

$$\tau'(\theta, m) = ((\theta + c)/(m(h' + h_0)))^{1/2};$$
$$\tau(\theta, m) = ((\theta + c)/(mh'))^{1/2} \qquad (2)$$

be the order intervals obtained by the EOQ formula with a fixed cost of $\theta + c$, a demand rate of two, and holding cost rates of $m(h' + h_0)$ and mh', respectively. For any given partition χ and for any $T > 0$ we partition the index set $\{1, \ldots, L\}$ into the following seven sets (some of which may be empty).

$$\begin{aligned}
G(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad T \leqslant \tau_l' \leqslant v/m_l \text{ and } \lambda < \tau_l'\} \\
E(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad \lambda < T < v/m_l \text{ and} \\
&\qquad \tau_l' \leqslant \mathrm{T} \leqslant \tau_l\} \\
S(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad \lambda < \tau_l \leqslant \mathrm{T} \text{ and } \tau_l \leqslant v/m_l\} \\
I_1(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad T < \lambda \text{ and } \tau_l' \leqslant \lambda\} \\
I_2(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad \tau_l < \lambda < T\} \\
I_3(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad T \leqslant v/m_l \text{ and } v/m_l < \tau_l'\} \\
I_4(T) &= \{l \mid 1 \leqslant l \leqslant L, \quad T > v/m_l \text{ and } v/m_l < \tau_l\}.
\end{aligned}$$

Observe that the sets $G(T)$, $E(T)$ and $S(T)$ consist of the regions l for which neither the capacity nor the frequency constraints are binding; i.e., for the value of T_l which minimizes D_T we have $T_l > (=, <) T$ for $l \in G(T)$ ($E(T)$, $S(T)$). Here $I_1(T)$ and $I_2(T)$ ($I_3(T)$ and $I_4(T)$) consist of those routes for which the frequency (capacity) constraints are binding:

$$\lambda = T_l > (<) T \quad \text{for } l \in I_1(T)(I_2(T)),$$

$$v/m_l = T_l \geqslant (<) T \quad \text{for } l \in I_3(T)(I_4(T)).$$

Note that if $\lambda = 0$ ($v = \infty$), then $I_1(T) = I_2(T) = \varnothing$ ($I_3(T) = I_4(T) = \varnothing$) for all T. A vector $\mathbf{T}$ is said to preserve the order of T^* if the sets $\{T_0, T_1, \ldots, T_L, \lambda, v/M^*\}$ and $\{T_0^* T_1^*, \ldots, T_L^*, \lambda, v/M^*\}$ are ranked in the same way.

Lemma 1. *For any given partition $\chi = \{X_1, \ldots, X_L\}$ of X, $U(\chi) \overset{def}{=} \inf\{C_\chi(T)$: $T > 0$ and $\lambda \leqslant T_l \leqslant v/|X_l|$, $l = 1, \ldots L\}$ is a lower bound for the minimum long-run average costs over all feasible policies that employ the regions in χ.*

Proof. The long-run average cost associated with the partition χ and any order-preserving vector $\mathbf{T}$ may be written in the form

$$C_\chi(\mathbf{T}) = K/T_0 + HT_0 + \sum_{l \notin E(T_0)} (K_l/T_l + H_l T_l),$$

where

$$K(T_0) = K_0 + \sum_{l \in E(T_0)} K_l \qquad (K_l = \theta_l + c, l = 1, \ldots, L),$$

$$H(T_0) = \sum_{l \in E(T_0)} m_l(h' + h_0) + \sum_{l \in S(T_0) \cup I_2(T_0) \cup I_4(T_0)} m_l h_0$$

$$H_l(T_0) = \begin{cases} m_l(h' + h_0) & l \in G(T_0) \cup I_1(T_0) \cup I_3(T_0) \\ m_l h' & l \in S(T_0) \cup I_2(T_0) \cup I_4(T_0), \end{cases}$$

Let

$$M(T_0) = 2(K(T_0)H(T_0))^{0.5}$$

$$M_l(T_0) = \begin{cases} 2(K_l H_l(T_0))^{0.5}, & l \in G(T_0) \cup S(T_0) \\ K_l/\lambda + H_l(T_0)\lambda, & l \in I_1(T_0) \cup I_2(T_0) \\ K_l|X_l|/v + H_l(T_0)v/|X_l|, & l \in I_3(T_0) \cup I_4(T_0). \end{cases}$$

Similarly to Lemma 1 *in Anily and Federgruen* (1991*b*) *we get for the vector* $\mathbf{T}^*$ achieving the minimum in the definition of $U(\chi)$:

a. $U(\chi) = M(T_0^*) + \sum_{l \notin E(T_0^*)} M_l(T_0^*)$

b. $T_0^* = \sqrt{K(T_0^*)/H(T_0^*)}$; $T_l^* = \sqrt{K_l/H_l(T_0^*)}$

$$l \in G(T_0^*) \cup S(T_0^*).$$

The remainder of the proof is based on a modification of the proof of Theorem 1 in Roundy as presented in Anily and Federgruen (1991b).

We now describe how $U(\chi)$ may be evaluated. For any partition $\chi = \{X_1, \ldots, X_L\}$ of X and $T_0 > 0$ we define

$$U_{T_0}(\chi) = K_0/T_0 + \sum_{l=1}^{L} f_{T_0}(\theta_l, m_l), \qquad (3)$$

where $f_{T_0}(\theta, m) = \inf_{\lambda \leqslant T_l \leqslant v/|X_l|} D_{T_0}(T_l, \theta, m)$.

Note that $U(\chi) = \inf_{T_0} U_{T_0}(\chi)$. The infimum to the right of (3) is easy to evaluate. If $T_0 \leqslant \lambda$ then,

$$f_{T_0}(\theta, m) = \begin{cases} m(\theta + c)/v + v(h' + h_0) & \text{if } v/m < \tau' \\ 2[(\theta + c)m(h' + h_0)]^{1/2} & \text{if } \lambda \leqslant \tau' < v/m \\ (\theta + c)/\lambda + m(h' + h_0) & \text{if } \tau' < \lambda. \end{cases} \quad (4)$$

If $\lambda < T_0 \leqslant v/m$ then,

$$f_{T_0}(\theta, m) = \begin{cases} m(\theta + c)/v + v(h' + h_0) & \text{if } T_0 \leqslant v/m < \tau' \quad (5a) \\ 2[(\theta + c)m(h' + h_0)]^{1/2} & \text{if } T_0 < \tau' \leqslant v/m \quad (5b) \\ (\theta + c)/T_0 + m(h' + h_0)t_0 & \text{if } T' \leqslant T_0 \leqslant \tau \quad (5c) \\ 2[(\theta + c)mh']^{1/2} + mh_0T_0 & \text{if } \lambda \leqslant \tau < T_0 \quad (5d) \\ (\theta + c)/\lambda + mh'\lambda + mh_0T_0 & \text{if } \tau < \lambda \leqslant T_0, \quad (5e) \end{cases}$$

and if $(\lambda \leqslant)\ v/m < T_0$ then,

$$f_{T_0}(\theta, m) = \begin{cases} m(\theta+c)/v + h'v + mh_0T_0 & \text{if } v/m < \tau \\ 2[(\theta+c)mh']^{1/2} + mh_0T_0 & \text{if } \lambda \leqslant \tau \leqslant v/m \\ (\theta+c)/\lambda + mh'\lambda + mh_0T_0 & \text{if } \tau < \lambda. \end{cases} \quad (6)$$

(Note that the function D_{T_0} is convex in T_l; its unconstrained minimum is thus obtained at τ_l' if $T_0 < \tau_l$, at T_0 if $\tau_l' \leqslant T_0 \leqslant \tau_l$, and at τ_l otherwise. Equations 4–6 represent the minimum of D_{T_0} under the restrictions $\lambda < T_l \leqslant v/m_l$.) We conclude from Lemma 1, (2) and (3) that

$$\underline{V}(X) = \inf_{T_0 > 0} \left\{ K_0/T_0 + \min \left[\sum_{l=1}^{L} f_{T_0}(\theta_l, m_l) : \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ and } m_l \leqslant M^* \right] \right\} \quad (7)$$

is a lower bound for $V^*(X)$. It follows that for any $T_0 > 0$, the minimization problem within the curled brackets in (7) reduces to the problem of partitioning the set X into L routes with minimal total cost, where the cost of a route with length θ and $m \leqslant M^*$ demand points is given by $f_{T_0}(\theta, m)$. This class of routing problems with general route cost function has been addressed in Anily and Federgruen (1990b) and in the remainder we draw on the results of this paper. Since θ_l represents the length of an optimal traveling salesman tour, (7) is still too complex to be evaluated. We therefore replace $\underline{V}(X)$ by a further lower bound. For any $T > 0$, consider the partitioning problem

$$\underline{P}_T^1 : \underline{V}_T^1(X) = K_0/T + \min \left\{ \sum_{l=1}^{L} f_T\left(\frac{2}{m_l} \sum_{i \in X_l} r_i,\ m_l \right) : \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ anc } m_l \leqslant M^* \right\},$$

and let $\underline{V}'(X) = \inf_T \underline{V}^1(X)$. Clearly $V^*(X) \leqslant V(X) \leqslant \underline{V}^1(X)$ (see also, Anily and Federgruen 1990b).

Note that the partitioning problem $\underline{P}_T^1$ depends on the parameter $T > 0$. Thus, even if $\underline{P}_T^1$ falls in the small class of polynomially solvable partitioning problems, the optimal partition may be expected to vary with T, so that $\underline{V}_T^1(X)$ as the finite minimum of convex functions may fail to be convex and the minimization over T required to evaluate $\underline{V}^1(X)$ can be expected to be cumbersome. The next theorem shows, fortunately, that $\underline{P}_T^1$ is extremal so a unique, easy to compute partition χ exists which optimizes $\underline{P}_T^1$ for all $T \geqslant 0$.

Theorem 1. *Fix $T > 0$.*

a. *The function $f_T(\theta, m)$ is concave in both arguments separately.*
b. *The function $h_T(\theta, m) \overset{\text{def}}{=} f_T(\theta/m, m)$ has antitone differences, i.e., $h_T(\theta_2, m) - h_T(\theta_1, m)$ is nonincreasing in m for all $\theta_2 > \theta_1$.*
c. *P_T^1 is extremal for all T; the same partition χ^* optimizes P_T^1 for all $T > 0$.*

Proof. See the Appendix.

As mentioned in Section 1, the partition χ^* which optimizes $\underline{P}_T^1$ for all $T > 0$ is easy to construct by the **EPA** algorithm in Anily and Federgruen (1991a).

Let $N = k_1 M^* + K_2$ with $0 \leqslant k_2 < M^*$. The partition generated by **EPA** for L unrestricted is the partition

$$\chi^* = \{X_1^*, \ldots, X_L^*\} = \{\{x_1, \ldots, x_{k_2}\}, \{x_{k_2+1}, \ldots, x_{k_2+M^*}\}, \ldots, \{x_{N-M^*+1}, \ldots, x_N\}\},$$

where $L = \lceil N/M^* \rceil$. In case an additional constraint on the number of sales regions is imposed (i.e., $L = \bar{L}$ with $\bar{L}$ a given parameter), the **EPA** generates the consecutive partition in which all groups, except possibly one, are either singletons or filled to capacity. The singletons consist of the lowest indexed elements in X; for details see Anily and Federgruen (1991a).

It follows from Theorem 1 that

$$\underline{V}^1(X) = \inf_T \underline{V}^1_T(X)$$
$$= \inf_T\left\{K_0/T + \sum_{l=1}^{L} f_T\left(2 \sum_{i\in X_l^*} r_i/|X_l^*|,|X_l^*|\right)\right\}$$

and in view of the shape of the function $f_T(.\,,.)$:

$$\underline{V}^1_T(X) = K_0/T_0 + \sum_{l=1}^{L}\left[\frac{\alpha_l(T_0)}{T_0} + \beta_l(T_0) + \gamma_l(T_0)T_0\right]$$

with $\alpha_l(T_0)$, $\beta_l(T_0)$ and $\gamma_l(T_0)$ piecewise constant functions for all $l = 1, \ldots, L$. It is easy to verify that each of the functions

$$[\alpha_l(T_0)/T_0 + \beta_l(T_0) + \gamma_l(T_0)T_0] \quad (l = 1, \ldots, L)$$

is convex and continuously differentiable in T_0 except possibly for capacitated systems at $T_0 = \lambda$, $T_0 = \upsilon/M^*$ and if $k_2 > 0$, then also for $T_0 = \upsilon/k_2$, so that $\underline{V}^1_{T_0}(X)$ is *strictly* convex and continuously differentiable in T_0 everywhere, except possibly for the points λ, υ/M^* and υ/k_2.

Thus, $\underline{V}^1(X) = \inf_{T_0>0}\underline{V}^1_{T_0}(X)$ is achieved at the unique point T^* where either $d\underline{V}^1_{T_0}(X)/dT_0 = 0$ (i.e., $T_0^* = [\alpha(T_0^*)/\gamma(T_0^*)]^{1/2}$ or $T_0^* \in \{\lambda, \upsilon/M^{**}, \upsilon/k_2\}$ in capacitated systems. Also $\underline{V}^1_T(X)$ is of the form $\alpha(T)/T + \beta(T) + \gamma(T)T$, where $\alpha(T)$, $\beta(T)$ and $\gamma(T)$ are piecewise constant functions changing values only when T crosses one of at most $2N + 3$ values in

$$\{\tau_l^{*\prime}: l = 1, \ldots, L\}$$
$$\cup \{\tau_l^*: l = 1, \ldots, L\} \cup \{\lambda, \upsilon/M^*, \upsilon/k_2\},$$

where $\tau_l^{*\prime}$ and τ_l^* are defined as in (2) replacing θ_l by $2 \sum_{i\in X_l} r_i/|X_l^*|$ and m_l by $|X_l^*|$. These observations suggest a simple $O(N \log N)$ algorithm for the minimization of $\underline{V}^1_T(X)$ (see Appendix A in Roundy). Queyranne (1987) proposes an alternative linear time procedure which uses a linear time median finding procedure.

Remark 1. As in Anily and Federgruen (1990a, b), one may consider the alternative lower bound $V^2(X)$, where $\underline{V}^2(X) = \inf_T V^2_T(X)$, and

$$\underline{V}^2_T(X)$$
$$= K_0/T + \min\left\{\sum_{l=1}^{L} f_T(2 \max_{i\in X_l} r_i, m_l):\right.$$
$$\chi = \{X_1, \ldots, X_L \text{ partitions } X$$
$$\left.\text{and } m_l \leqslant M^*\}\right\}.$$

Clearly, $V^*(X) \geqslant \underline{V}(X) \geqslant \underline{V}^2(X) \geqslant \underline{V}^1(X)$, i.e., $\underline{V}^2(X)$ is a better lower bound than $\underline{V}^1(X)$. Evaluation of $\underline{V}^2(X)$ is, however, significantly more complex than that of $\underline{V}^1(X)$ because the corresponding partitioning problem $\underline{P}^2_T$ may fail to be extremal (see Anily, pp. 135–137).

3. UPPER BOUNDS AND FEASIBLE SOLUTIONS

The partition χ^* associated with the lower bound $\underline{V}^1(X)$ represents regions in which the points have similar radial distances to the depot but may otherwise be far apart. In this section, we describe the construction of a heuristic partition χ^H and associated inventory strategy which is asymptotically ϵ − optimal for $\epsilon = 0.06$ under general conditions regarding the stochastic model used to generate $\{x_1, x_2, \ldots\}$. For uncapacitated systems we also describe a variant which is asymptotically ϵ − optimal for $\epsilon = 0.02$. The corresponding cost values of the heuristics represent upper bound for $V^*(X)$ which are asymptotically ϵ − accurate for the same values of ϵ.

Let $X^{(m)} = \{x_i \in X: \chi^* \text{ assigns } x_i \text{ to a set of cardinality } m\}$, $m = 1, \ldots, M^*$ denote the set of demand points which are assigned to a region with m points (in total).

Recall the unique representation of $N = k_2 + k_1M^*$ $(0 \leqslant k_2 < M^*)$. Thus, for L unrestricted, at most two of the sets $X^{(m)}$ are nonempty ($m = k_2$ if $k_2 \neq 0$ and $m = M^*$). If L is restricted there may be up to three nonempty sets. Anily and Federgruen (1990b) propose a general partitioning scheme (the so-called Modified Circular Regional Partitioning (MCRP) procedure) which is applicable to routing problems with general route cost functions. This procedure operates on each set $X^{(m)}$ $(m = 1, \ldots, M^*)$ *separately and clusters the points in* $X_{(m)}$ into regions of cardinality m each.

Note from the above characterization of the partition χ^* optimizing $\underline{P}^1$ that only the points of $X^{(M^*)}$ need to be partitioned into regions because at most one of the sets $X^{(m)}$ with $1 < m < M^*$ may be nonempty, and this set consists of exactly m points which necessarily need to be assigned to the same region.

The MCRP generates a collection of regions $\chi^H = \{X_1^H, \ldots, X_L^H\}$. Given this set of regions, and as described in Section 1, the system reduces to a one-warehouse multiretailer system, with each region acting as a (super) retailer. An optimal power-of-two policy for this system is easy to determine following the procedure in Roundy with appropriate modifications in capacitated systems. The first step in determining such a power-of-two policy is the determination of a vector $\mathbf{T}^H$ that achieves $\hat{V}(X) \overset{\text{def}}{=} \min_{T>0} C_{X^H}(\mathbf{T})$. This can be done with the same $O(L \log L)$ or $O(L)$ procedure required to compute the lower bound $\underline{V}^1(X) = \min_{\mathbf{T}>0} C_{\chi^*}(\mathbf{T})$ (see Section 2).

Next the following rounding procedure is employed to round $\mathbf{T}^H$ to a feasible power-of-two policy $\bar{\mathbf{T}}^H$ with respect to the base planning period $T^B = \lambda$. (If $\lambda = 0$, choose $T^B = v/M^*$.) In the uncapacitated case where $\lambda = 0$ and $v = \infty$, T^B may be chosen arbitrarily. Let $G = G(\mathbf{T}^H)$, $E = E(\mathbf{T}^H)$, $S = S(\mathbf{T}^H)$, $I_r = I_r(\mathbf{T}^H)(r = 1, \ldots, 4)$. (For the sake of simplicity, we assume from hereon that L is unrestricted, so that all regions, except possibly X_1^H, contain M^* points.)

Rounding Procedure (RP)

STEP 0. Determine the (unique) integer k such that $T_0^H \epsilon T^B[2_{k-1}, 2_k)$. *If* $T_0^H \leq T^B 2^{k-1}\sqrt{2}$, $\bar{T}_0^H := T^B 2^{k-1}$; otherwise $\bar{T}_0^H := T^B 2^k$.

STEP 1. For $l \in I_1 \cup I_2$, $\bar{T}_l^H := T_l^H := \lambda$. For $l \in I_3 \cup I_4$, $\bar{T}_l^H := T_l^H = v/|X_l|$. For $l \in E$, $\bar{T}_l^H := \bar{T}_0^H$.

STEP 2. For all $l \in G \cup S$ determine the (unique) integer k such that $T_l^H/\bar{T}_0^H \in [2^{k-1}, 2^k)$. If $T_l^H/\bar{T}_0^H < 2^{k-1}\sqrt{2}$, $\bar{T}_l^H := \bar{T}_0^H 2^{k-1}$; otherwise $\bar{T}_l^H := \bar{T}_0^H 2^k$.

STEP 3. If $|X_1| < M^*$ and $\bar{T}_1^H \geqslant v/|X_1|$, then $\bar{T}_1^H := \max\{2^k\lambda\colon 2^k T^B \leqslant v/|X_1|\}$.

(Step 3 is required because $v/|X_1|$ is not necessarily a power-of-two times λ.) Let $\bar{V}(X) = C_{\chi^H}(\bar{T}^H)$.

Lemma 2

a. *The rounding procedure generates a power-of-two vector* $\bar{T}^H = (\bar{T}_0^H, \bar{T}_1^H, \ldots, \bar{T}_L^H)$. *Moreover, the vector* $(\bar{T}_0^H, \bar{T}_2^H, \ldots, \bar{T}_L^H)$ *is order-preserving with respect to* $(T_0^H, T_2^H, \ldots, T_L^H)$.

b. $\bar{V}(X)/\hat{V}(X) \leqslant 1.061 + M^*/N[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2]$. *Moreover, if N is a multiple of* M^* *or* $v = \infty$, $\bar{V}(X)/\hat{V}(X) \leqslant 1.061$.

Proof. a. We write $\bar{T}$ instead of $\bar{T}^H$ and X_l insead of X_l^H; $\bar{T}$ is clearly a power-of-two policy since $m_l = M^*$ for $l > 1$. It is easy to verify that $(\bar{T}_0, \bar{T}_2, \ldots, \bar{T}_L)$ preserves the order of $(T_0^H, T_2^H, \ldots, T_L^H)$ and that $\bar{T}$ is feasible by comparing $\bar{T}_l$ with $\bar{T}_0$, λ and v/M^* for $l \in G\backslash\{1\}$, $l \in E\backslash\{1\}$, $l \in S\backslash\{1\}$ and $l \in I_r$ $(r = 1, \ldots, 4)$, respectively.

b. Let $q_l = \bar{T}_l/T_l^H$ for $l \neq 1$. Clearly, $\sqrt{0.5} \leqslant q_l \leqslant \sqrt{2}$ for $l \neq 1$. If $1 \in I_1 \cup I_2$, $q_1 = 1$. If $1 \in I_3 \cup I_4$, then $\bar{T}_1$ is first set in Step 1 and reset in Step 3; note that in this case $\tfrac{1}{2}T_1^H \leqslant \bar{T}_1 \leqslant T_1^H$. If $1 \in G \cup S$ and T_1^H is rounded down in Step 2, Step 3 is not executed and $\sqrt{0.5}\ T_1^H \leqslant T_1 \leqslant T_1^H$. Finally, if $l \in G \cup S$ and T_1^H is rounded up in Step 2, $\tfrac{1}{2}T_1^H \leqslant \bar{T}_1 \leqslant \sqrt{2}T_1^H$. We conclude that in all cases, $\tfrac{1}{2} \leqslant q_1 \leqslant \sqrt{2}$. These inequalities and $\sqrt{0.5} \leqslant q_0 \leqslant \sqrt{2}$ imply that with $K_l = \mathrm{TSP}(X_l^0) + c$ $(l = 1, \ldots, L)$,

$$\begin{aligned}
&D_{\bar{T}_0}(\bar{T}_1, \mathrm{TSP}(X_{1'}^0), |X_1|)\\
&\quad = K_1/\bar{T}_1 + m_1 h\bar{T}_1 + m_1 h_0 \max(\bar{T}_0, \bar{T}_1)\\
&\quad \leqslant \max\{K_1/\tau + m_1 h'\tau + m_1 h_0 \max(\bar{T}_0, \tau)\colon\\
&\qquad\qquad \tfrac{1}{2}T_1^H \leqslant \tau \leqslant \sqrt{2}T_1^H\}\\
&\quad \leqslant 2K_1/T_1^H + m_1 h'\sqrt{2}T_1^H + m_1 h_0 \max(\sqrt{2}T_0^H, \sqrt{2}T_1^H)\\
&\quad \leqslant 2[K_1/T_1^H + m_1 h' T_1^H + m_1 h_0 \max(T_0^H, T_1^H)]\\
&\quad = 2f_{T_0^H}(\mathrm{TSP}(X_1^0), |X_1|).
\end{aligned}$$

It follows from Theorem 3 in Haimovich and Rinnooy Kan (1985) that

$$\begin{aligned}
K_1 &= \mathrm{TSP}(X_1^0) + c \leqslant 2r_{|X_1|}[\sqrt{\pi}\sqrt{|X_1|} + 1 + \pi/2] + c\\
&\leqslant [\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2](2r_{|X_1|} + c)\\
&\leqslant [\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2]K_l, \quad l = 1, \ldots, L.
\end{aligned}$$

It thus follows from (1) and $m_1 \leqslant m_l$ $(l = 1, \ldots, L)$ that

$$\begin{aligned}
&D_{\bar{T}_0}(\bar{T}_1, \mathrm{TSP}(X_1^0), |X_1|)\\
&\quad \leqslant 2[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2]\min_{1\leqslant l\leqslant L}\{f_{T_0^H}(\mathrm{TSP}(X_l^0), |X_l|)\}\\
&\quad \leqslant \frac{2}{L}[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2] \sum_{l=1}^{L} f_{T_0^H}(\mathrm{TSP}(X_l^0), |X_l|)\\
&\quad \leqslant \frac{2M^*}{N}[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2]C_\chi^H(T^H).
\end{aligned}$$

In view of part a, we have by the proof of Lemma 1, with M, M_l $(l \notin E)$ and the index defined for the partition χ^H and T_0^H that

$$\begin{aligned}
&\bar{V}(X)/\hat{V}(X) = C_{\chi^H}(\bar{T})/C_{\chi^H}(T^H)\\
&\quad \leqslant D_{\bar{T}_0}(\bar{T}_1, \mathrm{TSP}(X_1^0), |X_1|)/C_{\chi^H}(T^H)\\
&\quad + \frac{\mathscr{A}}{[M + \sum_{l\notin E\cup\{1\}} M_l]}\\
&\quad \leqslant 2\,\frac{M^*}{N}\,[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2]\\
&\qquad + \max\left(\tfrac{1}{2}\left(\frac{1}{x} + x\right)\colon \sqrt{0.5} \leqslant x \leqslant \sqrt{2}\right)\\
&\quad = \frac{2M^*}{N}\,[\sqrt{\pi}\sqrt{M^*} + 1 + \pi/2] + 1.061,
\end{aligned}$$

where

$$\mathscr{A}=\left[\frac{M}{2}\left(\frac{1}{q}+q\right)\right.$$

$$\left.+\textstyle\sum_{l\in G\cup S/\{1\}}\frac{M_l}{2}\left(\frac{1}{q_l}+q_l\right)+\sum_{l\in I_1\cup I_2\cup I_3\cup I_4\setminus\{l\}}M_l\right]$$

(See also Appendix A.1 in Anily.) Finally, if $|X_1| = M^*$ or $v = \infty$, $\bar{T}$ is order-preserving (including T_1) and the simpler bound arises.

Remark 2. If $\text{TSP}(X_1^0) = \min_{1\leqslant l\leqslant L}\text{TSP}(X_l^0)$, as is usually the case, the bound in part b of Lemma 2 may be simplified.

We now characterize the worst-case optimality gap of our proposed heuristic.

Theorem 2

a. $V^1(X) \leqslant V^*(X) \leqslant \bar{V}(X)$.

b. *Consider an infinite stochastic sequence of demand points* $\{x_1, x_2, \ldots\}$ *with i.i.d. radial distances distributed as the bounded random variable* $\underline{r}$ *such that*

$$E\left[\min_{m=1,\ldots,M^*} f_{T_0}(2r, m)\right] > 0 \quad \textit{for } T_0 > 0.$$

Then,

$$\lim_{N\to\infty}\frac{\bar{V}(X_{(N)})}{\underline{V}^1(X_{(N)})} = 1.061 \text{ a.s.}$$

Proof. In view of Lemma 2b, it suffices to show that $\lim_{n\to\infty}\hat{V}(X_{(N)})/\underline{V}^1(X_{(N)}) = 1$. As above, let T_0^* denote the unique value of T_0 with $\underline{V}^1(X) = \underline{V}^1_{T_0^*}(X) = \inf_{T_0}\underline{V}^1_{T_0}(X)$. Let

$$\tilde{V}(X) = K_0/T_0^* + \sum_{l=1}^{L} f_{T_0^*}(\text{TSP}(X_l^H), |X_l^H|).$$

It thus suffices to show that

$$\lim_{N\to\infty}\tilde{V}(X_{(N)})/\underline{V}^1(X_{(N)}) = 1 \text{ a.s.}$$

Consider now the Euclidean vehicle routing problem in which the cost of a route of (Euclidean) length θ, going through m points and the depot is given by $f_{T_0^*}(\theta, m) \geqslant f_{T_0^*}(\underline{\theta}, m)$ with $\underline{\theta} \leqslant \theta$ twice the average radial distance of the m points on the route. Clearly,

$$\tilde{V}(X) \geqslant V(X)$$

$$= \inf_{T>0} C_{X^H}(T)$$

$$= \inf_{T_0>0}\left[K_0/T_0 + \sum_{l=1}^{L} f_{T_0}(\text{TSP}(X_l^{0H}), |X_l^H|)\right]$$

$$\geqslant \inf_{T_0>0}\left[K_0/T_0 + \sum_{l=1}^{L} f_{T_0}(2\sum_{i\in X_l^H} r_i/|X_l^H|, |X_l^H|)\right.$$

$$\left.\geqslant \inf_{T_0>0}[\underline{V}^1_{T_0}(X)\right] = \underline{V}^1(X).$$

The theorem now follows from Theorem 2 in Anily and Federgruen (1990b) because $f_{T_0^*}$ is concave in its first argument; see Theorem 1a.

Remark 3. For uncapacitated systems, it is possible to guarantee that the heuristic comes asymptotically within 2% of optimality by optimizing over the base planning period T^B. This optimization can be performed in $O(N)$ time (see Roundy). In uncapacitated models with frequency constraints (in addition to sales volume constraints) $v = \infty$ and $\lambda = (f^*)^{-1} > 0$. In this case, **RP** generates a complete order-preserving policy and $\bar{V}(X)/\hat{V}(X) \leqslant 1.061$ (see Lemma 2, part b). For capacitated models without frequency constraints ($\lambda = 0$, $v = b/2 < \infty$), the rounding procedure results in a power-of-two policy which is order-preserving (with respect to T^H) with the possible exception of T_1^H, the replenishment interval of the first sales region. As demonstrated in Lemma 2, and in spite of this complication, the generated solution comes asymptotically within 6% of the true minimum cost and so do the computed lower and upper bounds.

An interesting situation arises in capacitated models with frequency constraints but *without* sales volume bounds. In this case, $\lambda = 1/f^*$, $v = b/2$, $M^{**} = \infty$, and $M^* = bf^*/2$. Thus, $\lambda = v/M^*$ so that the feasible intervals for the regions' replenishment intervals reduce to the single value $1/f^* = b/(2M^*)$ (with the possible exception of the first region). In this case, $\{2, \ldots, L\} \subset I_1(T) \cup I_2(T) \cup I_3(T) \cup I_4(T)$. It follows from the proof of Lemma 2 that the generated solution (bounds) are asymptotically fully optimal (accurate)!

The entire procedure required to obtain the lower bound, upper bound and heuristic solution is thus of complexity $O(N \log N)$ and can be summarized as follows.

Combined Routing and Replenishment Strategies Algorithm for One Warehouse Multiretailer Systems With Central Inventories (CRRSA*)

STEP 1. Use the **EPA** algorithm to determine the partition χ^*. Compute the lower bound $\underline{V}(X)$ by the procedure in Section 2.

STEP 2. Apply MCRP; let χ'' be the resulting collection of sales regions. Find the vector $\mathbf{T}''$ achieving $\min_{T>0} C_{\chi''}(T)$ via the procedure in Section 2.

STEP 3. Use the rounding procedure to round $\mathbf{T}''$ to a power-of-two vector $\bar{\mathbf{T}}''$.

Remark 4. The uncapacitated model may be extended to systems with a backlogging option, at a cost of h^- per unit and per unit of time, if all retailers face identical demand rates, i.e., each retailer is a single demand point.

Consider a fixed collection of sales regions $\chi = \{X_1, \ldots, X_L\}$. Mitchell (1987) has demonstrated that the class of power-of-two inventory policies needs to be enlarged to the so-called "near-integer ratio policies" if we wish to get close to optimality. In a near-integer ratio policy, only the warehouse's inventory is necessarily replenished at constant intervals (of length T_0).

Mitchell has shown that the cost of any near-integer ratio policy may be expressed as a function of $(T_0, T_1, \ldots, T_L)$ only (with T_l now interpreted as the *average* replenishment interval) and that the cost expression is identical to that obtained in the model without backlogging, provided the holding cost rates are appropriately transformed: let $\alpha = h^-/(h^- + h)$. Replace h_0 by αh_0 and h by $\alpha\alpha' h$, where $\alpha' = h^-/(h^- + h + h_0)$.

In view of the above observations, it is easy to verify that the analysis of the uncapacitated backlogging model proceeds along the lines of the uncapacitated model without backlogging. In particular **CRRSA*** may be applied to generate bounds and a heuristic solution, merely replacing h' and h_0 by the above stated expressions. In addition, all optimality and accuracy results continue to hold.

4. A NUMERICAL STUDY

In this section, we summarize a numerical study conducted to assess the performance of **CRRSA*** and associated bounds for problems of moderate size. For the complete report see Anily and Federgruen (1990c). The study thus serves to complement the asymptotic optimality and accuracy results derived in Sections 2-3. We have analyzed both capacitated and uncapacitated models, all without frequency constraints but with sales volume upper bounds. We have also assumed that each retailer consists of a single demand point.

In all uncapacitated models we assume that the base planning period is fixed and equal to b/M^*. The purpose of our study is to assess the computational requirements of the **CRRSA*** algorithm as well as the optimality gap of the generated solutions and to compare the performance of **CRRSA*** to that used in Anily and Federgruen (1990a) for systems without central inventories. The ratios of the computed upper and lower bound (UB/LB) serve as upper bounds for the optimality gaps.

We conclude that our procedures have modest computational requirements which grow roughly linearly with the number of locations. For example, for a problem with 1,000 demand points in which no route visits more than four distinct points, the entire solution procedure (i.e., computation of the lower bound, upper bound, routes and inventory strategies) requires no more than about 0.7 CPU seconds when encoded in FORTRAN (Tops 20—Version 2) and run on an Amdahl 170V8 computer.

The generated solutions come within a relatively small percentage of a lower bound for the minimal system-wide costs (within the class Φ), even for problems of moderate size. The observed (bounds for the) optimality gaps are almost always smaller than those computed for the corresponding systems without central inventories, even though the theoretical asymptotic bounds are worse. For example, for problems with $N = 100$ and $M^* = 4$, the average optimality gap is 9.5% in our systems versus 18.8% for systems without central inventories. For problems with 500 demand points and $M^* = 4$ the average optimality gaps are 6.3% and 7.3%, respectively; and for problems with $N = 1{,}000$ and $M^* = 7$, the gaps are 6.7% and 10.3% only. (Only for problems with 1,000 demand points and a maximum of 4 points per route, is the average optimality gap of 6.7% somewhat larger than the corresponding average of 5.3% in systems without central inventories.) The optimality gaps are only slightly larger than the asymptotic worst-case gap of 6.1% which also applies to all values of N in systems with separable replenishment costs (see Roundy).

Table I summarizes the performance of **CRRSA*** in all 136 problems. The retailers' locations are always randomly generated in a square of 200 × 200 with the depot placed in its center. The traveling salesman tours are obtained by complete enumeration. This

Table I
Summary Results

N	M^a	No. of Problems	UB/LB μ^a σ^b	CPU Time (μ)	σ (Seconds)
100	4	40	1.095/0.056	0.230	0.014
500	4	40	1.063/0.030	0.402	0.011
500	7	6	1.122/0.056	14.502	0.064
1,000	4	40	1.055/0.035	0.650	0.024
1,000	7	6	1.067/0.021	28.708	0.040

[a] The average value within category.
[b] The standard deviation of values within category.

step in the **CRRSA*** algorithm accounts for most of the CPU time as may be inferred by comparing the average CPU time between problem categories with an identical number of retailers (N) but different values for M^*: Increasing M^* from 4 to 7 leads to an increase in the average CPU time by a factor of 40 (approximately). Note from the description of the algorithm that the required number of elementary operations and evaluations of the function $f_{T_0}(\cdot, \cdot)$ depends largely on N and M^*. For values of $M^* \geqslant 7$ (say) one should determine the optimal traveling salesman tours by a more sophisticated exact method. Alternatively, a heuristic TSP method with bounded worst-case performance may be employed (e.g., Christofides 1976), maintaining all the (asymptotic) accuracy and optimality results.

The ratios UB/LB and the CPU times are predictable as a function of N and M^* only. Moreover, the ratios UB/LB are quite low even for small problems with only 100 demand points; they decrease with N and increase with M^*, which is consistent with the analyses in Anily and Federgruen (1990b) and Lemma 5, exhibiting the error gaps as a function of the number of sales regions $L = [N/M^*]$. Note, in addition, that the lower bound approximation for the length of a traveling salesman tour by two times the average value of the radial distances becomes increasingly less accurate as the number of demand points per region increases.

The models are evaluated with $N = 100$, $N = 500$ and $N = 1{,}000$ points and $M^* = 4$ or $M^* = 7$. Only for settings where $M^* = 7$ have we omitted the runs for $N = 100$ because the generated solution would consist of only $15 = (\lceil 100/7 \rceil)$ sales regions. (The MCRP scheme creates 11 sectors; 10 consist of exactly $M^* = 7$ points.) The models are systematically evaluated for several different values of all cost parameters. We also investigate the impact of progressively more severe vehicle capacity constraints. In Anily and Federgruen (1990c) we list for each of the scenarios the same performance measures as those reported in Anily and Federgruen (1990a).

5. CONCLUSIONS

We have shown how cost effective system-wide replenishment strategies can be computed for one-warehouse, multiretailer systems in which goods are distributed from the warehouse to the retailers by a fleet of vehicles, combining deliveries into efficient routes.

These strategies are chosen in a class Φ, as defined in the Introduction. We have shown that the gap between the cost of the proposed strategy and a lower bound for the minimum cost (among all strategies in Φ) is bounded by 6% for sufficiently large numbers of retailers, and this gap is small even for problems with a moderate number of retailers or outlets. Computation of the complete replenishment strategy (routes and inventory strategies), as well as the lower bound cost approximation, requires no more than $O(N \log N)$ time.

The restriction to the class Φ is clearly associated with some loss of optimality, the exact magnitude of which is as yet unknown. On the other hand, as explained in Anily and Federgruen (1990a), the restriction is often imposed by the sales/distribution system itself. In many systems the sales and delivery functions are integrated: A salesperson is assigned to a given region and each salesperson is required to visit the outlets in his/her region periodically in a given route, determining replenishment quantities (in the form of definite sales or unbinding consignments) and delivering them as well. (See the Introduction for additional discussion on the relative merits of class Φ.)

APPENDIX

Proof of Theorems 1

a. We first show that $\partial f_T/\partial\theta(\partial f_T/\partial m)$ exists and is continuous in θ (m). Since $\partial^2 f_T/\partial\theta^2(\partial^2 f_T/\partial m^2)$ exists and is nonpositive almost everywhere, we conclude that $\partial f_T/\partial\theta(\partial f_T/\partial m)$ is nonincreasing in $\theta(m)$, i.e., f_T is concave in θ and m. To verify existence and continuity in $\theta(m)$ of $\partial f_T/\partial\theta(\partial f_T/\partial m)$ we distinguish between the following three cases: $T \leqslant \lambda \leqslant v/m$; $\lambda < T \leqslant \theta/m$, and $\lambda \leqslant v/m < T$. In each of these cases, it suffices to establish that $\partial^+ f_T/\partial\theta = \partial^- f_T/\partial\theta$ and $\partial^+ f_T/\partial m = \partial^- f_T/\partial m$ at the points (θ^0, m^0) for which $\theta^0 + c$ is a breakpoint of f_T.

b. Here again, we distinguish between the following three cases: If $T \leqslant \lambda$ then,

$$h_T(\Theta, m) = \begin{cases} (\Theta + mc)/(m\lambda) + m(h' + h_0)\lambda & \text{if } \Theta + mc < \lambda^2 m^2(h' + h_0) \\ 2[(\Theta + mc)(h' + h_0)]^{1/2} & \text{if } \lambda^2 m^2(h' + h_0) \leqslant \Theta + mc < v^2(h' + h_0) \\ (\Theta + mc)/v + (h' + h_0) & \text{if } v^2(h' + h_0) < \Theta + mc. \end{cases}$$

If $\lambda < T \leqslant v/m$:

$$h_T(\Theta, m) = \begin{cases} (\Theta + mc)/(m\lambda) + mh'\lambda + mh_0T & \text{if } \Theta + mc < m^2h'\lambda^2 \\ 2[(\Theta + mc)h']1/2 + mh_0T & \text{if } m^2h'\lambda^2 \leqslant \Theta + mc < T^2m^2h' \\ (\Theta + mc)/(mT) + m(h' + h_0)T & \text{if } T^2m^2h' \leqslant \Theta + mc \leqslant T^2m^2(h' + h_0) \\ 2[(\Theta + mc)(h' + h_0)]^{1/2} & \text{if } T^2m^2(h' + h_0) < \Theta + mc \leqslant v^2(h' + h_0) \\ (\Theta + mc)/v + v(h' + h_0) & \text{if } v^2(h' + h_0) < \Theta + mc, \end{cases}$$

and if $T \geqslant v/m$:

$$h_T(\Theta, m) = \begin{cases} (\Theta + mc)/m\lambda + mh'\lambda + mh_0T, & \text{if } \Theta + mc < m^2h'\lambda^2 \\ 2[(\Theta + mc)h']^{1/2} + mh_0T, & \text{if } m^2h'\lambda^2 \leqslant \Theta + mc \leqslant v^2h' \\ (\Theta + mc)/v + h'v + mh_0T, & \text{if } v^2h' < \Theta + mc. \end{cases}$$

Since $\partial f_T/\partial\theta$ exists everywhere for the three cases (see the proofs of part a) it follows from the chain rule that $\partial h_T/\partial\Theta$ exists everywhere as well, moreover, $\partial h_T/\partial\Theta$ is continuous in m. Since $\partial^2 h_T/\partial m\partial\Theta$ exists and is nonpositive almost everywhere, it follows that $\partial h_T/\partial\Theta$ is nonincreasing in m; hence h_T has antitone differences.

To establish the continuity of $\partial h_T/\partial\Theta$ with respect to m, it suffices to verify that $\lim_{m\downarrow m_0}\partial h_T/\partial\Theta = \lim_{m\uparrow m_0}\partial \mathrm{h}_T/\partial\Theta$ for all points (Θ^0, m^0) for which $\Theta^0 + m^0c$ is a breakpoint of h_T. This is done by computing $\partial h_T/\partial\Theta$ for h_T as defined above and verifying that at each breakpoint of $\Theta^0 + m^0c$, the left limit value equals the right limit value.

c. This part follows immediately from parts a and b, f_T nondecreasing in θ for all $T > 0$, and Theorem 5 in Anily and Federgruen (1991a).

REFERENCES

ANILY, S. 1987. Integrating Inventory Control and Transportation Planning. Ph.D. Dissertation, Columbia University, New York.

ANILY, S., AND A. FEDERGRUEN. 1990a. One Warehouse Multiple Retailer Systems With Vehicle Routing Costs. *Mgmt. Sci.* **36,** 92–114.

ANILY, S., AND A. FEDERGRUEN. 1990b. A Class of Euclidean Routing Problems With General Route Cost Functions. *Math. Opns. Res.* **15,** 268–285.

ANILY, S., AND A. FEDERGRUEN. 1990c. Two-Echelon Distribution Systems With Vehicle Routing Costs and Central Inventories, Unabbreviated Version. Working Paper, Graduate School of Business, Columbia University, New York.

ANILY, S., AND A. FEDERGRUEN. 1991a. Structured Partitioning Problems. *Opns. Res.* **39,** 130–149.

ANILY, S., AND A. FEDERGRUEN. 1991b. Capacitated Two-Stage Multi-Item Production/Inventory Model With Joint Setup Cost. *Opns. Res.* **39,** 443–455.

ANILY, S., AND A. FEDERGRUEN. 1991c. Rejoinder to Comments on One Warehouse Multiple Retailer Systems With Vehicle Routing Costs. *Mgmt. Sci.* **37,** 1497 -1499.

CHIEN, T. W., A. BALAKRISHNAN AND R. T. WONG. 1989. An Integrated Inventory Allocation and Vehicle Routing Problem. *Trans. Sci.* **23,** 67–76.

CHRISTOFIDES, N. 1976. Worst-Case Analysis of a New Heuristic for the Traveling Salesman Problem. Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Penn.

DROR, M., AND M. BALL. 1987. Inventory/Routing: Reduction From an Annual to a Short-Period Problem. *Naval Res. Logist.* **34,** 891–905.

DROR, M., AND P. TRUDEAU. 1990. Split Delivery Routing. *Naval Res. Logist.* **37,** 383–402.

GALLEGO, G., AND D. SIMCHI-LEVI. 1990. On the Effectiveness of Direct Shipping Strategy for the One Warehouse Multi-Retailer R-Systems. *Mgmt. Sci.* **36,** 240–243.

HAIMOVICH, M., AND A. RINNOOY KAN. 1985. Bounds and Heuristics for Capacitated Routing Problems. *Math. Opns. Res.* **10,** 527–542.

HALL, R. 1991. Comments on One Warehouse Multiple Retailer Systems With Vehicle Routing Costs. *Mgmt. Sci.* **37,** 1496–1497.

MITCHELL, J. S. B. 1987. 98% Effective Lot Sizing for One Warehouse Multiretailer Inventory System With Backlogging. *Opns. Res.* **35,** 399–404.

QUEYRANNE, M. 1987. Finding 94% Effective Policies in Linear Time for Some Production/Inventory Systems. Working Paper, University of British Columbia, Vancouver.

ROUNDY, R. 1985. 98% Effective Integer Ratio Lot-Sizing for One Warehouse Multiretailer Systems. *Mgmt. Sci.* **31,** 1416–1430.