Distributing Nonstorable Items Without Transshipments

ANTHONY F. W. HAN and CARLOS F. DAGANZO

ABSTRACT

The research reported in this paper attempted to find optimal strategies for distributing items from one depot to many demand points without transshipments and within a limited amount of time. The objective was to find a near optimal partition of the region supplied by the depot into districts (the zones containing the points visited by one delivery route) and corresponding shipment sizes and costs. Initially, the average distribution cost per demand point on a single delivery route was studied using expressions that relate route length to the dimensions of a delivery district. Two routing strategies were considered: one that generates tours with nearly minimal local distance per point and another that generates tours with nearly minimal line-haul distance per point. Formulas were derived to estimate the optimal shipment size, district shape, and cost when the strategy yielding the least cost per point is used. Finally, the results were applied to develop guidelines for partitioning a whole supply region into nearly optimal districts; an example is given. For a constant demand density, an optimal district partition of the supply region should have bigger and fatter districts near the depot and smaller and thinner ones along the boundary of the region.

This research focuses on minimizing the cost of distributing nonstorable items (goods that must be delivered within a limited amount of time) from one depot to many demand points without transshipments. The distribution costs considered include driver wages, vehicle depreciation, and operating cost. Examples include not only perishable goods (fruits, vegetables, etc.) but also newspapers and parcels delivered through express mail or other express services.

One-to-many distribution problems with multiple tours (routes) are usually known as "single depot vehicle routing" problems. Substantial literature exists on minimizing transportation costs for vehicle routing problems [see Turner, Ghare, and Fourds (1) and Golden, Magnanti, and Nguyen (2) for a review]. Existing vehicle routing methods include the savings algorithm developed by Clarke and Wright (3), the "cluster first, route second" method by Tyagi (4), the sweep algorithm by Gillett and Miller (5), and the "seed first, route second" algorithm by Fisher and Jaikumar (6). These earlier works, however, are not concerned with the time required for delivery; they do not apply to the distribution of nonstorable items.

This analysis starts with a single delivery district (i.e., the area containing all demand points served by a single vehicle route). The district is assumed to be rectangular. The spatial density of demand points rather than their exact locations is considered. This eliminates the need to specify a network and allows detailed routing arrangements to be ignored.

The dual-strip strategy, a routing strategy that can generate tours of nearly minimal distances (7), is considered first. For this routing strategy, the district dimensions and shipment size that minimize the average distribution cost per point are derived. The optimal cost per point consists of three components: first, the per stop cost; second, the average local operating cost, which depends on the local distance traveled per point; and third, the average fixed-plus-line-haul cost, which depends mainly on the distance from the district to the depot and on the shipment size (or the number of points served by the vehicle). Because the dual-strip routing strategy yields a nearly minimal average local distance per point, it is appropriate for use when the local operating cost is the major component of the total delivery cost.

An alternative routing strategy, the single-strip strategy, is also considered. Although this strategy yields longer distances, it allows a nearly maximal number of points to be served within a given amount of time and reduces the line-haul cost per point. Thus it is preferred when the average fixed-plus-line-haul cost is the major component of the total delivery cost.

Comparison of the delivery costs of the two routing strategies shows that dual-strip routing should be applied when the delivery district is close to the depot (or, more precisely, if the local operating cost is larger than one-half of the fixed-plus-line-haul cost). Otherwise, single-strip routing is preferred. Then the overall optimum shipment size, district shape, and cost for the best of the two routing strategies in any given situation are derived.

The results are applied to develop guidelines for partitioning a large region supplied by one depot into nearly optimal districts. An example, in which a circular region that contains more than 1,000 points is partitioned into more than 100 districts, is given to demonstrate how the guidelines can be used.

The formulas developed in this paper can be used for sensitivity analysis. This is illustrated in the final section in which the cost impacts of changes in available delivery time, vehicle speed, and vehicle capacity are analyzed and discussed.

SINGLE DELIVERY DISTRICT

Assume that on any given day items must be delivered to demand points (customers) that are independently

Institute of Transportation Studies, University of California, Berkeley, Calif. 94720. Current address for A.F.W. Han: Department of Transportation Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan.

and randomly scattered in a region. On the next day (or after any other operation cycle), whether or not the number and location of the customers stay the same, service must be provided again. Distribution strategies are derived for one day; the strategies may or may not change daily.

Situations in which the density of customers (δ) varies spatially but is nearly constant within each delivery district are considered. Furthermore, it is assumed that the same number of items is required at all demand points and that the items to be distributed on a given day become available simultaneously at the depot. At that time the locations of all the customers (for that day) are known and distribution begins. Each vehicle must visit all customers in its district within a limited amount of time (τ_0) after the beginning of distribution. It is also assumed that vehicles are large enough to hold all the items that can be delivered in time τ_0. Note that, although these assumptions appear to be restrictive, the results of this paper can be applied to situations that are more general than those described here. This will be discussed in the last section of the paper.

Consider a rectangular district of sides ℓ and ℓ' ($\ell' > \ell$), as shown in Figure 1, where the distance between the depot and the gravity center of the district is ρ [$\rho > (\ell'/2)$]. Let

- C_0 = fixed cost (per day) of the delivery vehicle;
- C_d = delivery vehicle operating cost per unit distance;
- L = average route length (i.e., average total distance traveled per day by the delivery vehicle);
- D = average distance traveled from the depot to the last delivery point;
- S = time consumed at each demand point;
- U = average speed of the delivery vehicle;
- χ = average number of demand points contained in a square with sides equal to ℓ, $\chi = \delta\ell^2$; and
- N = shipment size of the delivery vehicle in terms of the number of demand points visited, $N = \delta\ell\ell'$.

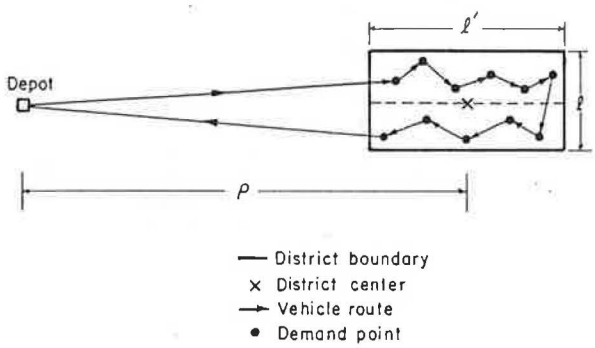

FIGURE 1 Delivery district with dual-strip route.

For the time being, it is assumed that vehicles are large enough that N can be as large as desired. Note that the district dimensions, ℓ and ℓ', can be expressed in terms of N and χ [i.e., $\ell = (\chi/\delta)^{1/2}$ and $\ell' = N/(\delta\chi)^{1/2}$]. N and χ can be thought of as dimensionless indicators of district width and district area. The use of such dimensionless variables, as it will be shown later, allows the development of numerical figures or tables that are applicable to different situations.

The design of the shortest route is not of concern; an attempt is made to find a near optimal partition of the region supplied by the depot into districts (the zones containing the points visited by one delivery route) and the corresponding shipment sizes and costs. Therefore expressions that relate route lengths (i.e., L and D) to district dimensions (in terms of N and χ) will be used when nearly optimal routes are used. Unless otherwise specified, the Euclidean distance will be used throughout this paper.

Early shortest tour length formulas (8,9) did not take district dimensions into account. Expressions that are sensitive to district shape have only been developed recently (7,10).

The routing strategy used by Daganzo (7), termed the "dual-strip strategy" in this paper, yields tours such as the ones shown in Figure 1 in which the rectangular district is divided into two equally wide strips. In each strip the delivery vehicle visits demand points from one end to another without backtracking. Daganzo (7) has shown that the average route length (L) can be expressed as a function of N and χ:

$$L(N,\chi) \cong 2\rho + N\delta^{-1/2}\,\phi(\chi) \qquad \rho > \ell'/2 \tag{1}$$

where

$$\phi(\chi) = (\chi^{1/2}/6) + (1/\chi^{1/2})\left(4/(\chi/4)^2\left\{[(1+(\chi/4)]\log[1+(\chi/4)] - (\chi/4)\right\} - 1\right) \tag{2}$$

The two terms in Equation 1 represent the line-haul distance and the local delivery distance, respectively.

Because the last demand point in the delivery district must be covered in time τ_0, the average distance (D) from the depot to the last delivery point has to be derived for the analysis. The difference between D and the whole route length (L), the back-haul distance, is approximately $\rho - \ell'/2 = \rho - N/[2(\delta\chi)^{1/2}]$. Thus,

$$D(N,\chi) \cong L(N,\chi) - \rho + N/[2(\delta\chi)^{1/2}] \qquad \rho > \ell'/2$$

$$= \rho + N\delta^{-1/2}\,\psi(\chi) \qquad \rho > \ell'/2 \tag{3}$$

where

$$\psi(\chi) = \phi(\chi) + 1/(2\chi^{1/2}) \tag{4}$$

The object now is to minimize the average cost of serving one demand point, that is,

$$c(N,\chi) = [C_0 + C_d L(N,\chi)]/N \tag{5}$$

Because this expression decreases with N, N should be chosen to be as large as possible. It has been assumed that vehicle capacity does not restrict N; however, the time constraint does. The number of stops should satisfy the following delivery time constraint:

$$NS + [D(N,\chi)]/U \leq \tau_0 \tag{6}$$

The left side of this inequality increases with N. Consequently, the optimum (minimum) cost (c^*) is obtained when N is so large that no more demand points can be visited within time τ_0. If N is approximated by a continuous variable, it should satisfy Equation 6 strictly:

$$NS + [D(N,\chi)]/U = \tau_0 \tag{7}$$

Thus substituting Equation 3 for $D(N,\chi)$ in Equation 7, and solving for N, the optimal shipment size (N) is obtained for a given zone width (χ):

$$N(\chi) = (\tau_0 - \rho/U)/[S + \delta^{-1/2}\,\psi(\chi)/U] \qquad \rho < U\tau_0 \tag{8}$$

In this expression the numerator and the denominator can be interpreted, respectively, as the time available for local delivery and the average time required to cover one demand point.

Before expressing the cost (c) as a function of χ, let us define the following dimensionless constant:

$$g = (C_0 + 2\rho C_d)/[(U\tau_0 - \rho)\,C_d] \qquad \rho < U\tau_0 \tag{9}$$

The term $U\tau_0 - \rho$ can be visualized as the maximum local distance that can be traveled within time τ_0; it will be called the local range. The parameter g thus can be interpreted as a ratio between the fixed-plus-line-haul cost and the operating cost required to cover the local range. If the fixed cost is considered a part of the line-haul expenses, g can legitimately be called the line-haul-to-local cost ratio. Note that g increases to infinity as ρ approaches $U\tau_0$ because then the local range (and the local cost) goes to zero. The line-haul-to-local cost ratio thus also indicates the district's distance from the depot.

The optimal cost per item [$c(\chi)$] is obtained by replacing N in Equation 5 by Equation 8. It can be written as

$$c(\chi) = C_d USg + C_d\delta^{-1/2}\,[\phi(\chi) + g\psi(\chi)] \tag{10}$$

As shown by this equation, the average cost per item has three components. The first, C_dUgS, is the portion of fixed-plus-line-haul cost per item associated with the time lost at one stop; fewer items can be carried in the time allowed because of this lost time. The second component, $C_d\delta^{-1/2}\phi(\chi)$, is the local vehicle operating cost per point and is proportional to the local distance traveled per point, $\delta^{-1/2}\phi(\chi)$. The third cost term, $C_d\delta^{-1/2}g\psi(\chi)$, is similar to the first; it is the portion of fixed-plus-line-haul cost per point that arises because vehicles do not travel infinitely fast and can only carry a finite number of items.

The problem now becomes one that has a single decision variable, χ. The optimal width, χ^*, is the one that minimizes $c(\chi)$ in Equation 10 or, simply, $f(\chi) = \phi(\chi) + g\psi(\chi)$. Let χ_1 and χ_2 be the solutions that minimize $\phi(\chi)$ and $\psi(\chi)$, respectively: $\chi_1 = 6.7$ and $\chi_2 = 9.2$.

Although a delivery district with width $\ell = 6.7^{1/2}\delta^{-1/2}$ yields a nearly minimal local distance traveled per point (as well as a nearly minimal local vehicle operating cost per point), a slightly wider district with $\ell = 9.2^{1/2}\delta^{-1/2}$ allows a nearly maximal number of points to be covered within time τ_0; it yields approximately the lowest average fixed-plus-line-haul cost per point. For districts near the depot, χ^* would be expected to be closer to χ_1; and for remote districts, χ^* should be close to χ_2. The following analysis confirms this expectation. Because both $\phi(\chi)$ and $\psi(\chi)$ are convex, $\chi^* \varepsilon [6.7, 9.2]$ for $0 < g < \infty$. As the aggregate line-haul-to-local cost ratio, g (i.e., the distance from the depot), increases, χ^* moves from the left to the right in [6.7, 9.2]. As the district distance from the depot increases, the optimal district becomes gradually wider, allowing more demand points to share the aggregate line-haul cost, although yielding a somewhat longer local distance traveled per point. These adjustments to district width are not very substantial; when $g \to \infty$, ℓ is only 17 percent larger than when $g \to 0$. The impact on cost of departures from the optimal ℓ is examined in the next section.

Because dual-strip routing can yield tours with nearly minimal local distances per point, it is appropriate for use when g is small and cost per point depends primarily on the local distance traveled. When g is large, radically different routing strategies may be better.

ALTERNATIVE ROUTING STRATEGY

Let us now consider the single-strip routing strategy shown in Figure 2. This strategy allows the delivery vehicle to serve more demand points within a given amount of time and thus reduces the average fixed-plus-line-haul cost per point. [To see this, simply consider a district half as wide but twice as long as for dual-strip routing and with the same center of gravity. The distance traveled between points is the same (on average) in both cases, but for single-strip routing the distribution stage begins and ends sooner. Additional points can thus be served.] Single-strip routing can be appropriate when the line-haul-to-local cost ratio is high. A recent study (11) also shows that, for distributing valuable goods, the single-strip strategy is better than the dual-strip strategy; these authors used an L_1 metric for their calculations.

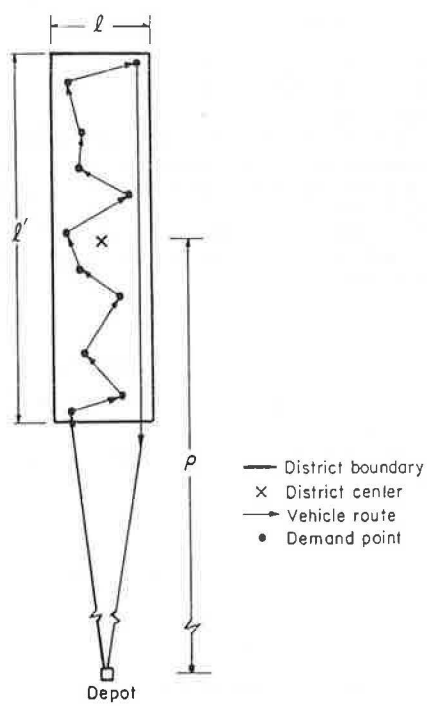

FIGURE 2 Single-strip routing.

Expressions for tour lengths, shipment size, and cost can also be derived for single-strip routing with a Euclidean metric (see the Appendix). An additional subscript (s) is used to denote single-strip variables and functions. All have a form similar to that given previously. For example, when the zone width (χ) is given, shipment size [$N_s(\chi)$] and cost [$c_s(\chi)$] are

$$N_s(\chi) = (U\tau_0 - \rho)/[US + \psi_s(\chi)\delta^{-1/2}] \qquad \rho < U\tau_0 \tag{11}$$

and

$$c_s(\chi) = C_d\left\{USg + \delta^{-1/2}\,[\phi_s(\chi) + g\psi_s(\chi)]\right\} \tag{12}$$

where

$$\phi_s(\chi) = (\chi^{1/2}/3) + (1/\chi^{1/2})\left\{(2/\chi^2)\,[(1+\chi)\log(1+\chi) - \chi]\right\}$$

and

$$\psi_s(\chi) = \phi_s(\chi) - 1/(2\chi^{1/2})$$

Note that $c_s(\chi)$, like $c(\chi)$ as given by Equation 10, also has three cost components: the cost per stop, the average local operating cost, and the cost of time constraint.

Unlike dual-strip routing, however, the minimum of $\psi_s(\chi)$, 1.9, is smaller than the minimum of $\phi_s(\chi)$, 2.7. This happens because $\psi_s(\chi)$ is obtained from $\phi_s(\chi)$ by subtracting a decreasing function. Thus with single-strip routing an optimal district becomes narrower instead of wider as the distance from the depot (g) increases.

Figure 3 shows how χ_s^* moves from 2.7 to 1.9 as g is increased; it compares χ_s^* and χ^* as well. The figure also reveals that, when $g \sim 2$, $\chi^* = 4\chi_s^*$. That is, single-strip districts should be half as wide as dual-strip districts; both routing schemes should use equally wide strips. This is approximately true for all the g's that can occur in practice. Even in extreme cases, when $g \to 0$ or $g \to \infty$, the optimal strip widths of the two routing strategies differ by less than 30 percent.

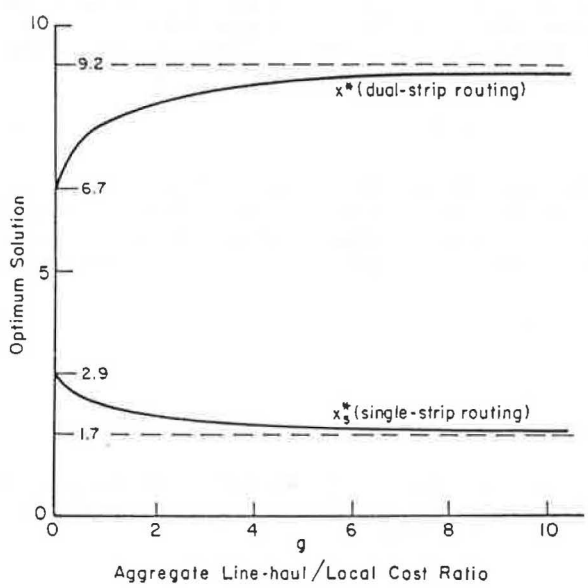

FIGURE 3 Optimum solutions of two routing strategies.

To compare the cost of the two routing strategies, f^* and f_s^* are plotted in Figure 4; they cross each other at the critical point $g_c \approx 2$ (this is exact for the L_1 metric). Thus the single-strip strategy should be applied when the aggregate line-haul-to-local ratio (g) is larger than 2 (this also implies that the strategy with the narrowest optimal strip is best). The farther g is from g_c, the more important it is to choose the proper strategy. For example, if the per stop cost is ignored, single-strip routing can reduce the delivery cost of dual-strip routing by about 10 percent when g = 4 and by about 24 percent when $g \geq 100$. For road transportation, g is most likely in the range of $0 < g < 50$ (12).

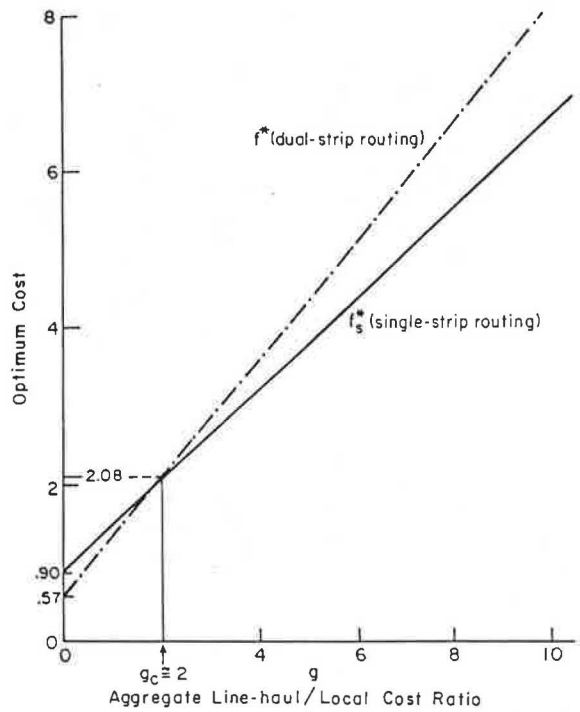

FIGURE 4 Optimum costs of two routing strategies.

According to Equation 9, a switch should be made from dual to single-strip with dual-strip routing when the (critical) distance between the depot and the district center of gravity (ρ_c) is

$$\rho_c = (2\,U\tau_0 - C_0/C_d)/4 \qquad \rho_c \geq 0 \tag{13}$$

Note that ρ_c does not depend on S. For $0 < \rho < \rho_c$, dual-strip routing should be used; beyond this range, single-strip routing is best. When the fixed cost (C_0) is zero, $\rho_c = 0.5U\tau_0$ (i.e., the critical distance is half the distance that can be traveled in time τ_0). As C_0 increases, ρ_c decreases; the application region of dual-strip routing shrinks. When $C_0 > 2U\tau_0 C_d$, only single-strip routing should be used. Figure 5 illustrates these phenomena.

OPTIMAL COST, SHIPMENT SIZE, AND DISTRICT DIMENSIONS

Let us now examine in more detail the properties of the optimal solution. Let us first define the overall optimum cost $\hat{c} = \min[c^*, c_s^*]$, which results from the best use of the two routing strategies considered. [Although hybrid strategies have dual-strips for only part of the way and strips of variable length can reduce the cost below $\hat{c}$, these reductions appear to be insignificant (11); hybrid strategies are not considered in this paper.] The circumflex is placed above any variable corresponding to $\hat{c}$. Let

$\hat{f} = \min[f^*, f_s]$; that is,

$$\hat{f} = f^* = f(\chi^*) \qquad \text{if } g < 2 \text{ (or } \rho < \rho_c)$$

and

$$\hat{f} = f_s^* = f_s(\chi_s^*) \qquad \text{if } g > 2 \text{ (or } \rho > \rho_c) \tag{14}$$

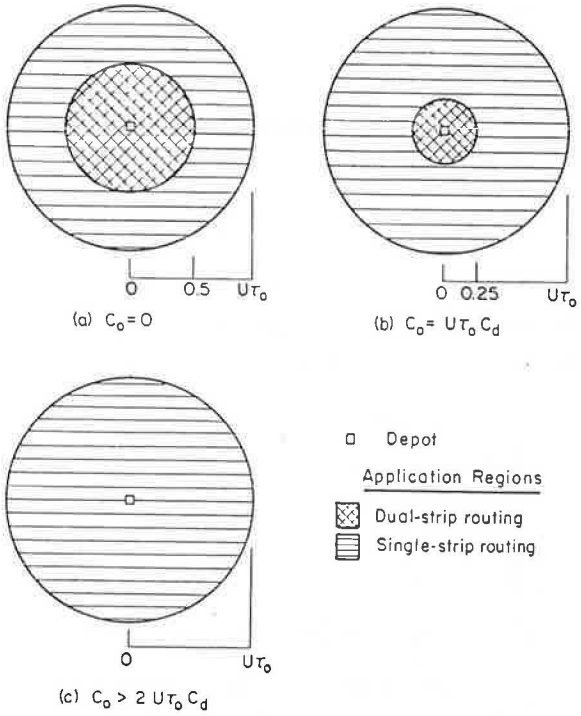

FIGURE 5 Application regions of two routing strategies.

For any g, the functions $f(\chi)$ and $f_s(\chi)$ are very flat around their minima. Thus χ^* and χ_s^* do not need to be chosen very precisely (as in Figure 3); simpler rules can be followed. For example, if dual-strip routing with $\chi^* = 8.0$ is chosen when $g < 2$ and single-strip routing with $\chi_s^* = 1.9$ is chosen when when $g > 2$, the resulting values of $f(\chi)$ and $f_s(\chi_s)$ only deviate from $f(\chi^*)$ and $f_s(\chi_s^*)$ by less than 0.5 percent (see Table 1). These approximations ($\chi^* \sim 8.0$ and $\chi_s^* \sim 1.9$) are reasonable; they will be used from now on. The $\hat{f}$ can be approximated as

$$\hat{f} \cong f(8.0) = 0.576 + 0.753g \quad \text{if } g < 2$$

$$\cong f_s(1.9) = 0.937 + 0.574g \quad \text{if } g > 2 \tag{15}$$

and $\hat{c}$ can be written as follows

$$\hat{c}(\rho) \cong \left\{ [\alpha_1(C_0 + 2\rho C_d)] / (U\tau_0 - \rho) \right\} + 0.576\delta^{-1/2} C_d \quad \text{if } \rho < \rho_c$$

$$\cong \left\{ [\alpha_2(C_0 + 2\rho C_d)] / (U\tau_0 - \rho) \right\} + 0.937\delta^{-1/2} C_d \quad \text{if } \rho > \rho_c \tag{16}$$

where

$$\alpha_1 = US + 0.753\delta^{-1/2} \tag{17}$$

and

$$\alpha_2 = US + 0.574\delta^{-1/2} \tag{18}$$

Note that α_1 and α_2 are the distances that the vehicle can cover during the time it takes to serve one point (US plus the local distance per point). Figure 6 (bottom) shows how $\hat{c}$ depends on distance; except at the critical point, where $\rho = \rho_c$, the optimum cost ($\hat{c}$) increases at an increasing rate. This phenomenon does not occur for storable items; the cost of distributing storable items increases at a decreasing rate with distance (13). With storable items, inventory-plus-transportation cost is minimized when the largest loads are dispatched to the remotest districts. This cannot be done with non-storable items; when ρ increases, the local range decreases and fewer points can be covered.

The overall optimum shipment size ($\hat{N}$) becomes

$$\hat{N}(\rho) \cong (U\tau_0 - \rho)/\alpha_1 \quad \text{if } \rho < \rho_c$$

and

$$\cong (U\tau_0 - \rho)/\alpha_2 \quad \text{if } \rho > \rho_c \tag{19}$$

These expressions are reasonable; both represent the ratio of the local range to the distance spent per point. Figure 6 (top) shows plots of $\hat{N}(\rho)$. As was just discussed, $\hat{N}(\rho)$ decreases (linearly) with ρ, except at the point of discontinuity ($\rho = \rho_c$). The vehicle load can be increased at this point because the switch from dual- to single-strip routing advances the time of the last delivery.

The variables V_d and V_s, defined in Figure 6 (top), represent the largest load that is carried with either routing strategy. There is no guarantee that $V_s > V_d$ (as in the figure).

The optimal size and district dimensions as functions of the distance ρ can also be derived. Such expressions, as it will be shown later, are useful for partitioning a region into nearly optimal districts. The optimal district size ($\hat{A}$) is

$$\hat{A}(\rho) = \hat{N}(\rho)/\delta \quad \rho < U\tau_0 \tag{20}$$

$\hat{A}$ exhibits the same properties as $\hat{N}(\rho)$.

The width of a delivery district is given by $\ell = (\chi/\delta)^{1/2}$. Thus similar to χ^* and χ_s^*, $\hat{\ell}$ remains constant when $\rho < \rho_c$ and when $\rho > \rho_c$. The expressions are

$$\hat{\ell} = 2.83\delta^{-1/2} \quad \text{if } \rho < \rho_c$$

and

$$= 1.38\delta^{-1/2} \quad \text{if } \rho_c < \rho < U\tau_0 \tag{21}$$

Districts are about half as wide when single-strip routing is used. In both cases the strip should be about $1.4\delta^{-1/2}$ distance units wide.

TABLE 1 Percentage Errors in Optimum Cost

$0 \leq g \leq 2$				$g \geq 2$			
	①	②	$\frac{① - ②}{②} \times 100\%$				
g	f(8)	f^*	Δ%	g	$f_s(1.9)$	f_s^*	Δ%
0	0.57599	0.57522	0.66	2	2.0849	2.0803	0.22
1	1.3287	1.3287	0	4	3.2331	3.2331	0
2	2.0815	2.0803	0.06	10	6.6776	6.6733	0.06
				100	58.345	58.178	0.28
				1,000	575.02	573.19	0.32

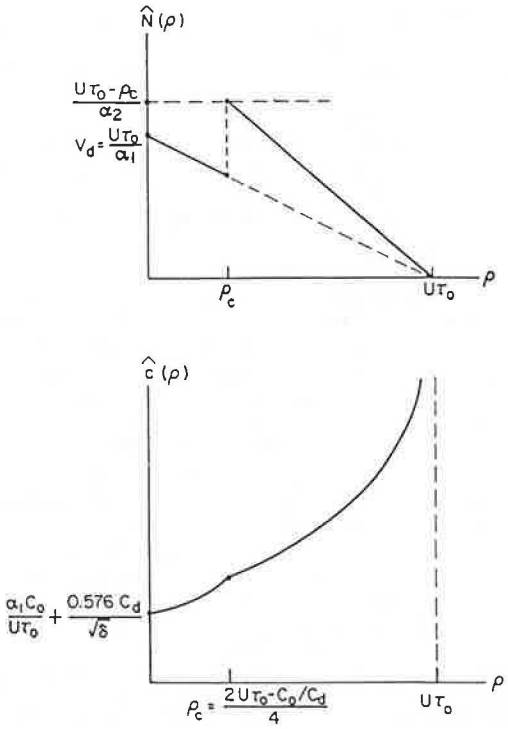

FIGURE 6 Optimal shipment size and cost.

The length of a delivery district is given by $\ell' = A/\ell$. Thus

$$\hat{\ell}' = (0.35)\,[(U\tau_0 - \rho)/\alpha_1]\,\delta^{-1/2} \quad \text{if } \rho < \rho_c$$

and

$$= (0.73)\,[(U\tau_0 - \rho)/\alpha_2]\,\delta^{-1/2} \quad \text{if } \rho_c < \rho < U\tau_0 \tag{22}$$

These expressions follow the same pattern as $\hat{N}(\rho)$. They decrease linearly with ρ, except for a jump when $\rho = \rho_c$. For a given distance, single-strip districts are 2.1 (α_1/α_2) times as long as dual-strip districts.

The shape of a rectangular district can be represented by the ratio of its width and length, $\beta = \ell/\ell'$. This ratio was called the slenderness factor by Daganzo (7). From Equations 21 and 22

$$\hat{\beta}(\rho) \cong 8.0\alpha_1/(U\tau_0 - \rho) \quad \text{if } \rho < \rho_c$$

and

$$\cong 1.9\alpha_2/(U\tau_0 - \rho) \quad \text{if } \rho_c < \rho < U\tau_0 \tag{23}$$

For storable items, β remained constant with distance. Now, however, $\hat{\beta}(\rho)$ increases with ρ except, of course, when $\rho = \rho_c$.

SERVING A REGION: AN EXAMPLE

Consider now a region that contains many delivery districts and define the optimal district partition of the region as that which yields the minimum cost of serving the whole region. Although such a partition is difficult to derive, its desirable properties can be explored. Imagine an ideal district partition that is characterized by the following two properties:

P1. It is feasible; all districts pack well, cover the whole region, and each can be covered within time τ_0.

P2. For each district, both the size (A) and the slenderness factor (β) are optimal (i.e., $A = \hat{A}$ and $\beta = \hat{\beta}$).

Although such an ideal partition usually does not exist, a district partition that closely follows its properties should yield a cost that is not much larger than the minimum. [Problems involving the determination of the location of, or the spacing between, a set of points (depots, warehouses, bus stops, scheduled headways, etc.) usually have cost functions that are very flat near their optima (8, 14,15).] Therefore, in designing a desirable district partition, an attempt is made to follow P1 and P2 as much as possible. The following example shows how these guidelines can be applied.

Consider a circular region with a radius R and the depot located at its center. For such a region, P2 can be approximated with a ring-and-radial partition that satisfies P1. Let m be the number of equally big districts in the ring defined by two concentric circles with radii r_0 and r_1 that are such that $r_1 > r_0 > 0$ (Figure 7). Given r_0 or r_1, the other radius and $\hat{A}$ can be determined from the previously developed formulas. Specifically,

$$r_1 = r_0 + \hat{\ell}' \tag{24}$$

and

$$\hat{A} = [U\tau_0 - (r_0 + r_1)/2]/(\alpha_1\delta) \quad \text{if } r_0 + r_1 < 2\rho_c$$

and

$$= [U\tau_0 - (r_0 + r_1)/2]/(\alpha_2\delta) \quad \text{if } r_0 + r_1 > 2\rho_c \tag{25}$$

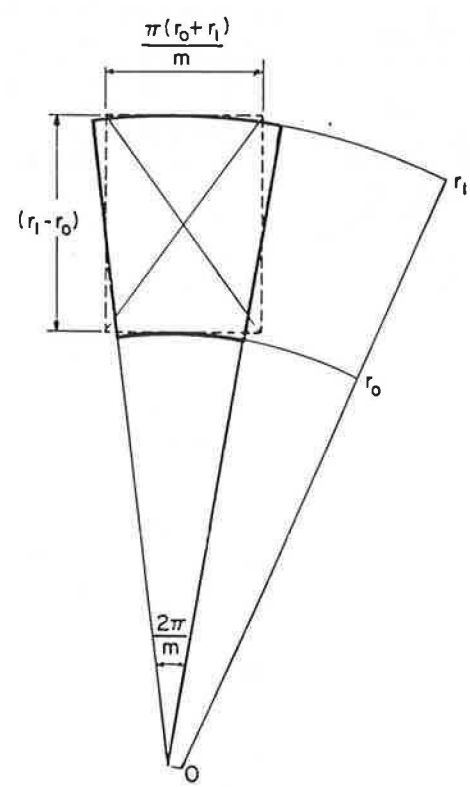

FIGURE 7 Partition sector.

The number of zones in a ring is

$$\hat{m} = \pi(r_1^2 - r_0^2)/\hat{A} \tag{26}$$

When the values $\hat{\ell}$ and $\hat{\ell}'$ are calculated for $\rho = (r_0 + r_1)/2$, Equation 24 yields

$$r_1 = r_0 + (U\tau_0 - r_0)k \tag{27a}$$

or

$$r_0 = (r_1 - U\tau_0 k)/(1 - k) \tag{27b}$$

where

$$k = (0.5 + 2.83\alpha_1\delta^{1/2})^{-1} \quad \text{if } r_0 + r_1 < 2\rho_c$$

and

$$k = (0.5 + 1.38\alpha_2\delta^{1/2})^{-1} \quad \text{if } r_0 + r_1 > 2\rho_c \tag{28}$$

These expressions are reasonable if the districts that result are approximately rectangular. For the first ring of districts, however, this is not the case. For storable items, first-ring districts should be approximately 40 percent longer than predicted for rectangular shapes (11); other rings do not need a correction. A similar phenomenon should occur now; it appears reasonable to increase $r_1 = kU\tau_0$ (see Equation 27a with $r_0 = 0$) by 40 percent while simultaneously maintaining $\hat{N}(\rho)$ and $\hat{A}(\rho)$ at the previous level (with $\rho = kU\tau_0/2$) for the first ring of districts. The optimal radius of the first ring thus should be approximately $1.4kU\tau_0$.

Suppose that customers and vehicles have the following characteristics: $U\tau_0 = 15$, $\delta = 4$, and $C_0/C_d = 10$ and consider two cases for US = 0 and US = 1. For a circle of radius R = 13, there are more than 2,000 customers.

Table 2 gives numerical results that show how (starting with $r_1 = 13$ and proceeding inwards) the sequence of ring radii can be obtained for both cases by repeated use of Equation 27b. Usually, as in the example, one of the innermost dual-strip rings will have $r_1 \sim 1.4kU\tau_0$. Any rings inside this ring should be eliminated. The resulting partition should be nearly optimal, even when the first radius is significantly different from $1.4kU\tau_0$; the first ring usually does not contain a large portion of all the customers, nor does it account for a large fraction of the vehicle-miles. Still, if desired, the boundary between the first and second rings can be shifted a little so that the inaccuracy in zone lengths is spread over two rings; Newell and Daganzo (11) discuss this for storable items. Table 2 also gives the (unrounded) $\hat{N}$, $\hat{A}$, and $\hat{m}$ corresponding to each ring.

Figures 8 and 9 show the districting patterns that result. Figure 10 shows the two typical routing patterns for $\rho < \rho_c (\rho_c = 5)$ and $\rho > \rho_c$ when US = 0. For US = 0, districts are larger and more elongated than for US = 1. Vehicles also make more stops. When US = 1, more rings are needed to cover the same area than when US = 0.

In cases in which the average number of stops per district calculated for a ring ($\hat{N}$) is not large (as with the outer rings for US = 1), the boundaries of the ring should be modified so that the resulting N is an integer. [Note in particular that an $\hat{N}$ smaller than 1 cannot be used. In some instances, the farthest customers may receive individual service.] A similar modification is needed to make the number of districts in a ring ($\hat{m}$) an integer when $\hat{m}$ is small. These modifications can be made as the calculations for Table 2 are being done or can be left to experienced judgment. Human intervention is hard to avoid in any case (7). For example, when customer locations change daily, the most cost-effective way of defining the final routes is often through human dispatching. The dispatcher can follow the guidelines, but he must ensure that districts pack, that the time constraint is not likely to be violated, and that routes are network feasible and balanced.

DISCUSSION

The example in the previous section illustrated how the formulas developed in this paper can be used for operational planning purposes. Although the example was idealized, it is not difficult to see how realistic cases should be addressed. If the customer demand density varies, the district dimensions should change with it; a districting pattern that follows these dimensions closely, and yet fits within the irregular boundaries of a service region, can usually be found. [See Newell (14), Clarens and Hurdle (15), Daganzo (7), and Daganzo and Newell (16) for additional discussion of this issue and several examples.]

Seven properties of a near-optimal operations plan are

1. Districts should be elongated toward the depot.
2. Vehicles should cover districts near the depot with two laps (dual-strip routing) and districts far

TABLE 2 Numerical Results

r_1	r_0	$\rho = (r_1 + r_0)/2$	$\hat{N}(\rho)$	$\hat{A}(\rho)$	$\hat{m}$	Routing
US = 0 ($1.4kU\tau_0 = 8.2$)						
13	6.2	9.6	18.9	4.7	87.2	Single-strip
6.2	0.8	3.5	30.6	7.6	15.5	Dual-strip
0.8	NA[a]					Dual-strip
US = 1 ($1.4kU\tau_0 = 2.5$)						
13	12.3	12.7	1.8	0.5	115.5	Single-strip
12.3	11.4	11.9	2.4	0.6	108.1	Single-strip
11.4	10.2	10.8	3.3	0.8	98.5	Single-strip
10.2	8.6	9.4	4.3	1.1	85.8	Single-strip
8.6	6.5	7.6	5.8	1.4	68.8	Single-strip
6.5	3.7	5.1	7.7	1.9	46.5	Single-strip
3.7	2.1	2.9	8.8	2.2	13.0	Dual-strip
2.1	0.3	1.2	10.0	2.5	5.4	Dual-strip
0.3	NA[a]					Dual-strip

[a]NA = not applicable ($r_0 < 0$).

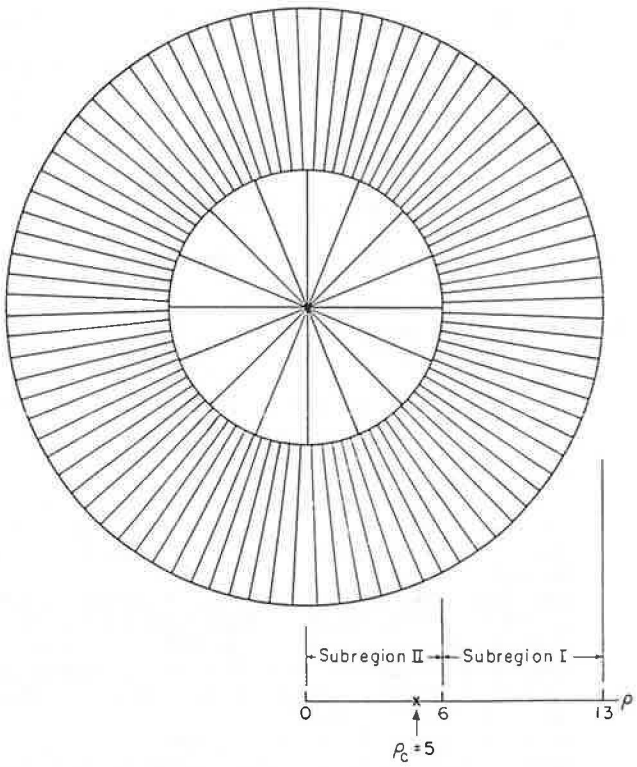

FIGURE 8 District partition, US = 0.

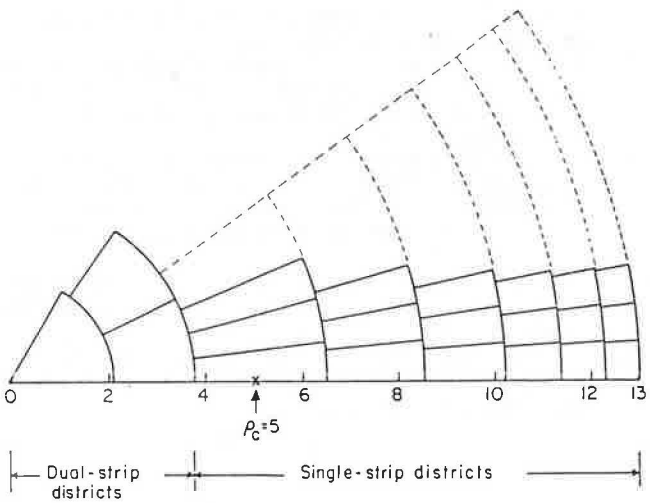

FIGURE 9 District partition, US = 1.

from the depot with only one outbound lap (single-strip routing).

3. If the fixed-plus-operating cost of a vehicle is proportional to distance, dual-strip routing should be used for districts the center of which can be reached in less than one-half the time available for delivery.

4. All districts of a given type, regardless of their locations relative to the depot, should have approximately the same width. Dual-strip districts should be about twice as wide as single-strip districts.

5. The number of stops, length, and size of both single- and dual-strip districts, however, should decline with the distance from the depot.

6. All else equal, an increase in the time needed per stop diminishes the size of the district that

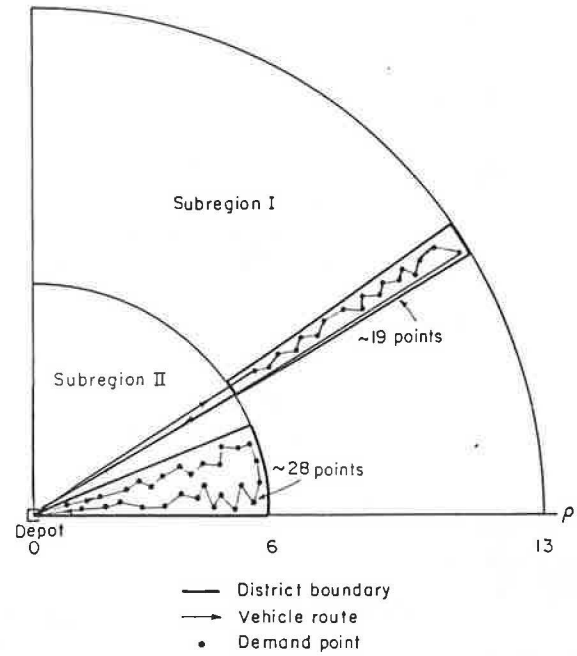

FIGURE 10 Two typical routing patterns.

can be covered but does not change either its width or the type of routing that should be used.

7. The cost per item increases at an increasing rate with distance.

For land transportation problems, where the Euclidean metric with constant speed everywhere is not reasonable, these principles still apply, albeit in a somewhat modified fashion. For example, "elongation toward the depot" should be interpreted to read "perpendicular to the equi-travel time contours." Newell and Daganzo (11) refined the principles and formulas for the case of storable items; a similar refinement can and should be sought for nonstorable items.

The formulas in this paper also quantify conveniently the cost impact of changes in demand, vehicle operating characteristics, and the service standard of a time-constrained distribution system. Therefore, they can be used for strategic planning purposes. For example, take the available delivery time (τ_0). Figure 11 shows how the entire cost curve shifts to the lower right when the available time increases from τ_0 to τ_0'. Vehicle speed (U) has a similar effect on optimum cost.

The capacity of the vehicles is another operating characteristic that affects cost. In this paper it was assumed that vehicles are large enough to hold all the items that can be delivered in time τ_0. When this is not reasonable, the expressions should be modified (12). Then, as the number of stops that can be made by a vehicle is reduced, the size of the region where dual-strip routing is preferred and the cost both increase. This is logical. The attractive feature of single-strip routing--that more stops can be made--is negated when vehicles are not large enough to make all the stops.

The results in this paper can be applied to scenarios more general than those described at the outset. For example, problems in which all the items distributed in one day are not produced simultaneously can be studied. If the items are not destination specific, loads can be made as soon as batches of the right size become available; there is no need for inventories at the depot. When the items

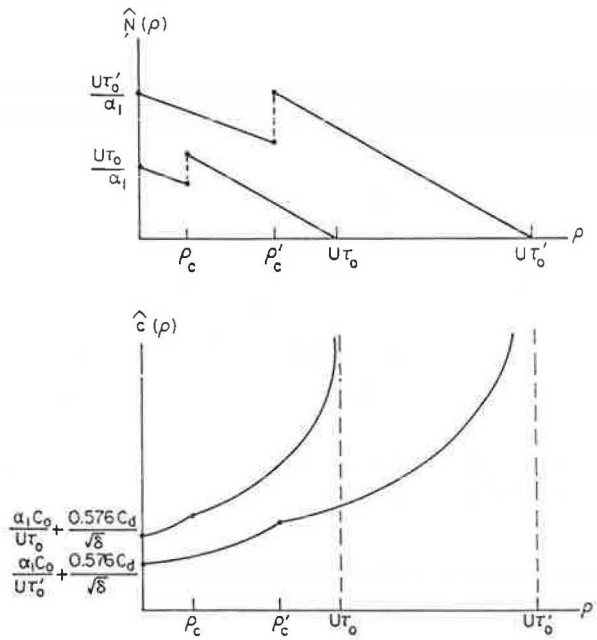

FIGURE 11 Influence of available delivery time.

shipped are perishable, the available delivery time for an item is measured from its time of production. Because all batches should be delivered in about the same amount of time, the strategies described in this paper apply verbatim. On the other hand, when the items have to be distributed to beat a deadline (e.g., newspapers and magazines as well as items taken to businesses that open and close at fixed hours), less time is available for the last batches that become available. In such cases the first batches should be sent to the remotest districts and the last loads to the nearest customers. To apply the results of this paper, the time available for delivery to each ring of the service area should first be determined. This is possible because the production schedule is known, each ring has a known demand, and rings are served from the outside in. The formulas presented in this paper can then be used to determine the best way of supplying each ring.

The research reported in this paper can be used as a building block for analyzing more complicated time-constrained distribution problems. Among possible applications are

- Determining either the optimal spacing between two adjacent depots (production plants or transshipment terminals) or the optimal location of additional depots, given the cost of setting up a depot, and
- Identifying optimal ways of distributing nonstorable items through transfer points.

ACKNOWLEDGMENT

This research was supported by a grant from the National Science Foundation to the University of California, Berkeley.

REFERENCES

1. W. Turner, P. Ghare, and L. Fourds. Transportation Routing Problem--A Survey. AIIE Transactions, Vol. 6, No. 4, 1974, pp. 228-301.
2. B. Golden, T. Magnanti, and H. Nguyen. Implementing Vehicle Routing Algorithms. Networks, Vol. 7, 1977, pp. 113-148.
3. G. Clarke and J.W. Wright. Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. Operations Research, Vol. 4, 1964, pp. 568-581.
4. M. Tyagi. A Practical Method for the Truck Dispatching Problem. Journal of the Operations Research Society of Japan, Vol. 10, 1968, pp. 76-92.
5. B. Gillett and L. Miller. A Heuristic Algorithm for the Vehicle Dispatch Problem. Operations Research, Vol. 22, 1974.
6. M.L. Fisher and R. Jaikumar. A Generalized Assignment Heuristic for Vehicle Routing. Networks, Vol. 11, 1981, pp. 109-124.
7. C.F. Daganzo. Distance Traveled to Visit N Points with a Maximum of C Stops per Vehicle: An Analytic Model and an Application. Transportation Science, Vol. 18, 1984, pp. 331-350.
8. S. Eilon, C.D.T. Waston-Gandy, and N. Christofides. Distribution Management: Mathematical Modeling and Practical Analysis. Hafner Press, New York, 1971.
9. D.M. Stein. An Asymptotic, Probabilistic Analysis of a Routing Problem. Mathematics of Operations Research, Vol. 3, No. 2, 1978, pp. 89-101.
10. C.F. Daganzo. The Lengths of Tours in Zones of Different Shapes. Transportation Research, Vol. 18B, 1984, pp. 135-145.
11. G. Newell and C.F. Daganzo. Design of Multiple Vehicle Delivery Tours. Transportation Research B (in press).
12. A. Han. One-to-Many Distribution of Nonstorable Items: Approximate Analytic Models. Ph.D. dissertation, Dissertation Series 84-3. University of California, Berkeley, 1984.
13. C.F. Daganzo and G.F. Newell. Physical Distribution from a Warehouse: Vehicle Coverage and Inventory Levels. Transportation Research, Vol. 19B, 1985, pp. 397-408.
14. G.F. Newell. Scheduling, Location, Transportation and Continuum Mechanics: Some Simple Approximations to Optimization Problems. SIAM Journal of Applied Mathematics, Vol. 25, No. 3, 1973, pp. 346-360.
15. G. Clarens and V.F. Hurdle. An Operating Strategy for a Commuter Bus System. Transportation Science, Vol. 9, 1975, pp. 1-20.
16. C.F. Daganzo and G.F. Newell. Configuration of Physical Distribution Networks. Networks, Vol. 16, 1986, pp. 113-132.

APPENDIX--Derivations Associated with Single-Strip Routing

Let d be the average Euclidean distance traveled between two succeeding points in the district. This distance is a function of strip width, which for single-strip routing is also the district width, ℓ (10).

$$d \cong (\ell/3) + (\delta\ell)^{-1} h(\delta\ell^2) \tag{A1}$$

where

$$h(\chi) = (2/\chi^2)\,[(1+\chi)\ln(1+\chi) - \chi] \qquad (\chi = \delta\ell^2 \text{ as defined in the text})$$

If the distances between the district boundary and the first and last points in the district are ignored,

$$L_s \cong N_s d + 2\rho \tag{A2}$$

and

$$D_s \cong N_s d + \rho - (\ell'/2) \tag{A3}$$

Substituting d in Equations A2 and A3 by Equation A1, and remembering that $\ell' = N_s/(\delta\chi)^{1/2}$ and $\ell = (\chi/\delta)^{1/2}$ yields

$$L_s \cong 2\rho + N_s\delta^{-1/2}\phi_s(\chi) \tag{A4}$$

and

$$D_s \cong \rho + N_s\delta^{-1/2}\psi_s(\chi) \tag{A5}$$

where

$$\phi_s(\chi) = (\chi^{1/2}/3) + (1/\chi^{1/2})\left\{(2/\chi^2)\,[(1+\chi)\ln(1+\chi) - \chi]\right\}$$

and

$$\psi_s(\chi) = \phi_s(\chi) - 1/(2\chi^{1/2})$$

For single-strip routing, the time constraint is

$$N_s S + (D_s/U) = \tau_0 \tag{A6}$$

Substituting D_s in Equation A6 by Equation A5 gives

$$N_s(\chi) = (U\tau_0 - \rho)/[US + \psi_s(\chi)\delta^{-1/2}] \quad \rho < U\tau_0 \tag{A7}$$

Replacing L and N in Equation 5 with L_s and N_s, Equations A4 and A5, respectively, yield

$$c_s(\chi) = C_d\left\{USg + \delta^{-1/2}[\phi_s(\chi) + g\psi_s(\chi)]\right\} \tag{A8}$$

where g, as defined in Equation 9, is the line-haul-to-local-cost ratio.

Equations A7 and A8 are the same as Equations 11 and 12, respectively.

Publication of this paper sponsored by Committee on Freight Transportation Planning and Marketing.