Method for Optimizing Transit Service Coverage

LAZAR N. SPASOVIC AND PAUL M. SCHONFELD

A method is presented for determining the optimal length of transit routes that extend radially from the central business district (CBD) into low-density suburbs. In addition to the route length, the route spacing, headway, and stop locations are also optimized. The equations for the route length, route spacing, headway, and stop spacing that minimize the sum of operator and user costs are derived analytically for many-to-one travel patterns with uniform passenger trip density. These equations provide considerable insight into the optimality conditions and interrelations among variables. The equations are also incorporated within an efficient algorithm that computes the optimal values of decision variables for a more realistic model with vehicle capacity constraints. The algorithm is applied to rectangular and wedge-shaped urban corridors with uniform and linearly decreasing passenger densities. The results show that in order to minimize the total cost, the operator cost, user access cost, and user wait cost should be equalized. At the optimum, the total cost function is rather shallow, thus facilitating the tailoring of design variables to the actual street network and particular operating schedule without substantial cost increases. The actual stop spacing pattern is determined for each corridor type. For a uniform passenger density, the stop spacing increases along the route in the direction of passenger accumulation toward the CBD. For a linearly decreasing passenger density, the stop spacing first decreases and then increases along the route toward the CBD. The sensitivity of design variables to some important exogenous factors is also presented.

One of the main problems in designing transit services is to provide appropriate transit service coverage and particularly to determine how far outward to extend transit routes into low-density suburbs. Service operators and users have somewhat conflicting objectives regarding the transit route length. Operators prefer short routes in order to minimize their costs. Passengers, especially those from the outer suburbs, prefer longer routes in order to minimize their access impedance. Since the route length has a significant impact on both operator costs and passenger impedance, its value should be carefully selected.

The purpose of this paper is to develop a method for optimizing the length of transit routes that extend radially outward from the central business district (CBD). However, this problem may not be considered independent of route location and service scheduling. Therefore, the problem considered here is finding optimal combinations of route length, route spacing, headway, and stop location and spacing that minimize the sum of operator and user costs for rectangular and wedge-shaped urban corridors with uniform and linearly decreasing passenger trip densities.

L. N. Spasovic, School of Industrial Management, Center for Transportation Studies and Research, New Jersey Institute of Technology, Newark, N.J. 07102. P. M. Schonfeld, Department of Civil Engineering, University of Maryland, College Park, Md. 20742.

LITERATURE REVIEW

Several previous studies sought to optimize various elements of transit network design and service using calculus and, to a lesser extent, mathematical programming methods (*1–18*). The summary of pertinent analytical models that are classified according to the design variable(s) optimized is presented in Table 1. The table shows that in most studies the travel demand was fixed and uniformly distributed over the service area. The usual travel pattern was many-to-one, which is typical for suburb-to-CBD commuting. The most common objective function was minimization of the sum of operator cost and user time cost.

A literature review revealed only one published paper (*15*) that optimized the radial length of a transit route in an urban transportation corridor, which is the focus of this research. Given the significant impact of the route length on cost, it is rather surprising that this research topic has not been given more attention in the literature. Wirasinghe and Seneviratne presented an analytical method for deriving the optimal length of a rail transit line in an urban corridor currently served by bus (*15*). The objective function to be minimized included the total cost of rail fleet, rail and bus operating cost, and passenger time cost. The authors found that for nonuniform rail line cost there could be several line lengths at which the total transit system cost is locally minimized (or maximized). For uniform rail line cost, an optimal line length existed if the net gain in travel time and operating cost of transporting the total demand a unit distance by rail when compared with bus exceeded the marginal line and fleet cost per unit length. The authors developed closed-form solutions for the line length for sectorial and rectangular corridors with uniformly distributed demand.

That paper did not consider stations along the line and related access cost. Furthermore, by basing the minimum rail fleet size on peak-period passenger capacity requirements, the authors assumed that the route would operate at the maximum allowable headway. This assumption may be unwarranted even for the peak periods since the optimal headway may be heavily influenced by user waiting time. The present work not only optimizes route length but also jointly optimizes the headway, route, and stop spacing.

TABLE 1 Summary of Pertinent Analytical Models for Transit Network Design

Decision Variables	Objective Function	Transit Mode	Street Network Geometry	Passenger Demand	Authors
Route Length	Min. operator and user cost	rail	rectangular grid	General, inelastic, many-to-one	Wirasinghe and Seneviratne (1986)
Zone Length, Headway	Min. operator and user cost	bus	rectangular grid	Piecewise uniform, inelastic,	Tsao and Schonfeld (1984)
Route Spacing, Lengths and Headway	Min. operator and user cost	bus and rail	rectangular grid	Uniform, inelastic, many-to-one	Byrne (1976)
Route Spacing	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-many	Holroyd (1967)
Route Spacing and Headway	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-one	Byrne and Vuchic (1972)
Route Density and Frequency	Min. operator and user cost	bus	rectangular grid	General linear, inelastic, many-to-one	Hurdle (1973)
Route Spacing, Headway and Fare	Max. operator profit, Max. user benefit, etc.	bus	rectangular grid	Uniform Elastic, many-to-one	Kocur and Hendrickson (1982)
Route Spacing, Headway and Stop Spacing	Min. operator and user cost	feeder bus to rail	rectangular grid	General, inelastic, many-to-one	Kuah and Perl (1988)
Route Spacing, Headway and Fare	Max. profit, max. welfare, min. cost	bus	rectangular grid	Irregular, elastic, many-to-many, time dependent	Chang and Schonfeld (1989)
Route Spacing, Zone Length, Headway	Min. operator and user cost	bus	rectangular grid	Uniform, inelastic, many-to-one	Chang and Schonfeld (1993)
Station Location and Spacing	Min. total user travel time	rail	linear	Uniform, inelastic, many-to-one	Vuchic and Newell (1968)
Stop Location and Spacing	Min. operator and user cost	rail	rectangular grid	Uniform, inelastic, many-to-one	Hurdle and Wirasinghe (1980)
Stop Spacing	Min. operator and user cost	bus	radial	General, inelastic, many-to-many	Wirasinghe and Ghoneim (1981)

STUDY APPROACH

The problem is to provide optimal transit service coverage in an urban corridor shown in Figure 1. The corridor of length E and width Y is divided into two zones. Zone 1 consists of the area between the CBD and the route terminus. Zone 2 is area between the route terminus and the end of the corridor.

The basic approach in this research is to develop a total cost function in which the various operator and user cost components are formulated as functions of several decision variables, namely, route length, headway, route spacing, and stop spacing. Optimal stop locations as well as stop spacing are determined. The design objective in determining the optimal service area coverage is to minimize the total operator and user cost. The optimal values of the decision variables are found by taking partial derivatives of the objective function of all decision variables, setting them equal to 0, and solving them simultaneously. This approach, as it will be seen later, resulted in a simple model that offered considerable insight into the optimality conditions and interrelations among variables. The equations obtained are incorporated within an efficient algorithm that determines decision variable values for a more realistic model that includes a service quality constraint.

This analysis of optimal transit service coverage is based largely on Spasovic's master's thesis (*16*), in which more detailed derivations and results can be found.

Simple Model

The following assumptions are made in this model:

1. The corridor is served by a transit system consisting of n parallel routes of uniform length L, separated by a lateral spacing M.

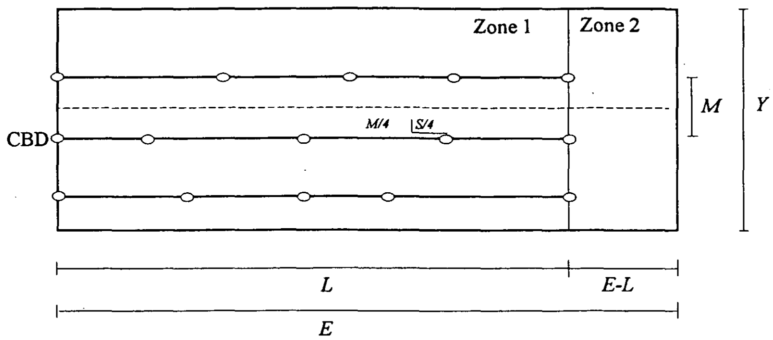


FIGURE 1 Corridor and transit network under study.

2. The routes extend from the CBD outward.
3. The total transit demand is uniformly distributed along the entire corridor, over time, and is insensitive to the quality of transit service.
4. The commuter travel pattern consists of many-to-one or one-to-many trips focused on the CBD.
5. Passengers board and exit transit vehicles only at stops along the route.
6. A very dense rectangular grid street network allows passengers orthogonal access movements (i.e., parallel and perpendicular to the route).
7. Transit vehicles operate in local service (i.e., all vehicles serve all stations).
8. The average access speed is constant. Walking is assumed to be the only access mode.
9. Average wait time is assumed to equal one-half of the headway. The headway is uniform along the route, as well as among all parallel routes.
10. Operator costs are limited to those for vehicles (i.e., infrastructure is freely available).
11. Demand does not exceed vehicle capacity.
12. There is no limit on vehicle fleet size.

The total cost objective function includes the operator cost C_o, and the user cost C_u. The operator cost represents the cost of resources used by the operator to provide the service. The user cost consists of the access, wait, and in-vehicle costs multiplied by their respective values of time.

The operator cost includes the maintenance and overhead as well as more direct costs of operation (driver wage, fuel, brake shoes, etc.). Vehicle depreciation might also be included as a portion of operator cost. In this paper, the operator cost is defined by the hourly operation cost c. The total hourly operator cost is the fleet size multiplied by the hourly operation cost. By definition, the fleet size is the number of on-line vehicles required to provide service and is obtained by dividing the total round-trip time (running time and layover time) by the headway. The average transit operating speed is selected to reflect running and layover times. Therefore, the total round-trip time is the round-trip route length divided by average speed. The stopping delay d is also included in deriving the operator cost. The delay d is a linear function of the number of people waiting for vehicles at stops and the passenger boarding rate:

$$d = \text{const.} + (\text{number of passengers})(\text{boarding time per passenger})$$

A constant delay due to acceleration and deceleration is assumed at each stop. The impact of these delays on the cost of operation is taken into account by multiplying the number of stops (given as $N = L/S$) by the stopping delay d and the operator hourly cost c. The total hourly operator cost is then

$$C_o = \frac{2cYL}{HM}\left(\frac{1}{V} + \frac{d}{S}\right) \tag{1}$$

where

c = vehicle operating cost (\$/veh-hr),
Y = corridor width (km),
L = length of transit route (km/route),
d = stopping delay (hr/stop),
H = route headway (hr/veh),
M = route spacing (km/route),
V = average transit speed (km/hr), and
S = average stop spacing (km/stop).

The hourly user cost, C_u, consists of the access (C_a), wait (C_w), and in-vehicle (C_{iv}) costs:

$$C_u = C_a + C_w + C_{iv} \tag{2}$$

Since the trip origins are uniformly distributed over the corridor served by parallel routes, an average passenger accessing the route perpendicularly walks one-quarter of the spacing between the two routes, for an access distance of $M/4$. The length of passenger access alongside the route depends on whether the trip originated within Zone 1 or Zone 2. A passenger from Zone 1 walks along the route one-quarter of the local stop spacing S before reaching the stop. Passengers originating in Zone 2 have no other choice but to board the route at the terminus, thus having an average access distance of $(E - L)/2 + M/4$. The total hourly user access cost, C_a, is then obtained by multiplying the average access distances by the value of access time (V_a) and the ridership, and dividing by the access speed (G). The total user access cost is then

$$C_a = \frac{V_a}{G}\left[\frac{(E - L)}{2}P(E - L)Y + \frac{S}{4}PLY + \frac{M}{4}PEY\right] \tag{3}$$

where E is corridor length in kilometers and P is passenger trip density in passengers per square kilometer hour.

The total waiting cost, C_w, equals the value of waiting time, V_w, multiplied by the average wait time per passenger ($H/2$) and by the total ridership (PEY).

$$C_w = \frac{V_w H}{2} PEY \tag{4}$$

The total user in-vehicle cost is obtained by multiplying the time that an average passenger spends in the vehicle by the value of in-vehicle time (V_{iv}) and the total number of passengers. In this model, the in-vehicle time consists of two parts: the actual riding time between the origin stop and CBD, and the additional delay due to stops at stations. The average in-vehicle riding time is obtained as the average distance traveled divided by the speed (V). Therefore, the passengers originating in Zone 1 travel about an average distance of $L/2$, and those from Zone 2 travel the whole length of the route L. Thus, the total user in-vehicle cost is given as

$$C_{iv} = V_{iv}\left[\frac{L}{2V} PYL + \frac{L}{V} P(E - L)Y + d\frac{L}{2S} PYL + d\frac{L}{S} PY(E - L)\right] \tag{5}$$

No out-of-pocket costs were included in the user costs. Transit fares are not part of the total cost since they are merely transfer payments from users to operators.

The hourly total system cost, TC, a sum of operator (Equation 1) and user costs (Equations 3–5) is then

$$\begin{aligned} TC(L,H,M,S) = {} & \frac{2cYL}{HM}\left(\frac{1}{V} + \frac{d}{S}\right) + \frac{V_a}{G} PY\left[\frac{(E - L)^2}{2} + \frac{S}{4} L + \frac{M}{4} E\right] + \frac{V_w H}{2} PEY \\ & + V_{iv} PY\left[\frac{L^2}{2V} + \frac{L}{V}(E - L) + d\frac{L^2}{2S} + d\frac{L}{S}(E - L)\right] \end{aligned} \tag{6}$$

The total cost function can be minimized by setting its partial derivatives with respect to the decision variables to 0. In this case, the partial derivatives of the optimization variables, the route length, headway, route spacing, and the stop spacing are

$$\begin{aligned} \frac{\partial TC(L)}{\partial L} = {} & \frac{2cY}{HM}\left(\frac{1}{V} + \frac{d}{S}\right) + \frac{V_a}{G} PY\left[2\frac{(E - L)}{2}(-1) + \frac{S}{4}\right] \\ & + V_{iv} PY\left[\frac{L}{V} + \frac{E}{V} - \frac{2L}{V} + d\frac{L}{S} + d\frac{E}{S} - d\frac{2L}{S}\right] = 0 \end{aligned} \tag{6a}$$

$$\frac{\partial TC(H)}{\partial H} = -\frac{2cYL}{H^2 M}\left(\frac{1}{V} + \frac{d}{S}\right) + \frac{V_w}{2} PEY = 0 \tag{6b}$$

$$\frac{\partial TC(M)}{\partial M} = -\frac{2cYL}{HM^2}\left(\frac{1}{V} + \frac{d}{S}\right) + \frac{V_a}{4G} PEY = 0 \tag{6c}$$

$$\frac{\partial TC(S)}{\partial S} = -\frac{2cYLd}{HMS^2} + \frac{V_a}{4G} PLY + V_{iv} PYL\left[-d\frac{L}{2S^2} - d\frac{1}{S^2}(E - L)\right] = 0 \tag{6d}$$

When Equations 6a–6d are solved independently, we obtain the following equations:

$$L^* = E - \frac{8cG\left(\frac{1}{V} + \frac{d}{S}\right) + PHMV_a S}{4\left[V_a - V_{iv} G\left(\frac{1}{V} + \frac{d}{S}\right)\right] PHM} \tag{7a}$$

$$H^* = \left[\frac{4cL\left(\frac{1}{V} + \frac{d}{S}\right)}{V_w MPE}\right]^{1/2} \tag{7b}$$

$$M^* = \left[\frac{8cLG\left(\frac{1}{V} + \frac{d}{S}\right)}{V_a HPE}\right]^{1/2} \tag{7c}$$

$$S^* = \left[\frac{4Gd(2c - V_{iv} PHM)\left(\frac{L}{2} - E\right)}{V_a HPM}\right]^{1/2} \tag{7d}$$

Several observations should be made here: when the route length, route spacing, headway, and stop spacing are optimized independently of each other, their relation to the other decision variables can be read directly from Equations 7a–7d. These equations provide the optimal value of one of the decision variables as a function of the other three variables and provide useful insights into the relations between the decision variables and parameters. For example, Equation 7a can be used to find the optimal route length when the headway, route spacing, and average stop spacing are given (e.g., to satisfy the minimum service standards). Such equations may be useful by themselves in some situations in which certain decisions variables such as the route length L or the stop spacing S cannot be modified. Unfortunately, Equations 7a–7d cannot be solved simultaneously using algebraic methods.

According to Equation 7a, the optimal route length varies directly with the corridor length E, passenger density P, operating headway H, route spacing M, value of access time V_a, transit speed V, and stop spacing S. It varies inversely with the vehicle operating cost c, and access speed G.

The optimal headway varies directly with the square root of operator cost, route length, and stopping delay. It varies inversely with the square root of the wait cost, passenger density, transit speed, corridor length, route spacing, and stop spacing.

The optimal route spacing varies directly with the square root of access speed, operator cost, route length, and stopping delay. It varies inversely with the square root of the access cost, passenger density, transit speed, corridor length, headway, and stop spacing.

Finally, the optimal average stop spacing varies directly with the square root of access speed, operator cost, and time lost per stop. It varies inversely with the square root of the access cost, passenger density, headway, route spacing, and in-vehicle cost.

More Realistic Model

Although the simple model provided valuable insights into the relationship among the decision variables and exogenous parameters, it is still too complex for the simultaneous optimization of all of the decision variables algebraically. To solve the model, a numerical algorithm was developed. In addition, some of the original assumptions were relaxed by the introduction of a vehicle capacity constraint. This constraint ensures that the total capacity provided on the routes satisfies the demand by restricting the maximum allowable headway; it is written as

$$PEY \le K \frac{Y}{MH} l \tag{8}$$

where K equals capacity of transit vehicle (in spaces), and l is the allowable peak load factor at the CBD.

Finally, the model can be written in the following form:

$$\begin{aligned} TC(L,H,M,S) = {} & \frac{2cYL}{HM}\left(\frac{1}{V} + \frac{d}{S}\right) \\ & + \frac{V_a}{G} PY\left[\frac{(E-L)^2}{2} + \frac{S}{4}L + \frac{M}{4}E\right] \\ & + \frac{V_w H}{2} PEY \\ & + V_{iv}PY\left[\frac{L^2}{2V} + \frac{L}{V}(E-L)\right. \\ & \left. + d\frac{L^2}{2S} + d\frac{L}{S}(E-L)\right] \end{aligned}$$

subject to

$$PEMH \le Kl$$

$$L,H,M,S \ge 0$$

OPTIMIZATION ALGORITHM

The preceding model is formulated as a constrained optimization problem with nonlinear objective function and linear constraints. The model can be solved by using a penalty method (*20*) as an unconstrained optimization problem by pricing the constraint out of the constraint set and introducing it into the objective function with a penalty.

Instead of using the penalty function method, an algorithm was developed that sequentially applied Equations 7a–7d in somewhat modified form to advance from an initial feasible solution toward the optimal solution. The algorithm, shown in Figure 2, starts with a trivial feasible solution to the problem and in each step improves the value of the objective function by computing an optimal value of one decision variable while keeping the others at their feasible levels. In computing the optimal values of decision variables, the algorithm first computes the number of stops, then route length, route spacing, and finally headway. In each step, the value of a newly computed variable is recorded and used in the next step for computing the optimal values of other decision variables. The algorithm keeps improving the objective function until it converges to an optimal solution. The algorithm decides to terminate on the basis of two criteria. The first criterion examines whether the newly obtained optimal headway satisfies the capacity constraint—that is, it checks whether the optimal headway is smaller than maximum allowable headway. If the optimal headway is greater than the maximum allowable headway, the algorithm terminates. The optimal headway is set equal to the maximum allowable headway, and the last set of decision variables is considered an optimal solution. The second criterion determines that the values of total costs from two successive iterations are sufficiently close that no significant further improvement can be expected. Assuming that the optimal set of decision variables is reached, the program computes the values of the total cost function for the optimal route length, number of stops, and spacing (i.e., N^*,L^*,M^*) allowing variations in the headway, H. The purpose of this is to investigate the shape of the total cost near the optimum. As discussed later, the total cost turned out to be a relatively flat (shallow, four-dimensional, U-shaped) function. Thus, small deviations from the optimal decision variables result in even smaller relative changes in total cost.

Besides computing the optimal values of the decision variables very quickly, the algorithm allows us to incorporate a scanning procedure for deriving the actual location of stops along the route. Note that Equation 7d calculates an optimal value of the average stop spacing, thus implying uniform spacing along the route. Although a uniform spacing is quite common on bus transit routes in urban areas with grid street networks (e.g., Philadelphia and Manhattan), it does not yield the optimal solution for our objective function that minimizes the total user and operator costs. Intuitively, one might see that the actual stop location, thus spacing, will vary along the line as a result of the trade-off between the delay cost at stops incurred by the operator and passengers aboard the vehicles and the access cost of passengers boarding the vehicle along the route. Therefore, a scanning approach is incorporated in the algorithm to optimize variable stop spacing along transit routes. This scanning algorithm is somewhat similar to a method presented by Newell (*4*) and Hurdle (*6*). They integrated the demand function over time and dispatched the vehicle each time the optimum condition was reached. Wirasinghe used a somewhat analogous integration procedure over space to locate stops on feeder bus routes when the function of cumulative number of stations reached an integer (*10*). Chang and Schonfeld used a similar approach to optimize the lengths of bus service zones (*18*).

The partial total cost equation, derivation of formulas for optimal decision variables that are used in the algorithm, and description of the scanning procedure for determining the actual location of the stops are described in the following.

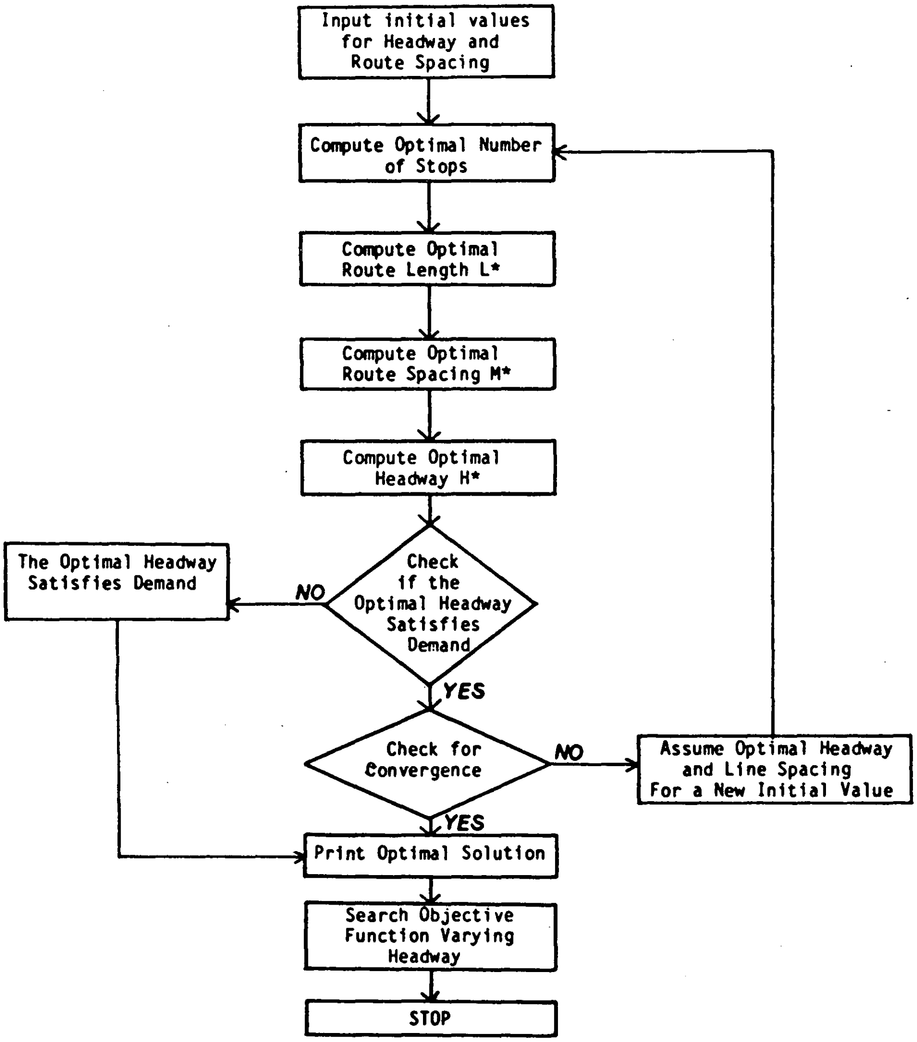


FIGURE 2 Optimization algorithm.

Optimal Number of Stops and Stop Spacing

The local station spacing on a route depends on trade-offs between delays to vehicles and passengers already on board versus the passengers' access cost. Clearly, under the objective of minimizing the total cost, the vehicle traveling on the route will not stop if the combined cost of delaying the passengers aboard the vehicle and operator cost outweighs the access cost of passengers waiting along the route.

To apply such logic in determining actual stop locations, the corridor was partitioned into a finite number of small areas, ΔX, and scanned from its end toward the CBD. At any point along the route at X distance from CBD, and for the small increment ΔX (e.g., 0.1 km), the number of people within the increment as well as the cumulative number of people (from the end of the corridor to X) aboard the vehicle entering the increment ΔX was computed. At any point along the route at X distance from CBD, the total cost function that affects the stop location consists of the three parts: the operator cost of vehicle currently at X stopping in the next increment ΔX, the cost of users along the route accessing the stop within increment ΔX, and the delay cost for the cumulative transit demand (i.e., the passengers aboard the vehicle) that originated at an area beyond the potential stop in the increment ΔX and the outer end of the corridor. The parts of the function that do not affect stop spacing are left out of the total cost equation because they drop to 0 in the derivatives of the cost function. The partial cost function at any point X along the corridor is

$$TC(Si) = \frac{2cdY}{HM}\int_{X-\Delta X}^{X}\frac{1}{Si}dx + \frac{V_aY}{4G}P\left[Si\int_{X-\Delta X}^{X}dx + V_{iv}d\Delta XY\frac{1}{Si}\int_{X}^{E}dx\right] \quad (9)$$

The partial derivative of Equation 9 with respect to stop spacing is

$$\frac{\partial TC(Si)}{\partial Si} = -\frac{2cdY}{Si^2HM}\Delta X + \frac{V_aY}{4G}P\Delta X + \frac{V_{iv}d\Delta XY}{Si^2}P(E - X) = 0 \quad (10)$$

The optimal stop spacing is

$$Si^* = \left\{\frac{4Gd[2c + V_{iv}HP(E - X)M]}{V_a PHM}\right\}^{1/2} \tag{11}$$

Equation 11 is used to compute the optimal fractional number of stops within each increment i, that is, Ni^*

$$Ni^* = \frac{\Delta X}{Si^*} \tag{12}$$

The total number of stops in the corridor is then obtained by summing incremental stops over all the increments i, namely,

$$N^* = \sum_{i=1}^{E/\Delta x} N^*i \tag{13}$$

The optimal number of stops is used in the next step to derive the optimal route length. After the optimal values of all decision variables are computed, the actual location of a stop is determined using Equation 13 by summing the stop increments on the route in the direction to the CBD. Each time an integer number is reached in the cumulative function of the number of stops, a true stop is established.

Optimal Route Length

The optimal route length is obtained as a result of the tradeoff between operator and user access costs. Intuitively, the route should end at the point at which the supplier marginal costs equals the marginal access cost for users accessing the route from an area beyond the terminus. Access along the route to the stop is omitted from consideration for the optimal number of stops has been determined in the previous step. Thus, the partial cost function is

$$TC(N^*,L) = \frac{2cYL}{VHM} + \frac{V_a}{G}\frac{(E - L)^2}{2}PY \tag{14}$$

Taking the partial derivative of Equation 14 with respect to the route length and setting it equal to 0 yields the following expression for the optimal route length:

$$L^* = E - \frac{2cG}{V_a PHVM} \tag{15}$$

Equations 13 and 14 are used as input for computing the optimal route spacing M^*.

Optimal Route Spacing

The optimal route spacing depends on the magnitude of user access cost via paths perpendicular to the route as well as on the operator cost per route. In Equation 7c the L/S is replaced by N^*, yielding the modified equation that is used within the algorithm:

$$M^* = \left[\frac{8G(L^*c + dN^*)}{V_a PEH^2}\right]^{1/2} \tag{16}$$

Optimal Headway

The optimal values of N^*,L^*,M^* are then input into the modified Equation 7d, yielding the optimal operating headway on the route:

$$H^* = \left[\frac{4c(L^* + VdN^*)}{V_w PEM^*V}\right]^{1/2} \tag{17}$$

NUMERICAL EXAMPLE—RECTANGULAR CORRIDOR WITH UNIFORM PASSENGER DENSITY

The input data for this example are shown in Figure 3. The parameter values were taken from Keeler and Small (*21*) for values of time and from Fisher and Viton (*22*) for costs. The algorithm generated optimal route length, route spacing, operating headway, number of stops, supplier and user cost, and the total cost of the transit system, which are given in Table 2. The table shows several iterations of the algorithm, which shows that it converges quickly toward the optimum solution. Note that the optimal headway of 8.5 min that minimizes the total cost is much smaller than the maximum allowable headway of 12 min that was used as an initial feasible solution. The total cost function is relatively flat near the optimum. This indicates that minor deviations away from the optimum will not increase the cost significantly. It is notable that at the optimum, the costs of user access time, of operating the service, and of waiting time are equal. This optimality condition is similar to the findings reported by Holroyd (*1*), Kocur (*12*), and Tsao and Schonfeld (*13,14*) for their particular models. The optimal route length, route spacing, and stop location are shown in Figure 3. This figure shows that the stop spacing increases along the route in the direction of passenger accumulation toward the CBD. At the outer end of the transit route, the delay cost of operator and passengers already on board is smaller than the access cost of passengers along the route who are trying to board. As access costs outweigh delay costs, more frequent stops are established. As the route approaches the CBD, the number of passengers aboard the vehicles increases so that the delay cost begins to outweigh the access cost to passengers along the route who are trying to board, thus increasing the optimal stop spacings.

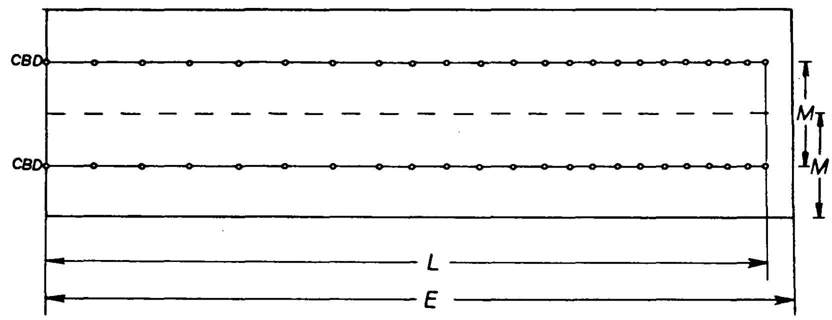
SPECIAL CASES

The optimization approach was also applied to a rectangular corridor with the passenger density decreasing linearly from the CBD and to a wedge-shaped corridor with uniform passenger density. The input data for these cases are the same as for the previous example.

Rectangular Corridor with Linearly Decreasing Passenger Density

To yield the same passenger volume as in the previous example so that the results can be compared, the passenger density of 77.2 passengers per kilometer per hour was assumed

transit corridor length:	8.045 km (5 miles)
transit corridor width:	4.827 km (3 miles)
average transit speed:	16.09 km/hour (10 miles/hour)
average access speed:	4.39 km/hour (2.73 miles/hour)
boarding (or alighting) time:	2 seconds
operator cost:	$30/vehicle-hour
value of access time:	$9/passenger-hour
value of in-vehicle time:	$3/passenger-hour)
value of waiting time:	$9/passenger-hour)
passenger density:	38.6 passengers/km^2-hour (100 passengers/mile^2-hour)
transit vehicle capacity:	50 seats/vehicle
allowable peak load factor:	1.0



o = Stop/Station

Optimal Values:

1 MILE

Route Length L: 7.78 km/route (4.83 miles/route)
Headway H: 8.5 minutes
Spacing M: 1.24 km/route (0.77 miles/route)

FIGURE 3 Optimal transit route configuration for rectangular area with uniform passenger density.

in the linear density function—$P = 77.2(1 - x/E)$. Figure 4 (*top*) shows the optimal transit route configuration. From the figure, it can be seen that the stop spacing is decreasing along the route toward the CBD. This is consistent with the passenger distribution along the route. Because passenger density decreases from the CBD, the passenger transit demand in the outer area is much smaller than it is near the CBD. At a certain distance from the CBD, the stop spacing starts increasing. As the route approaches the CBD, the number of passengers aboard vehicles rises rapidly so that these passengers' delay costs increase faster than the access costs of passengers along the route. As in the previous example the algorithm converges quickly to the optimal solution. A detailed discussion of this case study may be found in work by Spasovic (*16*).

Wedge-Shaped Corridor with Uniform Passenger Density

Figure 4 (*bottom*) shows the optimal transit configuration for the wedge-shaped corridor. In it, the stop spacing increases along the route in the direction of passenger accumulation toward the CBD. As the route approaches the CBD, the number of passengers aboard the vehicle increases so that their delay costs outweigh the access costs of passengers along the route waiting to board the vehicle. In addition, because of the wedge shape of the service area, the number of passengers is decreasing in the direction of the CBD. As in the previous examples, the algorithm converges quickly toward the optimal solution. At the optimum, the total supplier cost (including delay cost) and the user access cost are equal. The user wait costs are about 25 percent lower than either operator or user access cost. However, it should be pointed out that the operating cost (without stop delay) and the user wait cost are equal at the optimum. A detailed discussion of this case study also may be found in work by Spasovic (*16*).

Sensitivity Analysis

The sensitivity analysis is performed to show how changes in the more important exogenous parameters affect the values of the decision variables. The results are presented in the form of elasticities, which are convenient dimensionless measures of sensitivity. Two approaches for performing the sensitivity analysis were used. First, the sensitivity of one transit design element with respect to a particular parameter was examined without reoptimizing the system. This approach provides a very good insight into the relative changes in the dependent element of the transit system if a change in a particular parameter occurs. Second, the sensitivity of the groups of transit design variables with respect to the single parameter has been measured after reoptimizing the system. The elasticities of the design variables—namely, route length, spacing, headway and number of stops with respect to the corridor length, passenger density, transit and access speed, operator cost, and values of riding access and waiting times without and with reoptimization—are presented in Tables 3 and 4. The tables show that, for example, if the passenger density is increased by 10 percent the headway will be reduced by 4.8 percent. This result confirms that headway varies (ap-

TABLE 2 Optimal Cost and Design Variables

Solution No.	I	II	III	IV	V	VI	VII Optimal	VIII
Route Length (km/route)	7.9	7.69	7.8	7.76	7.78	7.77	7.78	7.77
Route Spacing (km/route)	1.609	1.054	1.303	1.197	1.252	1.230	1.241	1.236
Route Density (routes/km)	1.609	0.949	0.767	0.835	0.8	0.8138	0.806	0.768
Headway (minute/vehicle)	12.00	7.52	9.1	8.3	8.6	8.4	8.5	8.5
No. of Stops	24.96	21.20	23.15	22.32	22.71	22.54	22.62	22.59
Operator Cost ($/hr)	441.81	1046.22	709.47	846.3	781.44	810.3	797.13	803.07
Operator Delay Cost ($/hr)	89.25	194.01	135.03	158.58	147.24	152.19	149.88	150.90
Total Operator Cost ($/hr)	531.06	1240.23	844.5	1004.88	928.68	962.49	947.01	953.97
User Access Cost ($/hr)	1240.35	834.51	1013.25	933.3	974.73	959.19	967.44	964.29
User Wait Cost ($/hr)	1350.0	846.6	1024.8	930.3	966.9	947.7	954.9	951.0
Total System Cost ($/hr)	4451.19	4182.27	4172.49	4166.34	4162.92	4163.07	4162.71	4162.77
Fleet Size (vehicles)	17.702	41.341	28.15	33.496	30.956	32.083	31.567	31.799

proximately) with the square root of the passenger density. In addition, in both cases the optimal route length L is elastic (i.e., the absolute value of the elasticity exceeds 1.0) with respect to the corridor length E. The reason for this is that as the length of the corridor E is increased, the length of the area between the terminus and the end of the corridor, $E - L$, is increased very slowly, thus increasing L faster than E. The results of sensitivity analysis for the other two cases may be found in work by Spasovic (*16*).

CONCLUSIONS

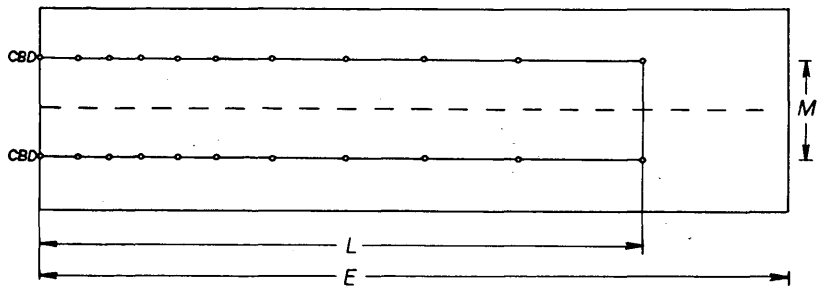
The model developed in this paper provides simple guidelines for optimizing the extent of transit routes and other major operating characteristics. Equations 7a–7d can be used to optimize separately route length, route spacing, headway, and stop spacing. The square root in Equations 7b–7d indicates that optimal solutions are relatively insensitive to changes in system parameters.

The algorithm provides an efficient and accurate method for simultaneously optimizing the decision variables. The results closely confirm that in a system optimized for minimum total cost, the vehicle operating cost, user wait cost, and user access cost should be equal. This finding is similar to those of previous studies (*1,12–14*) for somewhat different transit systems and provides a useful optimality guideline for designing real transit systems.

The total cost function is relatively flat near the optimum. For practical applications, this implies that a near-optimal cost can be attained while fitting the transit network to the particular street network or modifying its operating schedule.

FUTURE RESEARCH

Several simplifying assumptions should be relaxed in future models. More realistic and irregular distributions of demand over space and time should be used. The model should be improved to handle non-CBD trips and access modes other than walking. The cost of transit facilities (e.g., station cost) should be considered in order to make the methodology more applicable in planning fixed guideway modes. Demand elasticity should be explicitly considered in formulating demand as a function of level of service and fare. This will also allow optimization for objectives such as profit, revenues, and welfare.

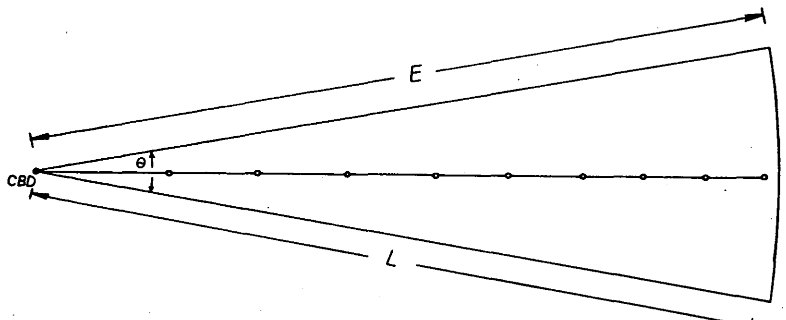


Optimal Values:

o = Stop/Station

1 MILE

Route Length L: 6.47 km/route (4.02 miles/route)
Headway H: 7.9 minutes
Spacing M: 1.14 km/route (0.71 miles/route)



o = Stop/Station

Optimal Values:

1 MILE

Route Length L: 7.9 km/route (4.91 miles/route)
Headway H: 7.5 minutes
Spacing Θ: 19°19' degrees/route

FIGURE 4 Optimal transit route configuration: ***top*****, rectangular area with linearly decreasing passenger density;** ***bottom*****, wedge area with uniform passenger density.**

TABLE 3 Elasticities of Design Variables with Respect to Various Parameters for Rectangular Area with Uniform Passenger Density, Without Reoptimization

	Design Variables			
	Route Length	Headway	Route Spacing	Number of Stops
Corridor Length	1.0286	- 0.0013	- 0.0010	0.6912
Passenger Density	0.0185	- 0.4807	- 0.4806	0.0625
Transit Speed	0.0185	- 0.4103	- 0.4103	0.0000
Access Speed	- 0.0328	0.0000	0.5052	- 0.5050
Operator Cost	- 0.0328	0.5051	0.5052	- 0.2447
Value of In-Vehicle Time	0.000	0.000	0.0000	- 0.2967
Value of Access Time	0.0185	0.000	- 0.5049	0.5050
Value of Wait Time	0.000	- 0.5050	0.000	0.0000

TABLE 4 Elasticities of Design Variables with Respect to Various Parameters for Rectangular Area with Uniform Passenger Density, with Reoptimization

	Design Variables			
	Route Length	Headway	Route Spacing	Number of Stops
Corridor Length	1.0278	- 0.0076	0.0068	0.7317
Passenger Density	0.0116	- 0.3247	- 0.3258	0.0083
Transit Speed	0.0146	- 0.2742	- 0.2751	- 0.1207
Access Speed	- 0.0282	- 0.3487	- 0.6596	- 0.3892
Operator Cost	- 0.0125	0.3382	0.3379	- 0.1296
Value of In-Vehicle Time	- 0.00005	0.0002	- 0.00003	- 0.2596
Value of Access Time	0.0209	0.3476	- 0.6618	0.3806
Value of Wait Time	- 0.0135	- 0.6819	0.3257	- 0.0289

APPENDIX A Notation

The following symbols are used in this paper:

c = vehicle operating cost ($/veh-hr)
C_a = total access time cost ($/hr)
C_{iv} = total in-vehicle travel time cost ($/hr)
C_o = total operator cost ($/hr)
C_u = total user time cost ($/hr)
C_w = total waiting time cost ($/hr)
d = average time lost per stop (hr/stop)
E = corridor length (km)
G = average passenger access speed (km/hr)
H = operating headway for a transit route (hr/vehicle)
K = vehicle capacity (seat/vehicle)
L = length of transit route (km)
l = allowable peak hour load factor at CBD
M = lateral route spacing for rectangular area (km/route)
n = number of routes
N = number of stops on route
θ = lateral route spacing for wedge area (degree/route)
P = passenger trip density (passenger/km^2-hr)
S = average stop spacing (km/stop)
TC = total cost of a transit system ($/hr)
V = average transit operating speed (km/hr)
V_a = value of access time ($/passenger-hr)
V_{iv} = value of passenger in-vehicle time ($/passenger-hr)
V_w = value of passenger waiting time ($/passenger-hr)
Y = corridor width (km)

REFERENCES

1. E. M. Holroyd. The Optimal Bus Service: A Theoretical Model for a Large Uniform Urban Area. In *Vehicular Traffic Science* (L. C. Edie, ed.), American Elsevier, New York, 1967, pp. 309–328.
2. V. R. Vuchic and G. F. Newell. Rapid Transit Interstation Spacing for Minimum Travel Time. *Transportation Science*, Vol. 2, No. 4, 1968, pp. 359–374.
3. V. R. Vuchic. Rapid Transit Interstation Spacing for Maximum Number of Passengers. *Transportation Science*, Vol. 3, No. 4, 1969, pp. 214–232.
4. G. F. Newell. Dispatching Policies for a Transportation Route. *Transportation Science*, Vol. 5, No. 1, 1971, pp. 91–105.
5. B. F. Byrne and V. R. Vuchic. Public Transportation Line Position and Headway for Minimum Cost. In *Traffic Flow and Transportation* (G. F. Newell, ed.), American Elsevier, New York, 1972, pp. 347–360.
6. V. F. Hurdle. Minimum Cost Locations for Parallel Public Transit Lines. *Transportation Science*, Vol. 7, No. 4, 1973, pp. 340–350.
7. B. F. Byrne. Public Transportation Line Positions and Headways for Minimum User and System Cost in a Radial Case. *Transportation Research*, Vol. 9, 1975, pp. 97–102.
8. B. F. Byrne. Cost Minimizing Positions, Lengths and Headways for Parallel Public Transit Lines Having Different Speeds. *Transportation Research*, Vol. 10, 1976, pp. 209–214.
9. V. F. Hurdle and S. C. Wirasinghe. Location of Rail Stations for Many-to-One Travel Demand and Several Feeder Modes. *Journal of Advanced Transportation*, Vol. 14, No. 1, 1980, pp. 24–45.
10. S. C. Wirasinghe. Nearly Optimal Parameters for a Rail/Feeder-Bus System on a Rectangular Grid. *Transportation Research A*, Vol. 14, No. 1, 1980, pp. 33–40.
11. S. C. Wirasinghe and N. S. Ghoneim. Spacing of Bus Stops for Many-to-Many Travel Demand. *Transportation Science*, Vol. 15, No. 3, 1981, pp. 210–246.
12. G. Kocur and C. Hendrickson. Design of Local Bus Service with Demand Equilibration. *Transportation Science*, Vol. 16, No. 2, 1982, pp. 149–170.
13. S. M. Tsao and P. M. Schonfeld. Optimization of Zonal Transit Service. *Journal of Transportation Engineering*, ASCE, Vol. 109, No. 2, 1983, pp. 257–272.
14. S. M. Tsao and P. M. Schonfeld. Branched Transit Service: An Analysis. *Journal of Transportation Engineering*, ASCE, Vol. 109, No. 2, 1984, pp. 257–272.
15. S. C. Wirasinghe and P. N. Seneviratne. Rail Line Length in an Urban Transportation Corridor. *Transportation Science*, Vol. 29, No. 4, 1986, pp. 237–245.
16. L. N. Spasovic. *A Method for Optimizing Transit Service Coverage*. Master's thesis. University of Maryland, College Park, 1986.
17. G. K. Kuah and J. Perl. Optimization of Feeder Bus Routes and Bus Stop Spacings. *Journal of Transportation Engineering*, ASCE, Vol. 114, No. 3, 1988, pp. 341–354.
18. S. K. Chang and P. M. Schonfeld. *Analytical Optimization of Parallel Bus Routes with Time Dependent and Elastic Demand*. Working Paper 89-22. Transportation Studies Center, University of Maryland, College Park, 1989.
19. S. K. Chang and P. M. Schonfeld. Optimal Dimensions of Bus Service Zones. *Journal of Transportation Engineering*, ASCE (in press).

20. M. S. Bazaara and C. M. Shetty. *Nonlinear Programming*. John Wiley & Sons, New York, 1977.
21. T. E. Keeler and K. A. Small. *The Full Cost of Urban Transport—Part III: Automobile Costs and Final Intermodal Cost Comparisons*. Institute of Urban and Regional Development, University of California, Berkeley, 1975.
22. P. Fisher and P. Viton. *The Full Cost of Urban Transport—Part I: Economic Efficiency in Bus Operations; Preliminary Intermodal Cost Comparisons and Policy Implications*. Institute of Urban and Regional Development, University of California, Berkeley, 1974.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.