Zonal Route Design for Transit Corridors

PETER G. FURTH

Northeastern University, Boston, Massachusetts

In "zonal express service," a transit corridor is divided into zones. Each inbound zonal express route picks up passengers in its zone only, then runs express to the CBD; outbound routes do the opposite. In "zonal local service," on the other hand, an inbound vehicle will stop between its service zone and the CBD to allow passengers to alight, but not to board. Outbound vehicles do the opposite; they will pick up passengers anywhere along the route, but will allow them to alight in the route's service zone only. Zonal express service design, i.e., the choice of zone boundaries and of service frequencies, has been studied by Turnquist for linear corridors using dynamic programming. These results are extended to zonal design for bidirectional local service, including light direction deadheading, and to branching as well as linear corridors. Application to a Boston area corridor shows considerable potential for reducing operator cost.

It is well recognized that transit service exhibits economies with respect to both the size and concentration of its market. The greater the size of a transit route's market, the higher the vehicle loads can be and the more frequent service can be, yielding improvements in both operator cost and passenger level of service. And when origins and destinations are more concentrated spatially, service can be more direct and speedy, lowering travel time for passengers and vehicle-hours for the transit operator. Thus it is to be expected that routes serving the radial corridors of large cities are among the most efficient, since these corridors typically have large transit markets concentrated around the radius of the corridor and, to a large extent, oriented toward the downtown destination. Yet transit operators have often found it advantageous to enhance the natural concentration that is found in urban corridors by segmenting the market. A common segmentation by service type is to separate some of the downtown-oriented traffic from the local traffic by offering it express service. A common spatial segmentation is achieved by forcing passengers traveling within the central city to use a short route while passengers beginning or ending their trips in the suburbs must use a different, longer route. With each market segmentation, the concentration of demand within each market segment increases, while the average market segment size decreases. Thus there is a trade-off between the economies of scale and the economies of concentration, and the question arises as to what is the optimal market segmentation, or equivalently, what is the optimal set of transit services for a given corridor, since each service creates its own market segment. The potential for improving the efficiency of transit service in a corridor through such an analysis can be substantial, depending on the demand structure, the road network, and the degree to which economies of concentration are already being exploited.

To avoid confusion, the term "corridor" in this paper should be understood to refer to a linear, radial network of one or more streets (in the case of branching corridors, a branching network) along which a transit route can operate, and the transit market along that street or streets.

1. ZONAL EXPRESS ROUTE DESIGN

A COMMON example of spatial market segmentation within a corridor is found in downtown-oriented express service. The part of a corridor to which express service is offered is called the express service area. Serving the entire express service area with a single route would require that every vehicle travel the full length of the service area, making many stops. Applying the strategy of zonal design, the service area is broken up into several nonoverlapping zones, and one express route is designed to serve each zone. Each such zonal express route collects passengers inbound and distributes passengers outbound within its service zone, and travels nonstop between the inner boundary of its service zone and downtown using the fastest

0041-1655/86/2001-0001 $01.25

path available. The main advantage of zonal service is that it reduces the number of vehicle-miles of travel in the outer zones of the corridor. It also reduces travel time since the number of stops made by each vehicle is only a fraction of all the stops in the service area. This strategy has been applied to both rail and bus systems, and has been modeled for rail by SALZBORN[1] and for bus by CLARENS AND HURDLE[2] and by TURNQUIST,[3] both of whom developed optimal design methods for one-directional express zonal service. This paper extends Turnquist's approach to deal with bidirectional local service, and with branching as well as linear corridors.

With some modifications, Turnquist's approach is summarized below. A local street paralleled by an expressway forms the spine of an express service area of corridor emanating from the central business district (CBD). Along the spine are points which could serve as service zone outer boundaries. (Bus stops just inbound from expressway access points are especially suitable for this purpose.) Each of these points is therefore a potential route terminus. These points are numbered from 1 to n with increasing distance from the CBD, where point n is the outer boundary of the service area, as shown in Figure 1. Sector i is defined as the segment of the corridor immediately inbound of point i; it includes point i but does not include point $i-1$. Route (i, j) is defined as the express route whose service zone is sectors, $i, \cdots, j$; buses on route (i, j) travel nonstop between downtown and the inner boundary of sector i, serve as collector/distributor within sectors i through j, and have their outer terminus at point j. Then we define:

b_{ij} = number of buses assigned to route (i, j)

p_{ij} = minimum number of buses that may be assigned to route (i, j) if route (i, j) is active

where p_{ij} is the minimum number of buses that will ensure tolerable headways and loading levels on route (i, j). The number of buses assigned to a route is assumed to be integer. Then the zonal system that requires the smallest number of buses to serve the corridor is found by solving for f_n, the minimum number of buses needed to serve all n sectors, using the equations

$$f_0 = 0 \tag{1}$$

$$f_j = \min_{i=1,\cdots,j}(p_{ij} + f_{i-1}) \quad \text{for} \quad j = 1, \cdots, n \tag{2}$$

where f_j is the minimum number of buses needed to serve sectors $1, \cdots, j$. The stage variable is j, the outermost sector served. (Turnquist's formulation uses number of zonal routes as stage variable and j as state variable, while the above formulation ignores number of zonal routes and therefore uses j as stage variable and has no state variable.) The decision variable i in Equation 2 represents the choice of the inner boundary of the service zone whose outer boundary is point j.

If the objective is to optimize a generalized cost of the route system including such factors as travel time or passenger demand, then we solve jointly for the optimal zoning system and bus allocation among routes by defining f_j to be the minimum generalized cost of serving sectors $1, \cdots, j$, and replacing Equation 2 with

$$f_j = \max_{\substack{i=1,\cdots,j,\\ b_{ij}=p_{ij},p_{ij}+1,\cdots}} [c_{ij}(b_{ij}) + f_{i-1}] \quad \text{for} \quad j = 1, \cdots, n \tag{3}$$

where $c_{ij}(b_{ij})$ is the cost incurred in serving the demands of sectors $i, \cdots, j$ with route (i, j) when b_{ij} buses are assigned to that route. If there is a fleet size constraint, Turnquist has shown how it can be accounted for by the addition of a state variable B_j, the number of buses available for use on routes whose outer terminus is points $1, \cdots, j$. If the number of buses being used in the corridor is large, or if there are several vehicle types to choose from, a close approximation can be obtained with less computational effort by dualizing the fleet size constraint for each vehicle type and searching for the optimal shadow price, as suggested by HAGBERG AND HASSELSTROM.[4]

The problem of finding the optimal route zoning strategy can be looked upon as a production and inventory scheduling problem (see, for example, WAGNER[5]), as each sector has a certain demand for which capacity must be provided, analogous to each period in a planning horizon having demands that must be met. For simplicity in this discussion, consider express service in the inbound direction. Just as the demand of period i in the production scheduling problem can be met by production in period i or by excess production in a prior period, so the demand of sector i in the route zoning problem can be met by capacity provided on a route originating in sector i or on a route originating in a sector upstream of i. When production and inventory holding costs are concave, it is a well known

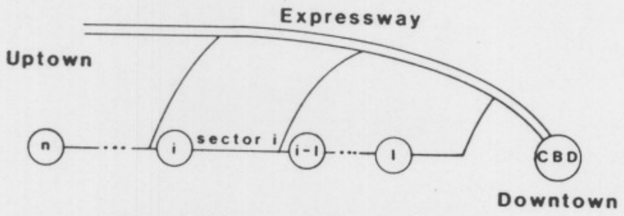

Fig. 1. Linear corridor configuration.

result that an optimal schedule for the production scheduling problem can be found for which production in any period will either be zero or exactly enough to meet the demands of an integral number of consecutive periods beginning with that period. In the route zoning problem, the analogous property exists: the capacity of a route beginning in sector j should either be zero (i.e., no route begins at sector j) or enough to meet the demands of sectors $i, \cdots, j$ for some $i \le j$. This property arises in the route design problem not because of the cost structure, but because of the exogenous constraint that service zones be nonoverlapping. In the production scheduling problem, as in the route zoning problem, this property gives rise to Equations 1 and 2, where stage variable j may be interpreted as the index of the period for which there remain j periods in the planning horizon, p_{ij} as the sum of production and holding costs incurred in meeting the demands of periods i through j through production in period j only, and f_j as the minimum cost of meeting the demands of periods, $1, \cdots, j$.

The modeling approach taken in this paper, which follows Turnquist[3] and Salzborn,[1] by treating the route as a set of discrete nodes and segments, may be contrasted with that taken by Clarens and Hurdle[2] who model space as a continuum. Using the discrete approach, demand rates at each stop and travel times on each link can be represented exactly. In the continuous space approach, which is also used in such works as KOCUR AND HENDRICKSON,[6] WIRASINGHE et al.,[7] and HURDLE,[8] demand is represented by a demand intensity function that must be assumed to vary slowly over space, and travel time must be represented by a slowly varying function of distance. Thus, the discrete approach can far more accurately represent discontinuities in demand, such as a major transfer stop in a neighborhood of otherwise low demand intensity, and discontinuities in travel time due to such factors as wide spacing between freeway access ramps and the varying alignment of the freeway vis-à-vis the corridor. It is also worth noting that nearly all modelers treat time as a continuum by assuming slowly varying demand rates; however, this assumption rarely presents a serious conflict with reality.

Continuous models also require a continuous solution space. The solution generated by Clarens and Hurdle is an optimal zone size for every point in the corridor. They suggest drawing on a map circles representing the optimal zone size for a grid of points along the corridor, and then asking a route planner to choose zone boundaries that will roughly match zone sizes to those indicated on the map.

The advantage of the continuous approach is in finding closed-form expressions for optimal control parameters, such as zone size. Such expressions yield insight into the relationships between these parameters and the data. From Clarens and Hurdle it may be seen that the optimal zone size, expressed in terms of market size (total demand) depends on available resources. When resources are most scarce and the objective is to minimize operator costs, the optimal market size is that which will just fill the buses at the greatest acceptable headway. With more resources available to improve level of service, buses remain full, but market size increases in proportion to the square root of the demand intensity, leading to smaller headways in the higher demand areas. With still more resources or in regions of unusually high demand, zone size increases less quickly (in proportion to the cube root of demand intensity) but headways decrease disproportionately, so that buses become no longer full.

In contrast to the continuous space approach, the discrete approach offers no immediate insights or rules of thumb regarding optimal control parameters. Yet it provides a much more exact representation and yields solutions that are immediately implementable. Thus, continuous space models can serve a useful function, albeit a function that is limited primarily to theoreticians, whereas discrete space representation is indispensable for practical planning purposes.

2. ZONAL DESIGN FOR LOCAL SERVICE: SYMMETRIC SERVICE ZONES

BECAUSE OF the need to accommodate interzonal travel, zonal service for the local transit market must differ from zonal express service. As with express service, the corridor is divided into nonoverlapping service zones, and one route is designed to serve each zone. Inbound, a local zonal route begins at the outer boundary of its service zone, operating locally within the zone. Between its service zone and downtown, boarding is prohibited, but buses remain on the local street, stopping to allow passengers to alight. Outbound, the mirror image policy is adopted—passengers are allowed to board but not alight at any stop before a route's service zone. This arrangement, also called "restricted zonal service," provides direct service between every bus stop pair within the corridor. However, it allows passengers no route choice, since inbound they may use only the route whose service zone includes their origin stop, and outbound they may use only the route whose service zone includes their destination stop. This strategy has proven useful in a number of American cities; one highly successful example is the Massachusetts Bay Transportation Authority's Route 77, a system of two zonal local routes that requires 33 vehicles in comparison with an estimated 40 that would be needed for unzoned local services (FURTH et al.[9]).

As with express zonal service, the advantages of the local zonal strategy are that it reduces vehicle-miles in the outer zones of the corridor, and that it increases vehicle speed slightly since each vehicle can be expected to make fewer stops than it would in conventional local service. However, the fact that inbound travelers who alight outside their route's service zone but before downtown cannot be replaced (because boarding is prohibited) means that this strategy will have more empty seat-miles of operation in the inner zones than would conventional local service, and therefore will require a greater aggregate service frequency than would conventional local service, assuming peak vehicle loads are held constant. The value of this strategy, then, depends both on the degree to which the load profile is tapered at its outer end, and the degree to which the demand is downtown-oriented. A suggested measure for screening corridors for their potential for efficient local zonal service is the "peak volume to uptown boardings" (PV/UB) ratio, defined as:

PV/UB = (peak volume)/(volume boarding before peak volume point)

This measure applies to the inbound direction and should be used when the peak direction is inbound. When the peak direction is outbound, the appropriate measure for the outbound direction is the "peak volume to uptown alightings" (PV/UA) ratio, the ratio of peak volume to the volume alighting after the peak volume point. For ratios above 0.85 or 0.90, local zonal service will likely be an efficient strategy. Corridors with a lower PV/UB (PV/UA) ratio lend themselves better to strategies that do not impose such rigid boarding and alighting restrictions.

In designing local zonal service, one first identifies potential uptown route termini along the corridor's spine, numbering them from 1 to n with increasing distance from downtown. As with express service, sector i is defined as the corridor segment immediately inbound of point i (including point i, but not including point $i - 1$), and route (i, j) is the zonal local route whose service zone is sectors $i, \cdots, j$. At this point, we assume that service zones will be identical in both the inbound and outbound directions. With b_{ij}, p_{ij}, and $c_{ij}(b_{ij})$ defined as they were for express service (but with reference to bidirectional local service), the optimal zonal configuration for local service can be found using the same formulas that were used for express service design.

3. ZONAL DESIGN FOR LOCAL SERVICE: ASYMMETRIC SERVICE ZONES

WHILE IT IS reasonable for local zonal routes to have the same service zone inbound as outbound if the zonal system is used all day long, in many cases local zonal service will be offered during peak periods only. Because of the directional asymmetry in demand that is typical of peak periods, it may be profitable in such a case to allow a route's light direction service zone to differ from its peak direction service zone. Of particular interest is an empty service zone, i.e. one which allows a route to deadhead (carry no passengers) in the light direction, enabling returning vehicles to use the fastest, most reliable path available. To avoid configurations that would be unacceptably complex, we specify the following restrictions:

Restriction "X." A route's light direction service zone may not extend farther out than its peak direction service zone (to avoid deadheading in the peak direction).

Restriction "Y." If a route provides service in both directions of travel, its service zone in each direction must have the same outer boundary.

(Restriction Y is not imposed in the author's original report.[10]) Then we define route (i, j, k) as the zonal local route that operates between point j and the CBD whose peak direction service zone consists of sectors $i, \cdots, j$ and whose light direction service zone consists of sectors $k, \cdots, j$. We employ the convention that $k = D$ if the route deadheads in the light direction. If the route does not deadhead, then a corollary of Restrictions X and Y is that $k \leq i$; for if on a particular route k were greater than i, sectors $i, \cdots, k - 1$ would go unserved by that route in the light direction, and there would be no other route satisfying Restrictions X and Y that could serve them. We define b_{ijk} as the number of buses assigned to route (i, j, k) and p_{ijk} as the minimum number of buses that must be assigned to route (i, j, k) if it is active.

At any given point in the dynamic programming procedure, we are concerned with optimizing service for the segment of the corridor market defined by the stage variable j and the state variable m. This market segment has two components, a peak direction submarket and a light direction submarket. The inner boundary of both submarkets is the CBD. The outer boundary of the peak direction submarket is sector j, the stage variable; for the light direction submarket, the outer boundary is sector m, the state variable. Furthermore, at stage j, j is the outermost route terminus allowed. Therefore, due to Restriction X, only market segments for which $m \leq j$ are feasible; that is, the light direction market may not extend farther from the CBD than the peak direction market. Then we define

$f_j(m)$ = minimum number of buses needed to serve sectors $1, \cdots, j$ in the peak direction and sectors, 1, ⋯, m in the light direction with routes whose outer terminus is no farther from the CBD than j.

(If j or m equals zero, no sectors in the corresponding direction are served.)

The entire corridor is served when $j = m = n$, and so the goal is to find $f_n(n)$, which is done recursively. We begin with $j = m = 0$, for which

$$f_0(0) = 0. \qquad (4)$$

At a particular stage j and state m, the route network must include a route, which we may call a "key route," whose outer terminus is j. The market segment defined by j and m is then split into two parts, the part served by the key route and the remainder, called the remaining market segment. While one parameter of the key route, j, is known, the other parameters, i and k, become decision variables. These parameters determine the inner boundary of the key route's service zone in each direction, and consequently also determine the outer boundaries of the remaining market segment.

Restrictions X and Y impose some constraints on the values that i and k may take. If $m = j$, the key route may not deadhead, but must serve sector j in the light direction. Thus the recursion for the case $m = j$ is

$$f_j(j) = \min_{\substack{i=1,\cdots,j,\\ k=1,\cdots,i}} [p_{ijk} + f_{i-1}(k-1)]$$

$$\text{for} \quad j = 1, \cdots, n. \qquad (5)$$

Here the minimum of buses needed to serve the market segment, $f_j(j)$, is the sum of the minimum number of buses needed on the key route, p_{ijk}, and the minimum number needed to serve the remaining market segment, which is $f_{i-1}(k-1)$ because the remaining market segment's peak direction outer boundary is sector $i - 1$ and its light direction outer boundary is sector $k - 1$.

On the other hand, if $m < j$, then by virtue of Restriction Y the key route *must* deadhead in the light direction, so $k = D$ and the only decision variable is i. Furthermore, i is limited to the range $m + 1, \cdots, j$ in order to avoid leaving the remaining market segment with a longer service zone in the light direction than in the peak direction. The recursion in the case of $m < j$ is then

$$f_j(m) = \min_{i=m+1,\cdots,j}[p_{ijD} + f_{i-1}(m)]$$

$$\text{for} \quad j = 1, \cdots, n-1 \quad (6)$$
$$m = 0, \cdots, j-1$$

where the number of buses needed on the key route is p_{ijD} and the number needed for the remaining market segment is $f_{i-1}(m)$ because the remaining market segment's peak direction outer boundary is $i - 1$ and its light direction outer boundary is m.

To minimize generalized cost, simply redefine $f_j(m)$ as a generalized cost, define the cost function $c_{ijk}(b_{ijk})$ analogous to $c_{ij}(b_{ij})$, and substitute it for p_{ijk} in (5) and (6), making b_{ijk} another decision variable (analogous to (3)). With a practical range of size G for b_{ijk} (i.e., $p_{ijk} \le b_{ijk} < p_{ijk} + G$), this procedure requires $O(n^3G)$ computations. For practical problems, the computational burden is moderate.

4. THE ROUTE MODEL

ONE ASPECT of this approach that has been ignored until now is the origin of the route-specific parameters p_{ijk} and $c_{ijk}(b_{ijk})$ for the $n(n+1)(n+5)/6$ potential routes in asymmetric local design (there are $n(n+1)(n+2)/6$ routes with service in both directions, and $n(n+1)/2$ routes that deadhead) and the values of p_{ij} and $c_{ij}(b_{ij})$ for the $n(n+1)/2$ potential routes in express or symmetric local design. Since a zonal route system as we have defined it allows passengers no route choice, each zonal route is independent, and hence may be analyzed separately. Thus, all that is needed to generate the route-specific parameters is a model of a single bus route. For completeness, the model used in this study is described in the remainder of this section.

Passenger arrivals are assumed Poisson. For convenience, we assume that the stops are numbered in the direction of the bus's travel. If λ_{jk} denotes the rate at which passengers destined for stop k arrive at stop j to use the route under study ($k > j$), then $\lambda_{j.} = \sum_{k>j} \lambda_{jk}$ is the arrival rate at stop j and $\lambda_{.j} = \sum_{i<j}$ ij is the rate at which passengers who will alight at stop j arrive upstream of j. Then if the headway between buses is h, the expected number of boardings at stop j is $M_j = h\lambda_{j.}$ and the expected number of alightings at j is $N_j = h\lambda_{.j}$. The probability that a bus will have to stop at stop j is

$$P_j(\text{stop}) = 1 - \exp[-(M_j + N_j)]. \qquad (7)$$

Expected delay at stop j has three components. The first component, the expected time spent for passenger movements, is assumed to have the form $w_{j1} = a_1M_j + a_2N_j$ where a_1 and a_2 are given parameters. (While this form is appropriate for single door operations, it is admittedly a crude approximation for two-door operations whenever there is a significant number of both boarders and alighters at a stop.) The second component, which is conditional on the bus stopping, is the time lost in deceleration and acceleration, given by $w_{j2} = v_0(1/\text{acc} + 1/\text{dec})/2$, where v_0 is the cruise speed and acc and dec are constant acceleration and decleration rates. The third component, also conditional on the bus stopping, is w_{j3}, a constant accounting for time to open and close doors and return to traffic. Expected dwell time at stop j, then, is

$$W_j = w_{j1} + P_j(\text{stop})(w_{j2} + w_{j3}). \qquad (8)$$

The other components of minimum cycle time are the link travel times, assumed deterministic, and minimum necessary layover, assumed a known, smooth function of run time. Thus, we can establish the minimum cycle time as a function $c(h)$.

There is usually a set of "permissible" headways in keeping with whole-minute, clockface, and policy headway constraints. There will also be a loading (capacity) constraint. To find the minimum number of buses required on a route, we first apply the loading constraint to determine a maximum allowable headway on the route, which is then decreased to its next permissible value, denoted $h_{\max}$. Then the minimum integer number of buses n required to serve the route is given by the fundamental relationship $n = [c(h_{\max})/h_{\max}]^+$, where $[\quad]^+$ means round up to the next integer. Then, in the interest of lowering wait times and loads, the minimum feasible headway $h_{\min}(n)$ is the solution to $h = c(h)/n$, which can be quickly solved by successive linear approximation; $h_{\min}(n)$ is then rounded up to its next permissible value to yield the final headway h. Finally, actual cycle time is nh.

The major level-of-service measures affected by decisions regarding n or h are wait time (including schedule inconvenience) and ride time. Average wait time per passenger is taken to be $a_3 h$, where a_3 is a constant reflecting schedule reliability. For perfectly regular service, $a_3 = 0.5$. A value of 0.6 perhaps better characterizes the typical radial route, and was used in the case study. Ride time can be computed straightforwardly from expected vehicle trajectories.

5. JOINT DESIGN OF EXPRESS AND LOCAL SERVICE IN A CORRIDOR

WHILE SECTIONS 1 to 3 show how to optimally design zonal express and local service for their given service areas, it is not clear how to best choose these service areas in a radial corridor. A "direct service plan" is the specification of which sectors are to belong to the local service area and which are to belong to the express service area. (A sector may belong to both service areas.) Under the assumption that the service area for either service type must be continuous, a particular direct service plan may be summarized as shown in Figure 2 by the variables u, the outermost sector with local service, v, the outermost sector with express service, and w, the innermost sector with express service, where sectors are, as usual, numbered from 1 to n beginning with the innermost sector. Since by definition express service cannot be offered in innermost sector, $w \geq 2$ and sector 1 must have lc service. A stricter lower bound on w can be impo by allowing a sector in the express service area onl express service to/from that sector offers a la enough travel time advantage over local service t CBD-oriented passengers will be attracted to the press route. In order to ensure that service is offe in every sector, either u or v must equal n, and u ≥ − 1. As shown in Figure 2, service plans can be divi into three types. Type (a) is when the outermost sec has both local and express service ($u = v = n$); th are a maximum of $(n - 1)$ such service plans. T (b) is when the outermost sector has only local ser ($u = n$ and $v < n$); there are a maximum of $(n - 1$ $- 2)/2$ such service plans. With type (b) service pla downtown-oriented travelers boarding in the ou most sector(s) will have to board a local route and then face the choice of remaining on that route all way to downtown or transferring at node v to express route. Type (c) is when the outermost sec has only express service ($v = n$ and $u < n$); there a maximum of $n(n - 1)/2$ such service plans. I type (c) service plan, only express service is be offered in one or more outer sectors, so that lo travelers originating in these sectors must be perr ted to use express service as far as the outer bound of sector u, where they would then transfer to lo service. For completeness, the plan in which no press service is offered at all should also be conside

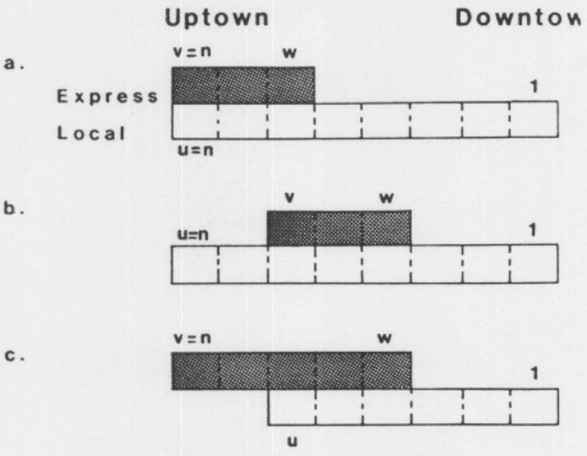

Fig. 2. Types of direct service plans.

Even when service areas overlap, assigning trave to either the express or local service market is straig forward for most practical direct service plans. (some service plans, there may be some passeng whose route choice is not clear, requiring some sor assignment rule and, possibly, equilibration of market segmentation with the supplied level of vice. These complications have not been not furt explored.) Once the market is thus segmented, local and express services can be designed independently. The suggested approach for the joint design of local and express service in a corridor is to evaluate either all or a limited number of feasible direct service plans, optimally designing local and express service for each. (Alternatively, one could jointly optimize express and local service with the direct service plan left as part of the optimization. This approach was taken in the author's original report,[10] but is omitted here because of its complexity and because the simpler two-stage approach is probably as efficient computationally for practical problems.)

6. A SECOND DYNAMIC PROGRAMMING APPROACH TO ZONAL DESIGN

A DIFFERENT dynamic programming approach to zonal route design is presented in this section. These algorithms are more efficient than those of Sections 1–3 for application to the design of branching routes. In this approach, potential route termini (called nodes) are numbered from 1 to n beginning with the outermost stop and ending with the potential terminus closest to the CBD. (This is opposite to the numbering scheme used previously.) Sector i is the sector just inbound from node i, including node i but not including node $i+1$.

For express or symmetric local design, route (i, j) is the route whose outer terminus is node i and whose service zone is sectors $i, \cdots, j$. For asymmetric local design, route (i, j, k) has its outer terminus at node i, its peak direction service zone is sectors $i, \cdots, j$, and its light direction service zone is empty if $k = D$ and is sectors $i, \cdots, k$ otherwise. Route variables such as b_{ij} are defined as before. Then we define, for express or symmetric local service,

$g_i(j)$ = minimum number of buses needed to serve sectors $1, \cdots, j$ with routes whose outer termini belong to the set of node $1, \cdots, i$.

and for asymmetric local service,

$g_i(j, k)$ = minimum number of buses needed to serve sectors $1, \cdots j$ in the peak direction and sectors $1, \ldots, k$ in the light direction with routes whose outer termini belong to the set of nodes $1, \cdots, i$.

To find the optimal express or symmetric local design with an objective of minimizing the number of buses required, solve for $g_n(n)$, the minimum number of buses needed to serve the entire corridor, using the relations

$$g_1(j) = p_{1j} \qquad \text{for } j = 1, \cdots, n \tag{9}$$

and

$$g_i(j) = \min[g_{i-1}(j), p_{ij} + g_{i-1}(i-1)] \qquad \text{for } i = 2, \cdots, n; \; j = i, \cdots, n. \tag{10}$$

Similarly, for asymmetric local design,

$$g_1(j, k) = p_{1jk} \qquad \text{for } j = 1, \cdots, n; \; k = j, \cdots, n \tag{11}$$

and

$$g_i(j, k) = \min[g_{i-1}(j, k), p_{ijD} + g_{i-1}(i-1, k), p_{ijk} + g_{i-1}(i-1, i-1)] \qquad \text{for } i = 2, \cdots, n; \; j = i, \cdots, n; \; k = j, \cdots, n. \tag{12}$$

Equation 10 embodies the choice of whether a route should begin at node i. If so, it must serve sectors $i, \cdots, j$. Equation 12 embodies first the choice of whether a route should begin at node i, and if so, then whether it should deadhead in the light direction or not. These equations may easily be modified for use with an objective of minimizing generalized cost.

The dynamic programming alogrithms represented by Equations 9–10 and by Equations 11–12 still require $O(n^2)$ computations for express or symmetric local design and $O(n^3)$ computations for asymmetric local design, but are more efficient for branching network design, as we shall see.

7. ZONAL ROUTE DESIGN IN BRANCHING CORRIDORS

THE DYNAMIC programming approach used for linear corridors can be extended to zonal design in a branching corridor as long as we continue to specify that service zones be nonoverlapping. A "branching corridor" here denotes a tree-shaped network of streets to which bus service is to be provided and whose root is at the downtown. The same algorithm can be used for express as well as for local service design (symmetric or asymmetric), provided that all variables are defined with reference to the desired service type.

First a reduced network is constructed with nodes only at the extreme points of the original branching network and at junctions, as shown in Figure 3. An arc on the reduced network is thus composed of one or more sectors of the original network. It is convenient to number the nodes on the reduced network in such a way that all the n_e extreme nodes have a lower index than any junction node, and that if node j is closer to the CBD than node i then $j > i$. Arc f is defined on the reduced network as the arc whose outer node (i.e., whose node farther from the CBD) is f. An

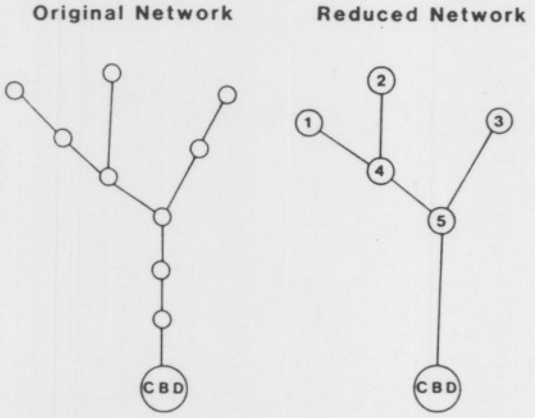

Fig. 3. Reduction of a branching corridor.

arc whose outer node is an extreme node is an extreme arc.

We define $z(e, f)$ on the reduced network as the minimum cost of serving all arcs on the path between extreme arc e and downstream arc f, inclusive. Values of $z(e, f)$ can be computed using one of the algorithms presented in Section 6 by treating the path from e to downtown as a linear corridor. (The algorithms of Sections 1–3 could be used instead; however, the algorithms of Section 6 are more efficient for this purpose because they begin at the outer end of the corridor and proceed inward, making it possible to obtain, for a given extreme node e, values of $z(e, f)$ for every downstream node f in a single pass of the algorithm.) Then

$$z(e, f) = g_j(j) \qquad (13)$$

for express or symmetric local design and

$$z(e, f) = g_j(j, j) \qquad (14)$$

for asymmetric local design, where j in Equations 13 and 14 is the innermost sector of the original network that belongs to arc f of the reduced network.

Next, we define $V(f)$ on the reduced network as the minimum cost of serving arc f and all arcs outbound from f. Then if node m is the junction node immediately upstream of downtown, we find the optimal zonal design for the branching network by solving for $V(m)$ using the relations

$$V(e) = z(e, e)$$
$$\text{for every extreme node } e = 1, \cdots, n_e \qquad (15)$$
$$V(f) = \min_{e \in E_f} [z(e, f) + \textstyle\sum_{k \in U_{ef}} V(k)]$$
$$\text{for} \quad f = n_e + 1, \cdots, m \qquad (16)$$

where

E_f = set of extreme nodes upstream of node f
U_{ef} = set of nodes immediately upstream of any node on path (e, f) excluding nodes on path (e, f).

In the final solution, the branching network will be decomposed into n_e paths, each originating at an extreme node, with service zones positioned on each path as if it were a linear corridor. The choice embodied in Equation 16 is the decision as to which path arc f should belong to. In this sense, then, Equations 15 and 16 represent a process of optimally decomposing a tree-shaped market into a number of nonoverlapping linear market segments. It may be noted that this algorithm requires that the same decomposition used in one direction of service be used in the other. However, it is unlikely that any transit operator would accept any other arrangement.

The restriction of nonoverlapping service zones admittedly limits the applicability of this approach to local service since often the routing strategy desired is to have the "trunk" of a branching corridor served by many branch routes, giving trunk passengers more frequent service. This service strategy and methods for its design are topics of further research.

8. APPLICATION

THE LINEAR corridor design algorithms were applied to the Watertown-Brighton corridor in the Boston area. As Figure 4 shows, this corridor follows a path of local streets from Watertown Square to Kenmore Square, and then follows the Green Line subway tracks through the Back Bay to downtown. The corridor is paralleled by the Massachusetts Turnpike. The four points in Figure 4 that are farthest uptown are presently used by the Massachusetts Bay Transportation Authority (MBTA) as route termini, and thus were chosen as potential route termini in designing a zonal route system. Access to the turnpike is at Newton Corner and at Linden Street. The MBTA presently operates four routes along the corridor: loca Route 57, running between Watertown Square and Kenmore Square, and three express routes, two o which form a zonal route system serving downtown while the other serves Copley Square, the heart of the Back Bay commercial district. Route 57 passengers destined for downtown must transfer at Kenmore Square to the Green Line. Data from morning peak period operations of these routes were used in the analysis, as were policies then in force at the MBTA

Both the entire corridor and Route 57 in isolation were analyzed. The results for Route 57 alone will be presented first. Route 57's load profile, shown in Figure 5, exhibits the characteristic taper at the uptown end and the directional imbalance in flows. However because the competing express routes capture much of the downtown-oriented demand, the PV/UB ratio is only 0.73, suggesting that restricted zonal servic will not be particularly efficient because of its inabilit to replace inbound passengers who alight outside thei

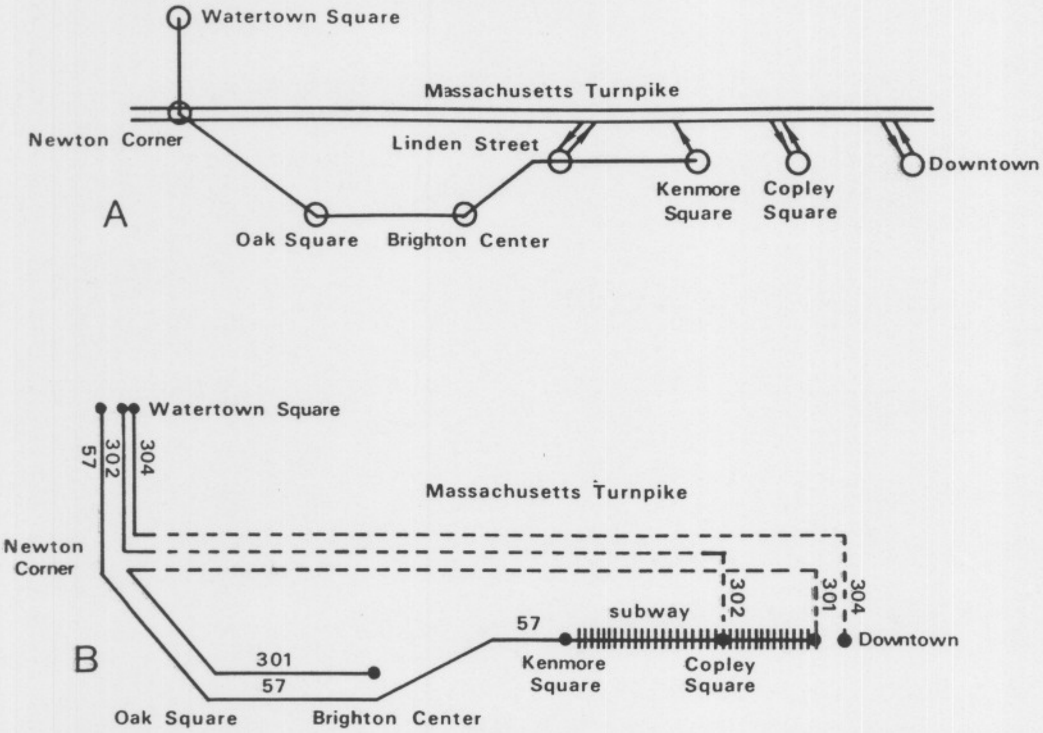

Fig. 4. Streets and existing routes in the Watertown-Brighton corridor. (A) Street network and (B) existing routes.

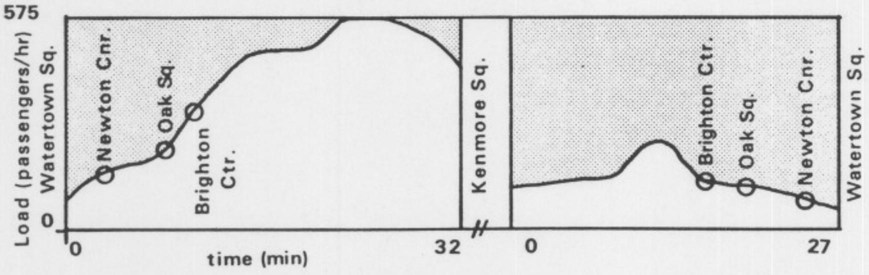

Fig. 5. Route 57 load profile.

service zone but before the peak volume point. As a conventional local route, Route 57 must operate 11 trips per hour, requiring 14 buses, and resulting in an average wait time of 3.2 minutes. The zonal configuration that minimizes the number of buses needed, illustrated in Figure 6, consists of two zonal routes with the shorter route beginning at Oak Square. The shorter route deadheads outbound while the longer route serves the entire outbound demand. The longer route operates 6 trips per hour and shorter route 8. This increase in the aggregate frequency of 3 trips per hour is due mainly to the many passengers on the longer route who alight in the inner zone and cannot be replaced. Because of this inefficiency, only one bus is saved while average wait time rises by 2.5 minutes and average in-vehicle time falls by 0.1 minute.

Next, zonal design methods were applied to the entire Watertown-Brighton corridor. The corridor was simplified by assuming that local buses would not terminate at Kenmore but would continue to downtown via Copley Square, thus eliminating the complication of the bus-to-subway transfer. The PV/UB ratio for the entire corridor is 0.91. If only local service were offered in the corridor, a conventional local route would require 34 buses, with an average wait time of 1.5 minutes and an average in-vehicle time of 22.3 minutes. Zonal design of local service would save 4 buses while raising average wait time by 1.3 minutes and lowering average in-vehicle time by 1.2 minutes. The local zonal configuration, shown in Figure 7, is a system of two routes, with the shorter one beginning at Newton Corner. The shorter route deadheads outbound while longer route serves the entire outbound demand.

If both express and local service are offered in the corridor (with express service provided to downtown

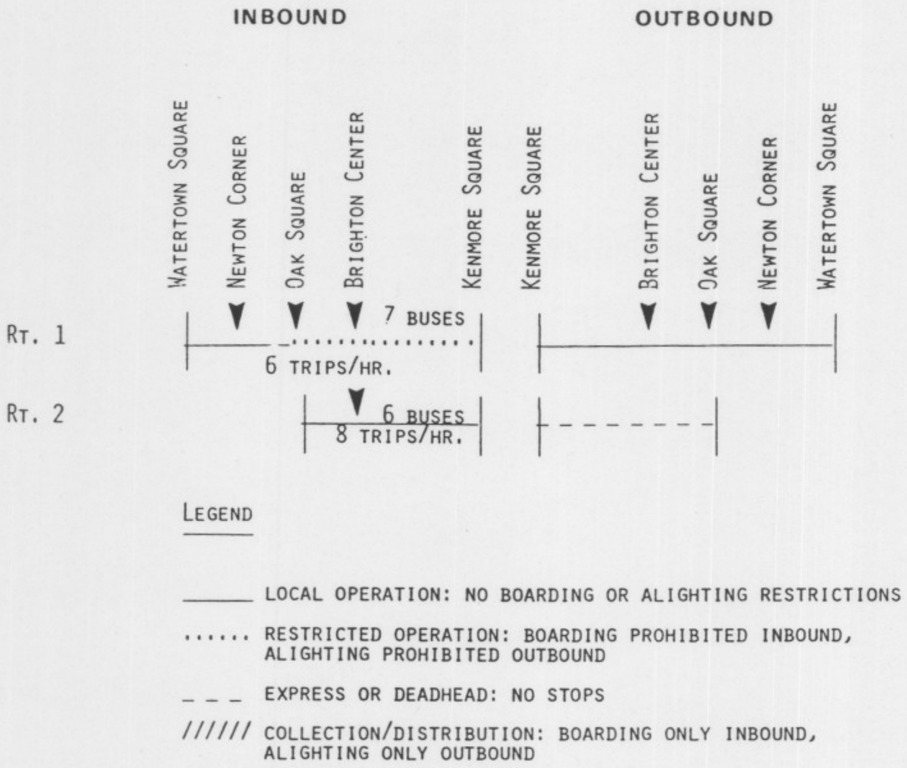
INBOUND
OUTBOUND
WATERTOWN SQUARE
NEWTON CORNER
OAK SQUARE
BRIGHTON CENTER
KENMORE SQUARE
KENMORE SQUARE
BRIGHTON CENTER
OAK SQUARE
NEWTON CORNER
WATERTOWN SQUARE
RT. 1
7 BUSES
6 TRIPS/HR.
RT. 2
6 BUSES
8 TRIPS/HR.
LEGEND
LOCAL OPERATION: NO BOARDING OR ALIGHTING RESTRICTIONS
RESTRICTED OPERATION: BOARDING PROHIBITED INBOUND, ALIGHTING PROHIBITED OUTBOUND
EXPRESS OR DEADHEAD: NO STOPS
COLLECTION/DISTRIBUTION: BOARDING ONLY INBOUND, ALIGHTING ONLY OUTBOUND

Fig. 6. Zonal design for local Route 57.

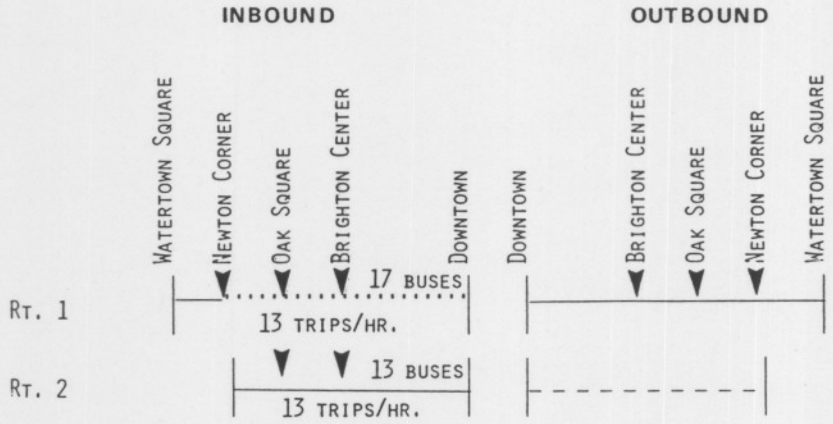

Fig. 7. Zonal design of local service for entire corridor.

only) and the express and local services are jointly designed, the vehicle requirement is reduced to 25 buses, with average wait time rising to 4.3 minutes and average in-vehicle time falling to 17.1 minutes. The routing configuration under this strategy, shown in Figure 8, resembles the existing configuration. It has one express route serving the Watertown Square-Newton Corner sector, a second express route serving the sectors between Newton Corner and Brighton Center, and one local route serving the entire corridor.

The route designs were repeated under an objective of minimizing a sum of operator cost, assumed proportional to vehicle-hours, and passenger wait plus in-vehicle time, valued at $3/hour. When applied t Route 57 only, the conventional local route was foun superior to any zoned configuration because of th adverse wait time effects of zonal segmentation. I designing for the entire Watertown-Brighton corrido the same configuration that minimized the number c buses was found to minimize operator plus passenge cost. Relative to a single conventional local rout serving the entire corridor, the local zonal strateg

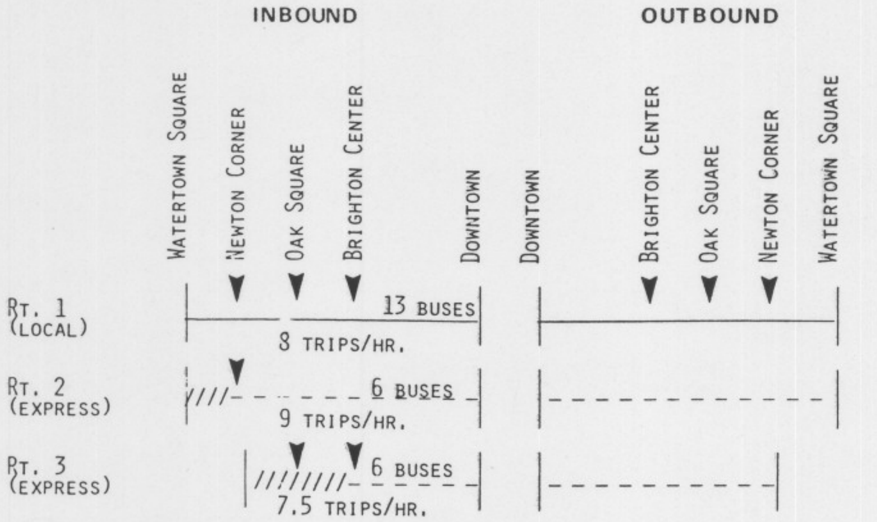

Fig. 8. Zonal design of express and local service for entire corridor.

reduced total operator plus passenger cost by 4% and the express/local zonal strategy reduced total cost by 16%.

The corridor was also modeled more realistically with its three downtown termini and its subway connection between Kenmore Square and downtown. Local routes were required to terminate at Kenmore Square, in keeping with MBTA policy, while express routes could serve either Copley Square or downtown. It was assumed that passengers bound for Copley Square or downtown would use an express route if available; if not, it was assumed that they would use a local route to Kenmore, where all downtown-bound passengers and 80% of the Copley-bound passengers would transfer to the Green Line while the remaining Copley-bound passengers would walk to their final destination. Because the peak load point on the Green Line is sometimes just before and sometimes just after Copley Square, the marginal subway operating cost was taken to be $0.30 per Copley-bound passenger and $0.60 for downtown passengers. Average transfer time at Kenmore Square was assumed to be 3 minutes. A 10-minute walk time was added to wait time for each Copley-bound passenger who walked from Kenmore to his destination.

Since MBTA policy permitted three types of routes (local to Kenmore, express to Copley, express to downtown), a direct service plan had to be chosen in order to specify the service area of each service type. Nine alternative direct service plans were identified as being practically feasible. They are listed in Table I. Alternative 4 is the direct service plan currently in place in the corridor. Zonal design was then performed for each service type under each alternative with an objective of minimizing the number of buses needed. The results are compiled in Table II. Omitted from Table II are operator and passenger costs incurred beyond Kenmore Square due to passengers originating downstream from Linden Street since these costs are unaffected by the design.

TABLE I

Alternative Direct Service Plans

Corridor segments:

1. Linden Street-Kenmore Square
2. Brighton Center-Linden Street
3. Oak Square-Brighton Center
4. Newton Corner-Oak Square
5. Watertown Square-Newton Corner

Direct Service Plan	Segments with Express Service to Downtown	Segments with Express Service to Copley	Segments with Local Service	Transferring Passengers[a]
1	5	—	1–5	1020
2	5	5	1–5	754
3	3–5	—	1–5	544
4[b]	3–5	5	1–5	278
5	3–5	3–5	1–5	135
6	2–5	—	1–5	450
7	2–5	5	1–5	184
8	2–5	3–5	1–5	41
9	2–5	3–5	1–5	0

[a] Downtown- and Copley-bound passengers per hour lacking direct service who must either transfer to subway or walk from Kenmore.

[b] Existing direct service plan.

As a benchmark for comparison, the routing configuration now existing in the corridor was optimized with respect to route headways; its costs appear also in Table II. The results for alternative 4 differ from this benchmark because the existing routing configuration is not the most efficient for the existing direct service plan.

As Table II indicates, the alternative with the small-

TABLE II

Impacts of Minimum Operator Cost Design

Direct Service Plan	Operator Impacts		Total Operator Cost[a]	Passenger Impacts		Percent Transferring[b]
	Buses needed	Subway operator cost[a]		Average wait + transfer time[a]	Average in-vehicle time[a]	
1	31	$450	$1444	4.5 min	18.1 min	41%
2	30	386	1384	4.7	16.9	30
3	32	164	1190	4.7	16.2	22
4	32	101	1126	5.2	15.2	11
5	33	66	1123	4.1	16.3	5
6	32	108	1134	4.7	16.4	18
7	31	44	1038	4.4	15.4	7
8	32	10	1036	4.3	16.5	2
9	33	0	1058	4.6	15.3	0
Existing route structure	33	101	1158	4.0	15.2	11

[a] Excludes subway costs of Copley- and downtown-bound passengers boarding in Segment 1.

[b] Of all passengers, percentage boarding in Segments 2–5 bound for downtown or Copley who lack direct service and hence must either walk or transfer to subway.

est operating cost is the eighth. This alternative, which is to extend the service area for downtown express service to Linden Street and to extend the service area for Copley express service to Brighton Center, requires one bus fewer than the existing configuration and reduces hourly subway costs by $91 because it relieves the Green Line of serving 237 passengers per hour. These passengers are also saved the trouble of transferring at Kenmore. Average wait time for this alternative is 0.3 minute higher than the existing configuration, and in-vehicle time is 0.2 minute longer. (The longer in-vehicle time results because the larger service area for the Copley express route lengthens travel time for many Copley-bound travelers.) The overall savings in operating cost is 10%.

Service design was also performed under the objective of minimizing operator plus passenger cost, which for this case included a transfer penalty of $0.20 per transferring passenger. It was found that little improvement could be made over the existing configuration. Alternative 9 offered the greatest improvement in overall cost, a decrease of only 2%. This small gain reflects the fact that the existing configuratio already well tailored to this objective.

It should be noted that revenue was ignored in analysis, whereas an operator would be likely to clude revenue in his objective function. Then if tl were a difference in the revenue received from a senger using an express bus and a passenger usir local bus and transferring to subway, operators m tend to favor alternatives that forced more pasen to use the higher priced path. (However, if prices different paths differed substantially or if new pa were created, a considerable demand response wc be expected.)

REFERENCES

1. F. J. SALZBORN, "Timetables for a Suburban Rail T sit System," *Trans. Sci.* **3,** 297–316 (1969).
2. G. C. CLARENS AND V. F. HURDLE, "An Opera Strategy for a Commuter Bus System," *Trans. Sc* 1–20 (1975).
3. M. A. TURNQUIST, "Zone Scheduling of Urban Routes," *Trans. Eng. J. ASCE* **105,** 1–12 (1979).
4. B. HAGBERG AND D. HASSELSTROM, "A Metho Assign Frequencies and Vehicle Types to Fixed R Urban Public Transport Systems," AB Volvo, Tr portation Systems, working paper 80-DH-33, Got burg, Sweden, 1980.
5. H. M. WAGNER, *Principles of Operations Research,* 2, pp. 314–320, Prentice-Hall, Englewood Cliffs, 1975.
6. G. KOCUR AND C. HENDRICKSON, "Design of Local Service with Demand Equilibrium," *Trans. Sci.* **16,** 170 (1982).
7. S. C. WIRASINGHE, V. F. HURDLE AND G. F. NEW "Optimal Parameters for a Coordinated Rail and Transit System" *Trans. Sci.* **11,** 359–374 (1977).
8. V. F. HURDLE, "Minimum Cost Schedules for a P Transporation Route," *Trans. Sci.* **7,** 109–157 (197
9. P. G. FURTH, F. B. DAY, AND J. P. ATTANUCCI, "O ating Strategies for Major Radial Bus Routes," DO 84-27, UMTA, 1984.
10. P. G. FURTH, "Designing Bus Routes in Urban C dors," Ph.D. thesis, MIT Department of Civil Engin ing, 1981.