Setting Frequencies on Bus Routes: Theory and Practice

PETER G. FURTH AND NIGEL H.M. WILSON

Since most transit systems have relatively stable route structures and politically determined levels of subsidy, one of the main recurrent decisions the transit planner must make is the service frequencies to be provided on each route in the system. Current practical and theoretical approaches to this problem are reviewed and, in light of their seeming inadequacies, a new model for setting frequencies is developed. The model allocates the available buses between time periods and between routes so as to maximize net social benefit subject to constraints on total subsidy, fleet size, and levels of vehicle loading. An algorithm is developed to solve this nonlinear program that can be applied by using a small computer program or, simplified in some generally acceptable way, by using a pocket calculator. In a case study the model is shown to produce results quite different from the existing allocation, which suggests changes that are insensitive to the specific set of parameters and objectives. It is shown that the model can readily be applied to evaluate the impacts of an alternative vehicle capacity and to investigate the value of changing service policies.

The North American public transit industry has, in the past decade, emerged from a long period of stagnation and decline to become a major focus in strategic planning to deal with the energy problem. Increasing attention is being given to the problem of using the ever-mounting public resources being devoted to transit more efficiently. This attention has revealed an apparent enigma: Although there is a wealth of academic research on how transit planning should be done, methods in use in the transit industry are generally crude and dominated by the planners' experience and judgment, sometimes codified into simple rules of thumb.

In this paper, one important part of the short-range transit planning process is selected and used to investigate whether significant differences exist between current practice and reasonable theory (1). The topic is setting frequencies on bus routes, a problem that must be addressed, either explicitly or implicitly, several times each year by all transit operators. After a discussion of existing industry practice in setting frequencies, prior research is briefly reviewed. In light of the weaknesses identified in this prior work, a new model is proposed that accurately reflects the objectives and constraints with which the transit industry must deal. Finally, a case study of part of the Massachusetts Bay Transportation Authority (MBTA) system shows the differences between the actual allocation of buses and that suggested by the theory.

CURRENT PRACTICE

Methods used by schedulers to set frequencies on routes are generally poorly documented and seem to vary among operators. Typically, however, only a small number of rules of thumb have been used that can be overridden by the judgment and experience of the scheduler. The best way of assessing industry practice is to refer to the service standards that have been widely adopted by many operators in the past five years. Service standards cover a broad range of planning, operations, and management and (of interest here) usually include specific guidelines on service frequencies. These service standards are a result of both codification of existing rules of thumb and a statement of policy. As such they do not always accurately reflect decisions made by schedulers (and others) but are likely to include factors traditionally used in decision making.

Based on a survey of existing service standards (2), the most frequently used methods for setting frequencies are policy headways, peak-load factor, revenue/cost ratio, and vehicle productivity. Each of these is described briefly below.

Policy Headways

Policy headways are used by virtually all operators and serve as a lower bound on the frequency. Routes are categorized by factors such as orientation (radial or crosstown), function (line-haul or feeder), and location (urban or suburban); and each category is assigned a set of policy headways for each period of the day. Policy headways are most effective in systems that operate principally as a low-demand social service. However, in large cities, during peak hours, and whenever demand is high, policy headways lose their relevance and other methods must be used to assign headways.

Peak-Load Factor

The ratio of the number of passengers on board at the peak-load point to the seating capacity of the vehicle is widely used under heavier demand. A lower bound on frequency is based on maximum peak-load factors established by route category and time period. These factors are based on the physical capacity of the vehicle and on comfort and operational considerations.

Revenue/Cost Ratio

The revenue/cost ratio is often used to define an upper bound on the amount of service to be provided on a route. This ratio is a rough measure of efficiency and equity in the distribution of service and has the important advantage of being readily understood by both elected officials and the general public.

Vehicle Productivity

Either in the form of passengers per vehicle mile or per vehicle hour, vehicle productivity is also occasionally used to set upper bounds on the frequency. As in the case of the revenue/cost ratio, vehicle productivity is used to approximate the benefit/cost ratio of a specific service and to guard against inefficient allocation of resources.

Although service standards are an advance in the state of the art of transit planning, this brief review shows that they fall far short of ensuring that transit resources are allocated most efficiently (3). Specifically, the standards focus on upper and lower bounds for setting frequencies but say nothing about setting frequencies to maximize efficiency within these constraints.

To better understand how frequencies are actually established, a set of MBTA routes was analyzed and a set of empirical relationships tested by linear regression. The 17 routes analyzed all belong to the Arborway Garage of the MBTA and include a wide variety: radial and crosstown, high- and low-frequency, that serve affluent and poor neighborhoods. Results show that the midday frequencies are heavily constrained by the policy headways; only four routes have a higher frequency.

The following three empirical relationships for setting frequencies were tested:

Equal load factor (2-h peak):

$$Q = b(PLP)$$

$$Q = a + b(PLP) \qquad (1)$$

Equal load factor (30-min peak):

$$Q = b(PLP)$$

$$Q = a + b(PLP) \qquad (2)$$

Square-root rule:

$$\ln Q = a + b\ln(r/T) \qquad (3)$$

where

a,b = coefficients,
Q = scheduled frequency (round trips/h),
PLP = peak-load-point count (riders/h that pass peak-load point),
r = ridership per hour (total boardings in both directions), and
T = round-trip run time (min).

The most important results for the morning peak period are shown below (all estimates of coefficients are significant at the 99 percent level):

Empirical Relationship	Coefficients a	b	R^2
Equal load factor	-	0.024	0.85
(2-h peak)	1.54	0.020	0.93
Equal load factor	-	0.018	0.90
(30-min peak)	1.32	0.016	0.95
Square-root rule	0.43	0.72	0.82

It appears that existing frequencies are very well explained by setting the peak-load factor equal on all routes, particularly during the peak half-hour. The average peak load on these routes during the peak half-hour was 1.2, which is about 13 percent below the policy peak-load factor of 1.4.

This case study suggests that schedulers do follow a clear decision-making process, which revolves around the rule of an equal peak-load factor. These results are strikingly similar to those found by Morlok in Chicago (4); the important point is that in both cases the equal peak loads were significantly below actual bus capacity. As demonstrated in the next section, this fact makes the rule inefficient with respect to the optimization of passenger service.

PREVIOUS THEORY

The best-known theory for setting frequencies on bus routes is the square-root rule, which is based on the minimization of the sum of total passenger wait-time costs and total operator cost. In the general case when routes of different lengths exist, the rule states that the service frequency provided on a route should be proportional to the square root of the ridership per unit distance (or time) for that route (5).

Major weaknesses of the square-root rule, which explains its lack of acceptance by the industry, are that it does not consider bus capacity constraints and that it assumes that ridership is fixed and independent of the service frequency. Ignoring the capacity constraint means that on some heavily used routes not enough capacity will be provided (i.e., the solution is infeasible). The assumption of fixed demand means that the user benefits are limited to minimization of wait time, which is probably only a minor part of the public benefit of transit service.

A second, almost trivial, theory is that if the objective is simply to minimize operator cost, the frequencies should be set so that the capacity will equal the peak load on each route. If the system is at capacity, each route will have the same peak-load factors, but if the system is operating below capacity, efficiency arguments do not lead to equal peak-load factors.

Guinn (6) used a linear-programming approximation to allocate buses to maximize revenue subject to a fleet-size constraint. Although his objective and constraint set are too approximate for direct application, the model presented later in this paper uses the same general optimization framework. Scheele (7) proposed a more complex mathematical programming approach to determine optimal service frequencies in the long-run case in which the distribution of trips (but not total trip generation and production) is allowed to vary in response to the service provided.

Several models have been developed for the simultaneous choice of routes and frequencies (8-10). The frequency components of these models typically minimize passenger wait time subject to capacity constraints under an assumption of fixed demand. Recent work at the Volvo Bus Corporation (11) has resulted in a package for choosing routes and frequencies that has been successfully applied in numerous cities in Europe and elsewhere.

Most of these models and theories are designed for one-time application when the entire transit network is redesigned--by definition a major and infrequent undertaking. Furthermore, only one of these models has been applied frequently, and none has been accepted for use by transit operators. This is both because of their orientation to large-scale system change and because they are either complex and hard to use or crude and hard to believe. There is need for a model that accurately reflects the frequency-choice decision, that is simple enough in its data and application requirements to be used frequently by operators, and that focuses on small changes so it can be applied repeatedly over the years. Such a model is developed in the remainder of this paper.

PROPOSED MODEL

The fleet-allocation problem can be formulated as an optimization in which an objective function is maximized (minimized) subject to a set of constraints. Before the formulation is presented and discussed in detail, however, it is useful to illustrate the style of solution by using a simple example.

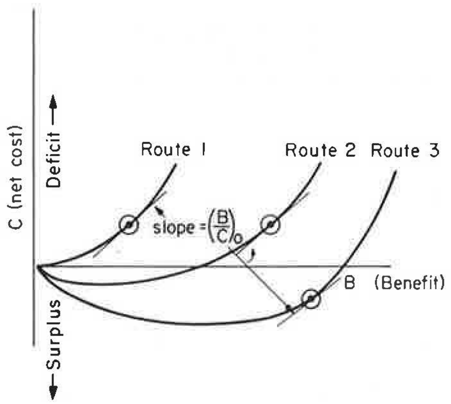
Suppose that a bus company operates three routes, charges a flat fare per passenger, and has allocated a fixed amount to cover the deficit that will result from providing the service. The single objective of the company is to maximize ridership by means of allocating buses, given that fares, routes, and operating speeds are fixed.

The problem can be viewed as a resource-allocation problem: How can the limited resources (subsidy) be allocated to maximize the benefit (ridership)? As shown in Figure 1, for each route a curve that relates net cost (deficit) to benefit can be obtained by varying the frequency of service on the route. At an optimal allocation, the ratio of marginal benefit to marginal cost should be the same for each route. Denoting the benefit of route i by B_i, its net cost by C_i, and its frequency by Q_i, this rule can be written as follows:

$$(dB_i/dQ_i)/(dC_i/dQ_i) = (dB_i dC_i) = (B/C)_o \qquad (4)$$

As suggested in Figure 1, at the optimum some routes may be operating at a profit and others at a loss; however, the total benefit cannot be increased by

Figure 1. Efficiency in subsidy allocation.



shifting resources from one route to another. The optimum occurs when the marginal rate of return on each route is the same and is sufficient to exhaust the available subsidy.

With this simple example in mind we will now turn to the real problem, which will be formulated as a mathematical program. In the next two sections the objective of the optimization is defined and the constraints are specified.

OBJECTIVE

Increasing attention has been paid to the objectives of transit operators since 1975 when London Transport enunciated its objective of maximizing passenger miles (12). Defining objectives is an important step in developing good management practice in public transport agencies. Since in general transit has been recognized to be an important social service, presumably the general objective should be maximization of the social surplus. In the case of determining service frequencies while holding all other attributes of the system fixed, the objective includes two distinct components--consumer surplus and externalities associated with transit ridership.

It can easily be shown that the consumer surplus is the saving in wait time that accrues to system riders who would have been prepared to ride at lower frequencies (and thus endure longer waits). In the remainder of this paper, the mean passenger wait time will be assumed to be half the mean headway--based on the sample model of random passenger arrivals, regular headways, and buses not operating close to capacity. If the demand function is $r = r(h)$, where r is ridership and h is mean headway, the saving in wait time at headway h* is as follows:

$$h^* = \tfrac{1}{2}\int_{h^*}^{\infty} r(h)dh \qquad (5)$$

It can be argued that the major motivation for subsidization of transit service is not saving in wait time. Other, probably more significant, public benefits include mobility for those without automobiles; reductions in congestion, pollution, and energy use; and land use effects. These positive externalities are largely collinear with the ridership, and so a social ridership benefit can be defined crudely as being proportional to the number of riders. This marginal social ridership benefit would logically vary between ridership classes and time periods, which reflects the extent to which attracting different types of riders contributes to the social objectives of providing transit service. This term must be weighted to reflect the value of an additional rider as it relates to saving in wait time.

The objective function then consists of these two components--wait-time saving and ridership. One of the important questions that will be addressed later is how sensitive the service frequencies are to changes in the relative importance of these two objectives.

CONSTRAINTS

Four sets of constraints are included directly in the mathematical program--subsidy, fleet size, policy headways, and loading. Typically, an operator has a fixed level of subsidy available for the planning period (e.g., one year), and the solution that maximizes total benefit will inevitably exhaust the entire subsidy. Since we are concerned with short-range planning, the operator has a limited fleet to allocate, which may vary between periods of the day because of preventive-maintenance needs. As discussed earlier in this paper, two constraints currently used by many operators in setting frequencies are policy headways (which stipulate a maximum allowable headway) and peak-load factors (which specify the maximum load at the most heavily loaded point on the route). For any route during any period clearly only one of these two constraints can be binding, and so the mathematical program includes constraints that require that each headway satisfy the more binding of these two constraints.

In addition to these formal constraints included in the model, there is another set of constraints not included in the model, which can be dealt with externally. In general some services may be mandated for reasons other than social benefit as narrowly defined above; buses and subsidy should be set aside for these required services before the optimization problem is solved, and these services are simply added to the solution to produce the recommended set. Often only an integer number of buses can be assigned to a route (although interlining is a common means to circumvent this requirement); this constraint can be handled by adjustment of the final frequencies. Finally, some interdependencies between routes can be incorporated directly into the objective function. For example, if two routes should have the same frequencies, e.g., for timed transfers, they can be included as a single decision variable; this would help to reduce the size of the problem.

PROBLEM FORMULATION AND DISCUSSION

The problem can be stated as follows: Find the frequencies on each of a number of routes that maximize net social benefit subject to constraints on total subsidy, fleet size, and maximum headways. In the following formulation, headway is used as the basic decision variable.

Maximize:

$$Z = \sum_{j=1}^{P} D_j \sum_{i=1}^{N_j} \left[(b/2)\int_{h_{ij}}^{\infty} r_{ij}(u)du + a_{ij}\, r_{ij}(h_{ij})\right] \qquad (6)$$

Subject to the following constraints:

Subsidy:

$$\sum_{j=1}^{P} D_j \sum_{i=1}^{N_j} \left[(k_{ij}/h_{ij}) - F_{ij}\, r_{ij}(h_{ij})\right] = S_o \qquad (7)$$

Fleet size:

$$\sum_{i=1}^{N_j} (T_{ij}/h_{ij}) \leq M_j, \quad j = 1, \ldots, P \qquad (8)$$

Headway:

$$h_{ij} \leq x_{ij}, \quad j = 1, \ldots, P \quad i = 1, \ldots, N_j \tag{9}$$

where

- P = number of time periods,
- N_j = number of routes operated during time period j,
- D_j = duration of period j,
- b = value of wait time,
- h_{ij} = headway on route i during period j,
- a_{ij} = surplus marginal ridership benefit on route i during period j (marginal ridership benefit minus fare),
- r_{ij} = ridership on route i during period j (a function of h_{ij}),
- k_{ij} = operating cost per run on route i during period j,
- F_{ij} = fare on route i during period j,
- S_o = subsidy available,
- T_{ij} = run time (round trip) on route i during period j,
- M_j = fleet size during period j, and
- x_{ij} = maximum headway for route i during period j.

The objective function (Equation 6) can easily be shown to be equivalent to maximizing the wait-time savings plus the social-ridership benefit minus operating cost; this is the net social benefit. Equation 7 simply states that the operating cost minus the revenue must be equal to the known subsidy. Equation 8 is the fleet-size constraint, and Equation 9 constrains the headway to be less than the policy headway and the headway at which the loading constraint is binding.

This general formulation could be simplified or made more complex (for example, by defining classes of riders, each of which has a separate marginal benefit) in specific applications, but all important facets of the problem are included. Before the method developed to solve this mathematical program is presented, it is necessary to recognize and discuss perhaps the most important limitation of the model--the assumption of the independence of all routes in the system.

It is because both costs and benefits due to a headway on a specific route have been assumed independent of headways on other routes that the problem formulation is so straightforward, but this assumption is not always true, at least on the benefit side. In this model, ridership on a route depends on the headway of only that route, whereas, in general, ridership will also depend on the headways on competing and complementary routes.

When passengers have a choice among several routes, an improvement in service on one of those routes will divert riders from the other routes. Such route competition is less common in North America than in other parts of the world, in which an approach that directly considers route competition is called for (13). An improvement in service on one route can also raise the demand on another route when there is a large transfer volume between the two routes. Care must be taken, therefore, both in applying the model and in interpreting its results in situations in which strong route competition or complementarity exists.

THE ALGORITHM

Optimality (Kuhn-Tucker) conditions can be derived as a set of equations that relate headways to the other variables in the model; these equations then become the optimal decision rules for the operator. These optimality conditions are applied in the following step-by-step algorithm to determine the optimal set of headways by route and by period:

Step 1: Relax the fleet-size and maximum-headway constraints on all routes and for all time periods not yet constrained and solve the following set of equations for the headways, h_{ij}:

$$(b/\lambda) \cdot (r_{ij}/2)\, h_{ij}^2 - \left\{F_{ij} + [(a_{ij}/\lambda)(dr_{ij}/dh_{ij})]\, h_{ij}^2 - k_{ij}\right\} = 0 \tag{10}$$

where λ is determined to exhaust the available subsidy. If no routes violate their maximum-headway constraint, go to step 3.

Step 2: For routes and periods for which the maximum-headway constraint (Equation 9) is violated, set $h_{ij} = x_{ij}$. Compute the deficit incurred on those routes and reduce the available subsidy by that amount. Go to step 1.

Step 3: Identify time periods in which the fleet-size constraint (Equation 8) is binding. For each of these time periods solve the following set of equations:

$$(b/\lambda) \cdot (r_{ij}/2) \cdot h_{ij}^2 - \left\{F_{ij} + [(a_{ij}/\lambda)(dr_{ij}/dh_{ij})]\, h_{ij}^2 - (k_{ij} + w_j T_{ij})\right\} = 0 \tag{11}$$

where w_j is the shadow price of run time during period j and is determined to use all available buses.

Step 4: If no routes violate their maximum-headway constraint (Equation 9), go to step 5. Otherwise, for every route that violates its maximum-headway constraint, set $h_{ij} = x_{ij}$. Compute the number of buses required by all such routes in each period j and reduce the number of available buses in period j by this amount. Go to step 3.

Step 5: Compute the deficit incurred by the fleet-constrained time periods and reduce the available subsidy by this amount. Let $\lambda_c = \lambda$.

Step 6: Repeat steps 1 and 2 for the unconstrained time periods to find a new value of λ, which is λ_u.

Step 7: If $\lambda_u \simeq \lambda_c$, stop; otherwise set $\lambda = \lambda_u$ and return to step 3.

The theory behind this algorithm will not be presented in detail here (1). However, the computational burden of the algorithm is very small, since it consists basically of a sequence of one-dimensional searches that are performed very rapidly. Equations 10 and 11 can be solved very efficiently by using the Newton method (provided the demand function has continuous second derivatives), and values of λ and w_j can be found by making successive linear approximations.

CASE STUDY

The Arborway Garage of MBTA, which serves 21 bus routes, was chosen to illustrate the capabilities of the model. Fifteen of these routes were included in the analysis; the others were excluded for one of the following reasons: incomplete ridership data, highly irregularly scheduled runs, and interdependence of routes.

The most important data and assumptions made in the study are summarized here:

1. Two time periods were examined--the morning peak (7-9 a.m.) and midday (10 a.m.-2 p.m.). Scheduled headways were approximately constant during each of these periods.

2. Scheduled round-trip running times were used with a layover time of 25 percent of the run time. This is above MBTA's policy of 10-20 percent of run time for layovers but slightly below the average observed figure of 28 percent.

3. Costs per run were based on MBTA's figures of cost per vehicle mile and per driver hour.

4. The systemwide average revenue per bus ride of 18¢ was used.

5. The current deficit incurred on this part of the system was used as the available subsidy, and the current number of vehicles used in the morning peak was used as the fleet-size constraint in each period.

6. Policy headways of 30 min in the peak period and 60 min in the off-peak period were used based on current MBTA service policies.

7. MBTA uses two peak-period load-factor standards: For the 2-h peak period, the peak-load factor should be no greater than 1.2; for the peak half-hour it should be no greater than 1.4. Actual peak loads were estimated for both periods for each route based on peak-point counts taken over a three-year period. Only one of these load-factor constraints will be binding for each route. The off-peak policy load factor of 1.0 adopted by MBTA was also used.

8. Bus seating capacity of 45 was used for the load-factor constraints.

9. Current route ridership was taken from a 1978 on-board survey.

More-detailed discussion is warranted about the demand model and the operator's objectives. A binary logit demand model was used that has assumed coefficients for wait time taken from another study. Estimates of the base transit market share were also made based on mode-split characteristics of the Boston area. These assumptions implied wait-time elasticities of demand of -0.2 in the peak period and -0.5 in the off-peak period, which are within the range observed in other U.S. cities (14). It is hoped that advances in the state of the art of demand forecasting at the route level will soon obviate the need for such assumptions. In this case study, sensitivity analyses demonstrated that the results were very robust with respect to these parameters.

The model allows an objective function that consists of a weighted sum of total passengers and total passenger wait-time savings. The absolute coefficients of these terms do not have to be exogenously specified, but their ratio does. The initial ratio chosen implied a trade-off of one passenger for 12 passenger-min of wait time.

Table 1 shows the resource allocation between routes and between periods as suggested by the model compared with the current MBTA allocation. The results in terms of deficit, number of buses, and changes in wait-time and ridership benefits are given in Table 2. The most striking result is that only 59 of the 70 available buses are used in the peak period, and the peak period's share of the deficit declines accordingly. Only 44 percent of the total subsidy is allocated to the peak period by the model compared with 58 percent in the current system. The peak period is heavily constrained by capacity; nine routes operate at the maximum load during the peak half-hour. In general, the shorter routes have the smallest loads and so do not necessarily have the highest revenue/cost ratio. As expected, midday loads are much lower than those during the peak period.

Several factors contribute to this large shift in

Table 1. Frequency on case-study routes: actual and recommended.

	Actual			Recommended		
Route	Frequency (buses/h)	Peak ½-h Load (passengers)	Revenue/Cost Ratio	Frequency (buses/h)	Peak ½-h Load (passengers)	Revenue/Cost Ratio
Morning Peak						
21	5.0	63[a]	0.55	5.0	63[a]	0.54
24	4.0	42	0.50	4.0	42	0.49
25	5.0	35	0.37	3.7	44	0.47
28	3.0	54	0.50	3.6	48	0.44
29	13.3	57	0.60	12.0	63[a]	0.65
31	4.0	39	0.30	2.9	49	0.37
32	15.0	60	0.52	14.2	63[a]	0.55
35	5.0	53	0.32	4.0	63[a]	0.36
36	10.0	51	0.49	8.0	63[a]	0.60
37	5.0	54	0.42	4.2	63[a]	0.49
38	5.5	34	0.39	2.6	59	0.32
41	6.0	56	0.67	5.3	63[a]	0.74
46	2.0	29	0.63	3.1	21	0.46
50	3.3	59	0.39	3.0	63[a]	0.41
51	4.0	55	0.26	3.3	63[a]	0.29
Midday						
21	1.3	13	0.22	1.8	13	0.21
24	1.5	15	0.37	2.5	13	0.30
25	3.0	11	0.35	3.0	11	0.35
28[b]	0.0	-	-	0.0	-	-
29	5.0	38	0.85	5.6	35	0.77
31	1.5	13	0.58	3.4	9	0.40
32	4.6	28	0.60	4.7	28	0.58
35	2.0	33	0.48	3.2	26	0.38
36	2.0	40	0.57	3.5	30	0.42
37	2.0	24	0.44	3.0	20	0.36
38	2.7	15	0.20	2.0	17	0.23
41	3.5	24	0.66	4.2	21	0.58
46	2.0	7	0.18	1.5	7	0.20
50	2.0	20	0.28	2.3	19	0.27
51	2.0	21	0.23	2.0	21	0.23

[a]Capacity constrained. [b]Route 28 is not operated in the off-peak period.

Table 2. Deficit, number of buses, and change in wait-time and ridership benefits: actual and recommended.

	Morning Peak		Midday	
Item	Actual	Recommended	Actual	Recommended
Deficit ($)	2175	1651	1588	2112
Buses operating	70	59	27	32
Change in benefit for				
Wait time ($)	-	−177	-	607
Ridership ($)	-	−103	-	431

resources from the peak period to midday relative to the current system. First, reducing headways in the off-peak period, when base headways are higher, is more effective in reducing total wait time than applying the same resources to reduce headways in the peak period. Second, demand elasticity with respect to wait time is higher in the off-peak period, and so reducing headways in that period is more effective in increasing ridership.

Experiments that vary both the demand parameters and the objective function weights, described in the following paragraphs, consistently suggested shifting resources from the peak to the midday period. This leads to the strong recommendation that midday services be expanded at the expense of the peak periods. Such a shift could be expected to reduce costs by 10-15 percent and result in a reduction in the deficit incurred of about 20 percent.

Several experiments were run that varied the ratio of the wait-time value to the marginal ridership benefit from zero to infinity to study the sensitivity of resource allocation between routes and periods. The resulting allocations between periods were almost identical; changes in number of buses and share of the deficit allocated to the peak period were less than 2 percent over the full range of objective function weights. Variations between routes within the same period were of similar magnitude, which supported the finding that the relative weights given to the two objectives have little impact on the optimal allocation. This is to be expected, since any action the operator takes to decrease wait time will also tend to increase ridership, and vice versa.

Perhaps the greatest weakness in this case study is the uncertainty about the demand function and its parameters. To test the importance of this uncertainty, a set of experiments was run that varied the demand parameters to see whether resource allocation changed significantly. As the headway coefficient was varied from zero to 150 percent of its base value, the total variation in resource allocation between time periods was less than 5 percent, and the variation between routes did not exceed 10 percent. Similar lack of sensitivity to the assumed absolute and relative transit-market shares between periods and between routes was observed.

An important observation from these experiments is that the results obtained when demand is assumed to be inelastic differ little from those when a more realistic demand model is used. This lack of sensitivity is not altogether surprising, since providing the best service for current customers is usually a good way to attract new customers. When demand is assumed fixed, the solution algorithm becomes computationally much less complex, since Equations 10 and 11 can then be solved in closed form, which makes this procedure one that could be performed by using a programmable calculator or even manually.

The model was also used to explore one policy question, Would higher-capacity buses be beneficial on some routes? Striking results were found by increasing bus seating capacity from 45 to 53; the total benefit (the value of the objective function) increases by one-third. Only on three routes was the capacity constrained, the number of buses required in the peak declined from 59 to 55, and the peak period's share of the deficit declined from 44 to 38 percent.

This analysis shows the value of the proposed allocation model in policy terms and also suggests that in this case real benefits may accrue from using larger vehicles.

CONCLUSIONS

In this paper the allocation of buses to routes, one component of the short-range transit-planning problem, has been discussed. A model was proposed that treats the problem as a constrained resource-allocation problem. The objective was that net social benefit, which consists of ridership benefit and wait-time savings, be maximized subject to constraints on total subsidy, fleet size, and acceptable levels of loading. An algorithm was developed to solve the resulting mathematical program, which can be implemented on a computer or on a programmable calculator.

The case study of one garage of the MBTA system produced a number of important findings:

1. The best allocation of buses (and resources) is very robust with respect to the objectives and parameters assumed,
2. Existing rules of thumb used in the transit industry may not be as efficient as a formal model that uses a consistent objective, and
3. The proposed model can be useful in policy analysis, for example, in the development of service policies and vehicle procurement.

This study leaves a number of important topics for further research:

1. Better understanding of the objectives and constraints currently used by transit schedulers,
2. Relaxation of the assumption of route independence embodied in the model, and
3. Pilot implementation of these ideas in a transit agency.

More broadly, with the encouraging results obtained in this study, a new look at the role of more-formal methods for improving short-range transit planning seems badly needed.

ACKNOWLEDGMENT

The research reported in this paper was funded in part by a grant from the Urban Mass Transportation Administration of the U.S. Department of Transportation.

REFERENCES

1. P.G. Furth. Setting Frequencies on Bus Routes. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Res. Rept. CTS-RAMP-80-2, Feb. 1980.
2. J.P. Attanucci, L. Jaeger, and J. Becker. Bus Service Evaluation Procedures: A Review. UMTA, U.S. Department of Transportation, Rept. UMTA-MA-09-7001-79-1, April 1979.
3. R.F. Kirby and M.A. Green. Policies and Procedures for Transit Service Development. Traffic Quarterly, Vol. 33, No. 3, July 1979, pp. 413-428.
4. E.K. Morlok. Supply Functions for Public

Transport: Initial Concepts and Model. In Traffic Equilibrium Methods (M.A. Florian, ed.), Proc., International Symposium on Traffic Equilibrium Methods, Montreal, 1974, Springer Verlag, Berlin, 1976, pp. 322-367.
5. H. Mohring. Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, Sept. 1972.
6. C.R. Guinn. Transit Optimization--Test 1 Analysis; Transit Optimization--Test 2 Analysis. New York State Department of Transportation, Albany, Preliminary Res. Repts. 2 and 12, 1967.
7. S. Scheele. A Mathematical Programming Algorithm for Optimal Frequencies. Institute of Technology, Linköping Univ., Linköping, Sweden, Linköping Studies in Science and Technology, Dissertation 12, 1977.
8. M. Lampkin and P.D. Saalmans. The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study. Operations Research Quarterly, Vol. 18, No. 4, Dec. 1967, pp. 375-397.
9. L.A. Silman, Z. Barzily, and U. Passy. Planning the Route System for Urban Buses. Computers and Operations Research, Vol. 1, 1974, pp. 201-211.
10. A. Last and S.E. Leak. TRANSEPT: A Bus Model. Traffic Engineering and Control, Vol. 17, No. 1, Jan. 1976.
11. O. Hasselström. A Method for Optimization of Urban Bus Route Systems. Volvo Bus Corporation, Göteborg, Sweden, working paper, 1979.
12. C.A. Nash. Management Objectives, Fares, and Service Levels in Bus Transport. Journal of Transport Economics and Policy, Jan. 1978.
13. A.F. Han and N.H.M. Wilson. The Allocation of Buses in Heavily Utilized Networks with Overlapping Routes. Transportation Research (in preparation).
14. Ecosometrics, Inc. Patronage Impacts of Change in Transit Fares and Services. Office of Service and Methods Demonstrations, UMTA, U.S. Department of Transportation, Rept. 135-1, 1980.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Strategies for Improving Reliability of Bus Transit Service

MARK A. TURNQUIST

Four major classes of strategies for improving reliability of bus transit service are analyzed: vehicle-holding strategies, reduction of the number of stops made by each bus, signal preemption, and provision of exclusive right-of-way. The principal findings are that (a) strategies to improve service reliability can have very substantial impacts on overall service quality, including improvements in average wait and in-vehicle time as well, and (b) the best strategy to use in a particular situation depends on several factors, but service frequency is the most important. For low-frequency services (less than 10 buses per hour), schedule-based holding strategies or zone scheduling is likely to work best. For midfrequency services (10-30 buses per hour) zone scheduling or signal preemption is likely to be most effective, although headway-based holding can also work well if an appropriate control point can be found. In high-frequency situations (more than 30 buses per hour), an exclusive lane combined with signal preemption should be considered.

The concept of service reliability has come into increasing prominence in recent years as an important characteristic of the quality of service provided by transportation systems. A basic definition of reliability, as the term is used here, is the variability of a system performance measure over time. The focus is on stochastic variation in performance rather than on more-traditional engineering concepts of probability of component or system failure. The level-of-service measure most clearly subject to variation is travel time, and this variability is often described in terms of nonadherence to schedule.

Service reliability is important to both the transit user and the transit operator. To the user, nonadherence to schedule results in increased wait time, makes transferring more difficult, and causes uncertain arrival time at the destination. The importance of some measure of reliability to trip-making behavior has been emphasized in several attitudinal studies. For example, Paine and others (1) found that potential users ranked "arriving when planned" as the single most important service characteristic of a transit system. This finding has been substantiated in further studies by Golob and others (2) and by Wallin and Wright (3).

In addition to its importance to transit users, unreliability in operations is a source of reduced productivity and increased costs for transit operators. This is due to the need to build substantial slack time into timetables in order to absorb deviations from the schedule. This leads to reduced use of both equipment and personnel. The recent report by Abkowitz and others (4) provides an excellent summary of the major issues in transit-service reliability from the perspectives of both the user and the operator.

In light of the current need for more cost-effective public transportation in urban areas, it is important to understand the sources of unreliability and to investigate the potential of several alternative control strategies to improve both the quality of service provided and the productivity of the equipment and the personnel in the system.

The research on which this paper is based has had four major objectives:

1. Investigation of the sources of service-reliability problems in bus transit networks,
2. Identification of potential strategies for improving reliability of service,
3. Development of models to allow these strategies to be analyzed and evaluated, and
4. General evaluation of the relative effectiveness of these strategies.