TRANSPORTATION RESEARCH RECORD 854

Bus Services

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1982

Transportation Research Record 854
Price $11.80
Edited for TRB by Scott C. Herman

mode
2 public transit

subject areas
11 administration
12 planning
13 forecasting
14 finance
15 socioeconomics
16 user needs
53 vehicle characteristics
54 operations and traffic control
55 traffic flow, capacity, and measurements

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Bus services.

(Transportation research record; 854)
Reports for the 61st annual meeting of the Transportation Research Board, 1982, in Washington, D.C.
1. Bus lines–United States–Congresses. I. National Research Council (U.S.). Transportation Research Board. Meeting (61st: 1982: Washington, D.C.) II. Series.
TE7.H5 no. 854 [HE5623] 380.5s 82-18882
ISBN 0-309-03358-6 [388.4'1322'0973] ISSN 0361-1981

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Kenneth W. Heathington, University of Tennessee, chairman

Public Transportation Section
John J. Fruin, Port Authority of New York and New Jersey, chairman

Committee on Bus Transit Systems
James C. Echols, Tidewater Transportation District Commission, chairman
Robert B. Deuser, Beiswenger Hoch & Associates, Inc., secretary
John J. Bakker, John W. Bates, Eugene T. Canty, Avishai Ceder, Frank Dicesare, John Dockendorf, Edward R. Fleischman, Marvin C. Gersten, Jack R. Gilstrap, Andrew Hollander, Ronald I. Hollis, Galen C. Larson, Herbert S. Levinson, Neil Craig Miller, Subhash R. Mundle, Martin C. Nizlek, Howard J. Simkowitz, Robert N. Taube, F. V. Webster, Douglas L. Wentworth, Robert J. Zerrillo

Committee on Rural Public Transportation
Douglas J. McKelvey, Federal Highway Administration, chairman
Jon E. Burkhardt, Ecosometrics, Inc., secretary
Robert H. Bruton, Bernard F. Byrne, Ken E. Clary, Charles A. Davis, Ronald E. Gargasz, Robert T. Goble, Lawrence J. Harman, Joyce H. Johnson, Andrew D. Jones, Robert G. Knighton, H. Thomas McCormick, Ervin Poka, Jr., Arthur Politano, Peter M. Schauer, Lawrence L. Schulman, Loretta E. Sharpe, Bruce S. Siria, Robert L. Smith, Jr., Sheldeon G. Strickland, Gregory C. Wilder

Wm. Campbell Graeub, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1981.

Contents

Authors of the Papers in This Record

Adebisi, Olusegun, Department of Civil Engineering, Ahmadu Bello University, Zaria, Nigeria
Albright, Richard, Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142
Amoruso, Raymond, Madison Avenue Dual-Width Bus Lane Project, New York City Department of Transportation, 40 Worth Street, New York, NY 10013
Brand, Daniel, Charles River Associates, Inc., 200 Clarendon Street, Boston, MA 02116
Cherwony, Walter, Abrams-Cherwony & Associates, PSFS Building, Suite 3139, 12 South 12th Street, Philadelphia, PA 19107
Clark, James W., Consultant, 2415 East McGraw, Seattle, WA 98112
Collura, John, Department of Civil Engineering, University of Massachusetts, Amherst, MA 01003
Cope, Dale Ferguson, Department of Civil Engineering, University of Massachusetts, Amherst, MA 01003
Couture, Michael, Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142
Fuhs, Charles A., Metropolitan Transit Authority, P.O. Box 61429, Houston, TX 77208
Gleichman, Gregory R., Simpson and Curtin Division, Booz, Allen and Hamilton, Inc., 400 Market Street, Philadelphia, PA 19106
Hollander, Andrew, Office of Surface Transportation Planning, New York City Department of Transportation, 40 Worth Street, New York, NY 10013
Hurdle, V.F., Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4
Kocur, George, Thayer School of Engineering, Dartmouth College, Hanover, NH 03755
Lam, Tenny N., Department of Civil Engineering, University of California, Davis, CA 95616
Lanman III, Maurice H., AVM Systems, Inc., 5050 Mark IV Parkway, Fort Worth, TX 76106
Louie, Charles, Madison Avenue Dual-Width Bus Lane Project, New York City Department of Transportation, 40 Worth Street, New York, NY 10013
Lyles, Richard W., Richard Lyles Associates, 51 Blackstone Street, Bangor, ME 04401
Mandel, Michael, Charles River Associates, Inc., 200 Clarendon Street, Boston, MA 02116
Meyberg, Arnim H., School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, 14853
Mundle, Subhash R., Simpson and Curtin Division, Booz, Allen and Hamilton, Inc., 400 Market Street, Philadelphia, PA 19106
Nelson, Michael, Charles River Associates, Inc., 200 Clarendon Street, Boston, MA 02116
Noguchi, Tomoki, City of Bellevue Planning Department, P.O. Box 1768, Bellevue, WA 98009
Paules, Granville E., Urban Mass Transportation Administration, U.S. Department of Transportation, 400 Seventh Street, S.W., Washington, DC 20590
Porter, Benjamin D., Simpson and Curtin Division, Booz, Allen and Hamilton, Inc., 400 Market Street, Philadelphia, PA 19106
Ritchie, Stephen G., School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853
Schuler, Harry J., Programming Department, Riverside County Transportation Commission, Riverside, CA 92501
Schwartz, Samuel L., Bureau of Traffic Operations, New York City Department of Transportation, 40 Worth Street, New York, NY 10013
Turnquist, Mark A., School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853
Ulberg, Cy, Municipality of Metropolitan Seattle, 821 Second Avenue, Seattle, WA 98104
Ward, Donald, Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142
White, Robert A., GM Engineering Staff, General Motors Technical Center, Warren, MI 48090

Innovative Transit Service Planning Model That Uses a Microcomputer

MARK A. TURNQUIST, ARNIM H. MEYBURG, AND STEPHEN G. RITCHIE

Transit service planning is the process of designing appropriate services, including considerations of area coverage, integration with other transit services, and the frequency of service that can be justified economically as well as socially and politically. A simple and usable analytic model to guide management in the search for, and evaluation of, operating strategies that meet local transit service objectives is described. This analysis system is intended primarily for use on single routes or in transit corridors that include a small number of parallel or serial routes. The model system includes as basic components models of supply (system performance), demand (mode and path choice), cost, and evaluation-measure prediction. The supply-and-demand components are linked in an explicit equilibration structure to include the important interactions between transit system performance and passenger volume. Design options that can be explored with the model system include fare and headway changes, scheduling changes such as turn-backs, etc. Two major aspects of this model system are that (a) it is designed to make maximum use of readily available data and (b) it has been implemented on a microcomputer (an Apple III) in order to minimize the investment in computer resources.

Today operators of urban transit systems are in the difficult position of facing demands for improved and more widespread services on the one hand and increasingly stringent financial constraints on the other. This requires the operator to design new services very carefully and to look critically at existing services on a route-by-route basis to determine whether the level of service (LOS) being provided is economically justifiable. Whether designing new services or examining old ones, it is vital that the operator have appropriate performance measures and analytical tools to aid the operations-planning activities.

The objective of the research reported here is to develop such tools. The analysis system that results from this effort is intended primarily for use on single routes or in transit corridors that include a small number of parallel or serial routes. It is important to emphasize that the analysis system is built around interacting supply-and-demand models. Existing supply models, which represent the impacts of routing and scheduling decisions, are not sufficient in themselves because they assume fixed travel demands, while many service changes are expected to create substantial demand changes. Existing demand models reflect the impacts of service changes but generally treat LOS measures in a way that does not support effective operations planning and management. In addition, these models usually ignore relevant service measures such as on-time reliability. Linking improved supply-and-demand models together in an equilibrium structure and rationalizing their level of detail and sensitivity enables the manager or analyst to obtain policy-relevant predictions of the impacts of alternative service strategies.

A major concern of this research is to ensure that the models developed can be used easily and effectively by transit operators. This has resulted in two major decisions regarding model structure. First, the models are implemented on a microcomputer (an Apple III) in order to minimize the investment in computer resources necessary for an operator to use the models. This is a major departure from other operations-planning models such as IGTDS (1) and TNOP (2), which require large mainframe computers and sophisticated interactive graphics hardware. Although such models are powerful, they are complex in structure and require expertise not readily available to many transit operators.

The second major aspect of this model system is that it is designed to operate primarily with data normally available at most transit properties. This minimizes the data-collection costs imposed on an operator to use the system. Although there is no substitute for good data if we wish to make precise predictions of the effects of various operational changes, it is important to provide options to the operator or analyst to allow rough estimates to be obtained with little detailed data. This improves the responsiveness of the model system because it reduces the cost of specifying new alternatives to be analyzed.

OVERVIEW OF MODEL SYSTEM

The analysis system includes basic component models of supply (system performance), demand (mode and path choice), costs, and evaluation-measure prediction. The supply-and-demand components are linked in an explicit equilibration structure to include the important interactions between system performance and passenger volume. The overall organization of the model system is illustrated in Figure 1.

The service specification determines the operating cost of the system and initial values for four basic performance measures. These performance measures both influence, and are influenced by, ridership. When equilibrium values of the performance measures and ridership have been achieved, the basic evaluation measures can be predicted. These basic evaluation measures can be used simply as a list of individual measures or combined in various ways. For example, several partial cost-effectiveness measures such as operating cost per trip, operating deficit per trip, passengers per vehicle mile, etc., may be formed.

Based on these evaluation measures, the service specification can be revised if necessary and the analysis redone with the new values. Iterations of this sort can be repeated until the user (transit operator or planner) is satisfied with the evaluation measure values.

The next section of this paper provides a summary of the component models. Additional technical details on the models are contained in Turnquist and others (3,4). The other sections describe data requirements of the model system and conclusions.

COMPONENT MODELS

Supply Model

The basic structure of the supply-model system is illustrated in Figure 2. The inputs are environmental characteristics (speed limits and numbers of signalized intersections along route segments), the service specification (route length, service frequency, stop spacing, and fare), and ridership (from the demand model). Outputs are an LOS vector (including fare, in-vehicle time, access or egress time, wait time, and transfer time) to be input to the demand model, as well as estimates of vehicle hours and vehicle miles for cost estimation.

Figure 1. Structure of model.

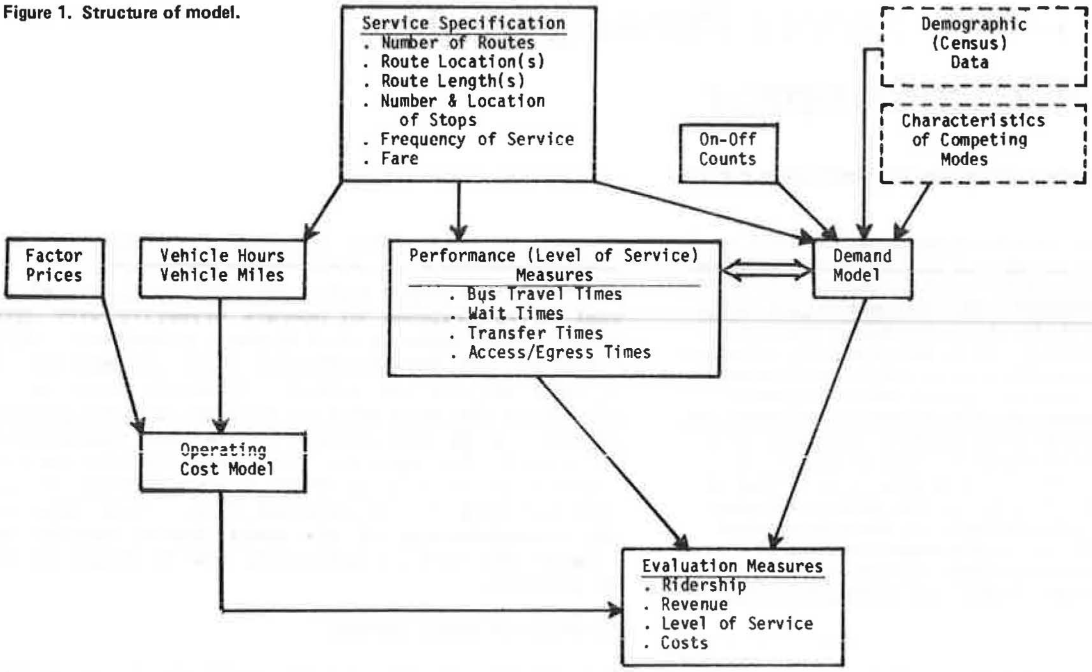
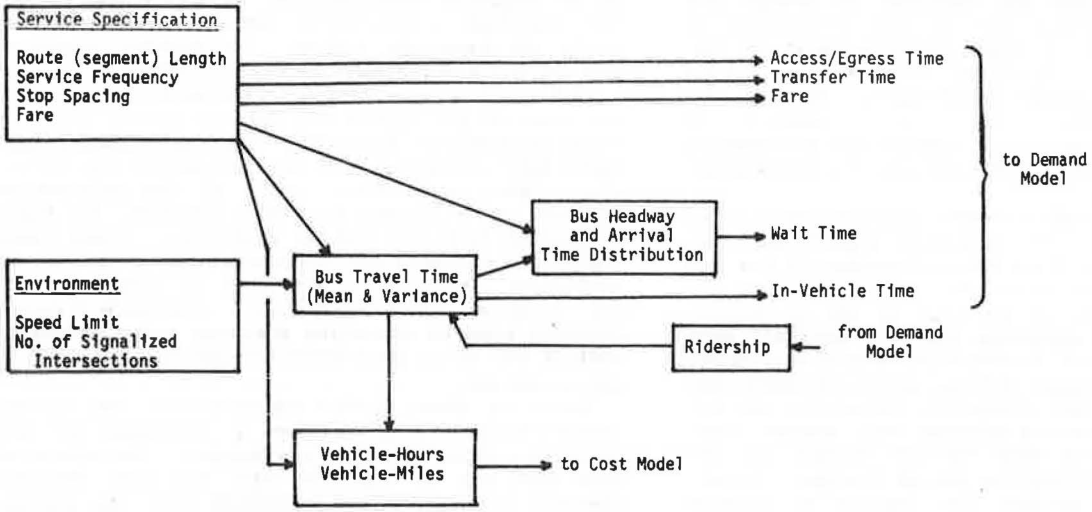


Figure 2. Overall structure of supply model.



Predictions of access or egress time depend primarily on stop spacing and the assumed maximum access distance. For most services, the access or egress mode will be walk. However, for some express services the predominant access or egress mode may be automobile (park-and-ride or kiss-and-ride). In this case the access or egress time may be substantially different.

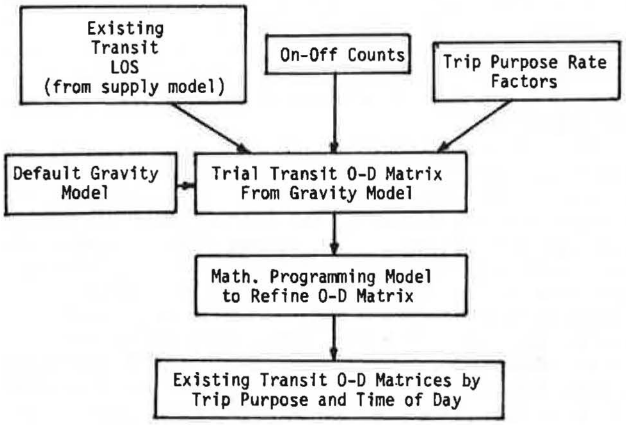
Maximum walking distance will typically be about 0.25 mile (400 m). [As an example, see the analysis done by Bakker (5).] This value was used as the maximum walking distance in our analysis. Together with an assumed average walking speed of 3 mph (4.8 km/h) and the stop spacing of the transit routes, this determines average access or egress time as an element of LOS. If desired, the maximum walking distance can be changed quite readily.

Prediction of mean bus travel times along route segments is important for average predictions of in-vehicle time. Also, prediction of the variability of bus travel times is necessary for use in the wait-time model. Our approach to travel-time prediction is based on the previous work of Turnquist and Bowman (6) and Jordan and Turnquist (7), although there are some minor modifications.

The mean travel time on a route segment is estimated as a function of segment length, speed limit, number of signalized intersections, average total boardings and alightings per bus, and average number of stops made to serve passengers. This functional specification is the result of analyses that use data from Cincinnati, Ohio. Note that the effect of ridership on LOS is clearly evident here, as mean time is affected by the number of passengers served and the number of stops made.

Once the mean travel time is estimated, the standard deviation is estimated simply by assuming that the distribution of travel times has a constant coefficient of variation. The standard deviation of

Figure 3. Model for estimating existing transit ridership.



travel time is an important input to the waiting-time model.

The model used to predict wait time is based on the model described by Bowman and Turnquist (8). This model predicts an arrival pattern of passengers at a stop by using information on expected headways between buses and the variability of arrival times (with respect to schedule) of buses at the stop. Thus, it includes the effects of both mean headway and service reliability on passenger wait time. It can be used over a wide range of headways; for very short headways it produces essentially the same results as simpler random-arrival models of wait time but matches observed waiting times much better than the random-arrival models for less-frequent services.

Finally, an expression for the expected transfer time when passengers board a first-leg vehicle without regard for the schedule of services on the second leg has been derived by Sen and Morlok (9). They term this the non-selective-arrivals case, and it is the most appropriate case for use in transit applications.

The total LOS vector, which includes access and egress time, wait time, transfer time, in-vehicle time, and fare, is then available for input to the demand models. Access and egress time, wait time, and transfer time can also be aggregated into a single variable (out-of-vehicle time) if desired.

Estimation of Operating Costs

The basic method for cost estimation in the model system relies on unit costs for vehicle hours and vehicle miles. The system-performance models produce estimates of vehicle miles and vehicle hours that result from the service change being tested. Vehicle ownership, maintenance, and fuel expenses are based on vehicle miles by using information on both physical relations (e.g., fuel consumption of buses) and current prices (fuel price, interest rate, maintenance labor wage rate, etc.). Vehicle hours are used to estimate driver and superintendence labor costs. The unit cost reflects wage rates and labor productivity factors, which may vary from peak to off-peak periods.

Total costs are computed as the sum of labor, vehicle ownership and maintenance, and fuel costs. Operating costs are given by total costs less the vehicle ownership (capital) cost. This procedure is intended to estimate the short-run incremental costs of service changes and thus does not include overhead costs that reflect general administration, insurance, etc., which would not be expected to change.

A second option on cost estimation is to base the analysis on more detailed timetable information rather than simply on a rough approximation of total vehicle hours and vehicle miles. If a complete timetable of services is provided for all routes under analysis, the number of vehicles required can be computed by using the concurrent scheduler described by Bodin and Dial (10). This algorithm also produces an accurate estimate of vehicle miles, including deadheading. A better estimate of labor costs can also be developed by using another method described by Bodin and Dial (10). This method is based on the out-of-kilter flow algorithm (11) and gives an accurate estimate of the number of drivers required to provide a given service.

The concurrent scheduler produces estimates of the number of vehicles required and the vehicle miles operated. This gives an accurate basis for estimating vehicle ownership, maintenance, and fuel costs. The out-of-kilter algorithm estimates the number of drivers, which is the basis for labor-cost estimation. This cost-estimation model is more complicated than the basic method and requires more data, but it gives more reliable estimates. This option is currently being implemented in the model.

Demand Model

The demand-model components are used in two ways. First, a procedure is provided to estimate base-case ridership by origin-destination (O-D) pair and trip purpose for the existing service. Second, an incremental analysis method is used to predict changes from that base-case ridership that results from changes in the service provided.

The establishment of existing ridership patterns entails derivation of transit O-D matrices by time of day and trip purpose. If the transit operator does not have O-D data available already and does not wish to collect such data through on-board surveys, the basic approach is to use transit on-off counts to obtain the required O-D matrices. As indicated in Figure 3, this involves calculation of a trial O-D matrix by using a gravity model; this matrix is then refined by a mathematical optimization model to yield a matrix that replicates the observed transit link loadings during the time period of interest.

Such on-off counts also provide actual values for two important independent variables in the bus travel-time prediction function that were mentioned earlier in the discussion of the supply model. These variables are the average number of total boardings and alightings per bus on each segment and the average number of stops made to serve passengers along each segment.

If desired by the user of the model system, the construction of the trial O-D matrix can be done by using multiple trip purposes. The total productions and attractions at each zone are split by trip purpose by using the percentages given by Sosslau and others (12), and the gravity model uses impedance coefficients that are trip-purpose specific. The resulting trial O-D matrices are then combined to form a total trip matrix before being input to the second-stage optimization model.

The basic approach to predicting ridership response to changes in LOS involves application of incremental demand techniques. These techniques pivot on the existing ridership levels and LOS to derive new ridership estimates based on the new LOS. The final ridership estimates and LOS are the result of equilibration between transit supply and demand.

Two incremental demand-modeling methods are employed. They implicitly assume that base ridership for existing transit service is accurately predicted and emphasize the use of relatively simple analytical techniques. The methods differ only in terms of the sophistication and underlying rationale of the incremental models used to predict the impact of service changes. The first method employs travel-demand elasticities, and the second method uses both full and incremental forms of the multinomial logit (MNL) model.

Elasticity analysis is the simpler of the two methods to apply. It avoids calculation of total corridor travel volumes and market shares and focuses only on the existing transit trip interchanges and existing and proposed LOS. Explicit treatment of competing modes can therefore be ignored under the assumption that transit service changes have no effect on the LOS of other modes.

The extreme simplicity of the elasticity method is both an advantage and a limitation. Note that not only are there various definitions of elasticities that affect their value, but elasticities are dependent on both the position and slope of the demand curve and are strictly appropriate to one point on the curve. Thus, the method should only be applied for relatively small changes in LOS variables. Also, it might be argued that elasticities are unlikely to be transferable over time and across locations and socioeconomic groups. Nevertheless, some consistency has emerged in empirically estimated elasticity values, and elasticity methods continue to be useful tools of analysis.

The second incremental demand method employs both full and incremental forms of the MNL model. This method therefore requires explicit treatment of modes that compete with transit and calculation of their market shares, but it is theoretically more satisfying than the elasticity method. Changes in the impedances of competing modes also may be incorporated with the logit analysis, if desired.

The MNL model used as a default model in the analysis system was developed for the Twin Cities area of Minneapolis and St. Paul (13). It is a three-mode model (transit, automobile drive alone, and automobile group ride). Three trip purposes are considered (home-based work, home-based nonwork, and nonhome based) and separate equations are used for each. If the user desires, an alternate model can be substituted.

The calibration results for the Twin Cities indicate that this model incorporates LOS and socioeconomic variables relevant for transit route-demand analysis. The data for these variables are also accessible to the operator. The estimated parameters are consistent with other comparable MNL model calibrations, which is encouraging for transferability of the models. In addition, the transit disutility equation is sensitive to automobile access to the transit system, which is a useful feature if park-and-ride lots are to be analyzed. The home-based-work model also distinguishes between transit wait times for the first vehicle and time spent transferring to subsequent vehicles.

Evaluation Measures

The key issue in the selection of a set of measures by which a transit design alternative may be evaluated is to choose the smallest set that still includes all the desired information. It is important to keep the set as small as possible so that the analyst can understand and retain the important information about each alternative. On the other hand, it is vital that the set of evaluation measures be rich enough to accommodate evaluation from several perspectives. The previous work of Fielding and others (14) on transit service quality and performance measures is of particular value in this regard, since it has highlighted the importance of including both measures of efficiency (i.e., doing things right), which are of primary interest to the operator, and measures of effectiveness (i.e., doing the right things), which are of primary interest to the community.

After computation of a new solution, the user of the model system is offered the output options shown in Figure 4. Each of the four summaries provides values for the existing service as well as values for the new design. This allows direct comparison for each measure. The ridership summary includes information on total trips, passenger miles, and passengers per vehicle mile. The LOS summary includes average in-vehicle time, wait time, transfer time, total passenger hours, average O-D passenger speed, and average passengers per seat as a measure of crowding. The operating-statistics summary includes vehicle hours, vehicle miles, seat miles, average vehicle speed, and average load factor. The financial summary includes revenue and cost information. Figure 5 shows an example of the financial summary to illustrate the information provided and the general format of all the summary outputs.

For a more detailed examination of route performance, options are provided for a route-load profile and detailed output of O-D ridership. Figure 6 shows an example of a route-load profile as it appears on the monitor screen. This type of graphical output conveys a great deal of information very quickly and can be very useful for consideration of detailed schedule changes on a route.

It should be emphasized that the exact form of the output reports and the measures reported can be modified easily to meet the specific needs of a particular operator. Finally, note that there is a certain level of flexibility inherent in most of the evaluation measures proposed here. For example, the time reference frame can be chosen by the evaluator from among peak period, nonpeak period, hour, day, week, etc.

DATA REQUIREMENTS FOR USE OF MODEL

The basic philosophy of the model system is not only to allow meaningful analyses to be carried out with a minimum of data but also to allow more detailed and precise information to be used if it is available. This is accomplished by having many default parameters in the model. These default values are realistic numbers based on a variety of studies but should be overridden whenever possible by values based on detailed local information.

The minimum data requirements for analysis of a route (or routes) are as follows:

1. On-off counts by stop,
2. Fare,
3. Length of each route segment (in kilometers),
4. Number of stops (by segment),
5. Number of signalized intersections (by segment),
6. Speed limit (by segment),
7. Capital cost per vehicle kilometer,
8. Operating cost per vehicle kilometer, and
9. Operating cost per vehicle hour.

On-off counts are the basis for construction of the existing ridership by origin and destination. The counts themselves do not provide the information to connect origins and destinations for individual riders, but the procedure that described the demand model solves this problem. This method produces an O-D trip table that is consistent with the observed on-off counts and passenger volumes on vehicles. By

Figure 4. Menu of output options.

```
OUTPUT OF RESULTS
!!!!!!!!!!!!!!!!!

OPTIONS:

   1 NO OUTPUT - RETURN TO MAIN MENU
   2 RIDERSHIP SUMMARY
   3 LEVEL OF SERVICE SUMMARY
   4 OPERATING STATISTICS SUMMARY
   5 FINANCIAL SUMMARY
   6 ROUTE LOAD PROFILE
   7 ORIGIN-DESTINATION RIDERSHIP

PLEASE ENTER DESIRED OPTION NUMBER: █
```

Figure 5. Example of financial summary output.

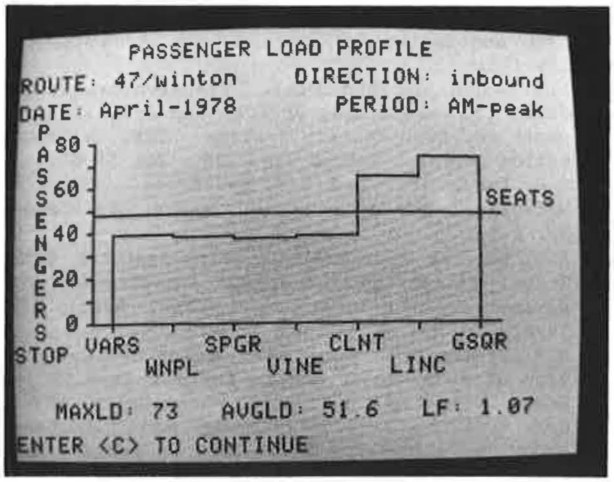
```
47/winton     April-1978    AM-peak
-----------------------------------

              FINANCIAL SUMMARY

        MEASURE               NEW VALUE   EXISTING

FARE                          $   0.50    [illegible]
REVENUE                       $ [illegible] [illegible]
 - INBOUND REVENUE            $ [illegible] [illegible]
 - OUTBOUND REVENUE           $ [illegible] [illegible]
REVENUE/PASSENGER-KM          $ [illegible] [illegible]

ROUTE COST                    $ [illegible] [illegible]
CAPITAL COST                  $ [illegible] [illegible]
OPNG & MAINT COST             $ [illegible] [illegible]
O&M COST/PASSENGER            $ [illegible] [illegible]
TOTAL COST/PASSENGER          $ [illegible] [illegible]

REVENUE/O&M COST                  1.87      1.71

ENTER (C) TO CONTINUE OR (P) TO PRINT HARDCOPY █
```

incorporating this method in the model, the transit operator is relieved of the need to supply a complete O-D table. This procedure is normally performed just once to establish existing (base-case) ridership patterns as a basis for predicting changes by using the incremental demand models as discussed earlier.

The route-segment information is used in the estimation of vehicle running times (mean and variance). These estimates, in turn, are used in computation of in-vehicle travel times, wait times, and operating costs.

Finally, unit cost data are the minimum information required for estimating changes in capital and operating costs that might result from various service changes. If the incremental logit model is used to predict ridership changes, additional information on automobile LOS and income data for the population must be provided. These data are not necessary if the simple elasticity model is used for predicting ridership changes.

Figure 6. Example of route-load profile.



```
    PASSENGER LOAD PROFILE
ROUTE: 47/winton     DIRECTION: inbound
DATE: April-1978        PERIOD: AM-peak

  MAXLD: 73    AVGLD: 51.6    LF: 1.07
ENTER <C> TO CONTINUE
```

CONCLUSIONS

This paper has described a simple and usable analytic model structure designed to provide the transit operator with a capability to search for and evaluate operating strategies that meet local transit service objectives. It has been shown that an effective supply-and-demand equilibration structure can be developed that is based on very limited and readily available data. Further, this model has been implemented on a relatively low-cost microcomputer.

The user can explore readily a number of changes in route-level operating policies. The cost model and evaluation measures allow the operator to assess the consequences of such operating policies. Thus, the model system represents an important new addition to transit operations-planning capability. Ongoing extension and development efforts will further enhance the attractiveness of the model system as a useful tool for transit operators.

Preparations are being made currently for field-testing of the model in two cities. These tests should begin in the summer of 1982, and results should be available by mid-1983.

ACKNOWLEDGMENT

This research was supported by the Office of University Research, U.S. Department of Transportation.

REFERENCES

1. Interactive Graphic Transit Design System (IGTDS): User's Manual for GM TSD Release No. 2. General Motors Transportation Systems Division, General Motors Technical Center, Warran, MI, Sept. 1978.
2. Interactive Graphic Transit Network Optimization System (TNOP): User's Manual for GM TSD Release No. 1. General Motors Transportation Systems Center, General Motors Technical Center, Warren, MI, Aug. 1980.
3. M.A. Turnquist, A.H. Meyburg, and S.G. Ritchie. An Equilibrium Model for Short-Range Transit Planning. School of Civil and Environmental Engineering, Cornell Univ., Ithaca, NY, Working Paper 8303-01, July 1981.
4. M.A. Turnquist, A.H. Meyburg, and S.G. Ritchie. Short-Range Operations Planning Models for Transit Systems. U.S. Department of

9. A.K. Sen and E.K. Morlok. On Estimating Effective Frequencies and Average Waiting Times for Indirect Connections. Transportation Planning and Technology, Vol. 3, No. 3, 1976, pp. 175-184.
10. L.D. Bodin and R.B. Dial. Hierarchical Procedures for Determining Vehicle and Crew Requirements for Mass Transit Systems. TRB, Transportation Research Record 746, 1980, pp. 58-64.
11. L.R. Ford, Jr., and D.R. Fulkerson. Flows in Networks. Princeton Univ. Press, Princeton, NJ, 1962.
12. A.B. Sosslau and others. Quick-Response Urban Travel Estimation Techniques and Transferable Parameters, User's Guide. NCHRP, Rept. 187, 1978.
13. R.H. Pratt; DTM, Inc. Development and Calibration of Mode Choice Models for the Twin Cities Area. Metropolitan Council, Minneapolis, 1976.
14. G.J. Fielding, R.E. Glauthier, and C.A. Lave. Applying Performance Indicators in Transit Management. Presented at the National Conference on Transit Performance, Norfolk, VA, Sept. 1977.

Transportation, Final Rept., in preparation.
5. J.J. Bakker. Transit Operating Strategies and Levels of Service. TRB, Transportation Research Record 606, 1976, pp. 1-5.
6. M.A. Turnquist and L.A. Bowman. Control of Service Reliability in Urban Networks. U.S. Department of Transportation, Final Rept. DOT-OS-80018, 1979.
7. W.C. Jordan and M.A. Turnquist. Zone Scheduling of Bus Routes to Improve Service Reliability. Transportation Science, Vol. 13, No. 3, 1979, pp. 242-268.
8. L.A. Bowman and M.A. Turnquist. Service Frequency, Schedule Reliability, and Passenger Wait Time at Transit Stops. Transportation Research, Vol. 15A, No. 6, 1981, pp. 465-472.

Publication of this paper sponsored by Committee on Bus Transit Systems.

State of the Art of Current Bus Transfer Practices

MICHAEL NELSON, DANIEL BRAND, AND MICHAEL MANDEL

The use and consequences of alternative bus transfer policies are examined. A bus transfer policy consists of a set of operator actions that involves vehicle routing and scheduling, transfer charges, information for passengers, and terminal facilities that affects the movement of passengers between buses as part of a continuing trip. In this paper, the bus transfer policies currently in use on U.S. transit properties are described and summarized. Reasons why properties use or do not use particular transfer policies are identified, and the specific consequences of alternative transfer policies in different settings on cost, ridership, revenue, and user satisfaction are assessed. Situations or settings in which particular transfer policies can be applied beneficially are then identified and analyzed.

Abstract goes here

Under ideal circumstances, transit would carry all users directly from their orgins to their destinations without requiring a change of vehicles. However, given the geographic and temporal distribution of trips, such direct service is of course uneconomical for transit to provide. Therefore, operators must undertake some set of actions that involves such factors as vehicle routing, scheduling, transfer charges, and/or information for passengers (a transfer policy) to accommodate transferring riders.

This paper examines the use and impacts of the following 11 bus transfer-policy components, which are listed under four main components:

1. Routing components--distance between routes at transfer points and through-routing;
2. Scheduling components--schedule coordination, dynamic control of departure times at transfer points, timed transfers, schedule adherence on connecting routes, and service frequency on connecting routes;
3. Pricing components--transfer charge and use of transfer slips; and
4. Information components--provision of schedule information and marketing initiatives.

Note that the 11 transfer-policy components examined here do not exhaust the list of possible operator actions that affect transfers. However, most of the remaining ones (such as transit shelters, terminal facilities, and temporal or directional restrictions) are reviewed at least briefly in conjunction with one or more of the above components.

The material presented in this paper is drawn from the results of a recently completed study conducted under the Service and Methods Demonstration program of the Urban Mass Transportation Administration (UMTA) for the Transportation Systems Center (1,2). Data for that study were drawn primarily from a series of telephone and on-site discussions with experienced transit professionals on 39 different properties.

On any particular transit property, the demand for transferring clearly influences the type of transfer policy adopted. Relevant transfer demand characteristics include the following:

1. The percentage of riders who transfer (i.e., transfer rate),
2. Their socioeconomic and trip purpose characteristics,
3. Transfer-point locations, and
4. Directional and temporal characteristics.

The transfer rate is the percentage of transit person trips that involve transfers between transit vehicles. Often, the transfer rate cannot be calculated directly from available data but rather must be estimated from transfer slip data, passenger counts, or special surveys. Data problems include transit pass users who do not use transfer slips or riders who transfer more than once in the course of a trip. In general, however, it is possible to obtain reasonable estimates of transfer rates on most properties.

For bus-to-bus transfers, the average transfer rate on the properties examined is approximately 21 percent. However, several bus properties have a transfer rate on the order of 5 percent, while transfer rates as high as 50 percent have been observed. The size of the property has a large effect on the overall transfer rate. Not including properties that currently use timed transfers extensively, large bus properties have a much higher average transfer rate than small properties (20 versus 12 percent). Bus properties that currently use timed transfers extensively are uniformly small properties and have a much higher transfer rate (28 percent) than either large or small non-timed-transfer cities. Also, bus properties that do not charge for transfers have a higher average transfer rate than those that do charge (22 versus 18 percent). In many of these cases, the causal relation is not clear. That is, rather than the action causing a higher transfer rate, the presence of a high transfer rate may cause a property to institute options such as timed transfers or no transfer charge.

Riders who transfer vary by socioeconomic and demographic groups. Low-income riders transfer more often than higher-income riders, and young people also have above average transfer rates. Elderly people, on the other hand, tend to have a lower transfer rate than other riders, perhaps because of the effort involved in changing vehicles.

In the following sections, the current practices of transit operators regarding each of the 11 transfer-policy components listed above are described. Reasons offered by operators for the use or lack of use of the policy component are examined, and the cost, user satisfaction, ridership, and revenue consequences of the component when used in different settings are provided. In this way, the types of properties and settings for which the transfer-policy component is most beneficial are identified, along with other components that work well when combined with the component in question.

TRANSFER-POLICY OPTIONS

Distance between Routes at Transfer Points

A basic attribute of transferring is the walk required between vehicles. There may be only a few feet or, alternatively, passengers may have to walk several blocks to transfer. The greater the distance, the less useful the transfer is for the passenger. A significant number of bus properties separate by 500 ft or more some routes between which transfers are expected to occur. On at least one property, passengers must walk up to 1500 ft to transfer. In contrast, an ideal transfer arrangement is one where the transfer walk distance is less than one or two blocks, with clear lines of sight between buses.

There are two major factors that determine how closely a given transfer point can approach this ideal situation--the number of vehicles that meet at the transfer point and the size of the central business district (CBD). Available information strongly suggests that the upper bound on the number of buses than can be present simultaneously in the area that surrounds a single bus transfer point is approximately 20. Above this number, even if all buses formally meet in the same area, there will necessarily be a significant transfer walk distance and obstructed lines of sight between at least some pairs of routes.

The size of the CBD is important because line-haul transit often serves as a downtown distributor if the CBD is large enough so that no single terminal area is within walking distance of all of it. Concentrating the termination points of all routes in one spot in such a CBD may cut down overall coverage and greatly increase transferring unless costly detours are made. Therefore, except where the layout of the CBD is well suited to single-point termination, transit systems with large CBDs cannot generally obtain the most desirable walking distance between all vehicles at a single transfer point.

It is therefore often necessary to consider alternatives to the most desirable transfer arrangement. Specific alternatives available to the operator that relate to spatial separation at transfer points include the following:

1. Building an off-street terminal facility. A terminal facility increases the number of buses that can meet at one point and reduces pedestrian obstacles. However, it may be necessary to use a non-central location, and there may be significant capital costs. This option is best implemented when there are non-transfer-related benefits as well (e.g., reduction of street congestion).
2. Establishing a bus transit mall or street. With a bus transit mall or street, the number of routes is not a constraint and transfer walk distance is effectively very low, since vehicles from different routes pass the same points. Depending on the shape of the CBD, transit malls may reduce CBD coverage and require more walking and more transfers. This option is most feasible when the CBD is narrow (e.g., four blocks), and can often be implemented without significant capital costs.
3. Collecting route termini into several subfoci such that all or most routes intersect, although not all at the same point. With the option of collecting route termini, the overall number of routes is not a limitation, transfers between routes within subfoci are easy, and the CBD is well covered. Operating costs increase due to the extra vehicle miles of travel (VMT) in the congested downtown area. These costs tend to restrict the use of subfoci in larger cities.
4. Laying out routes in a grid network. With a grid system, the distance between routes at transfer points is minimized. Unfortunately, dispersed transfer points may be less understandable and less safe in the evening than single or multiple route foci. This option is most often employed in cities with large central areas of high-density employment and population, where a grid system may be the only appropriate route structure.

The principal cost consequences of any of these strategies typically arise from the changes in bus VMT needed to move the routes closer together. On the demand side, key aspects of transfer distance are walk time, comprehensibility of the transfer system, and potential pedestrian obstacles. These factors are particularly important to the elderly, shoppers, and infrequent users.

One additional important factor to consider when examining the trade-offs involved in reducing spatial separation is that some transfer-policy options discussed below, such a through-routing, schedule coordination, and timed transfers, require or are greatly facilitated by the physical proximity of connecting vehicles. In these cases, the trade-offs connected with spatial separation cannot be treated independently of the trade-offs connected with consideration of the other options. Thus, although reducing spatial separation has its costs, it may also have benefits that go beyond those of the single transfer-policy component standing alone.

Through-Routing

Through-routing, also known as interlining, involves linking two routes so that the same vehicle travels on both routes. It eliminates transfers between the two routes, since a passenger can board a vehicle at a stop on one route and get off at a stop on the other without having to change vehicles. A number of U.S. transit properties use through-routing; some properties use it quite extensively.

Five types of bus through-routing are currently in use:

1. Interlining, or classic through-routing, is when two separately identified routes share the same vehicles;
2. Single route through-routing differs from classic through-routing only in that the two halves of the route are joined on a permanent basis and are formally treated as a single route;
3. Variable through-routing differs from classic through-routing in that buses are exchanged among multiple routes rather than just between pairs of routes;
4. Trippers are when buses are through-routed at particular times of the day, usually during the rush hour or to meet shift or school times; and
5. Overlap involves terminating a radial route on the opposite side of the CBD from which it came in.

Through-routing can be used for two distinctly different reasons: operations and ridership. Both types of through-routing are considered in detail below.

Through-Routing for Operations

Through-routing can produce significant cost savings through elimination of turnaround time and distance, opportunities for logical scheduling, and potential gains in service reliability (if layovers are preserved). Although headway matching may add costs and extra scheduling effort may be needed, the net effect of through-routing is generally to reduce costs. These cost savings are most likely to occur in cities with a congested CBD where routes enter from more than one direction. Through-routing is most applicable as an aid to logical scheduling when properties are constrained by service-area boundaries or when operators seek to maintain clock-face or pulse scheduling. The presence of clock-face or pulse scheduling will also tend to minimize the need for further headway matching to implement through-routing, thereby avoiding potentially adverse cost impacts. Nevertheless, it must be emphasized that the operational and cost consequences of through-routing are heavily dependent on the street layout and other conditions. For instance, on some properties the dominant reason for implementing through-routing might be the elimination of dangerous left turns.

Through-Routing for Ridership

Through-routing eliminates transfers, which thus eliminates waiting and walking time and produces significant benefits for riders. Through-routing for ridership is often profitably employed where there is a high volume of transferring passengers between two routes with a common terminus. For instance, properties with strong and definite flows to outlying shopping malls may want to interline the mall route with a route running through a densely populated residential area. The groups that tend to benefit from this would be shoppers and the elderly--groups whose user satisfaction is most increased by through-routing. Properties with periodic peak flows to particular points, on the other hand, might profitably run trippers. If there is a relatively dispersed flow of transferring passengers, variable through-routing is a possible option. This will principally benefit the elderly or others who are made aware of and can afford to wait until a particular time of day for service. Properties that have a large amount of transferring to reach destinations within the CBD may consider overlap.

A reasonable range for the increase in ridership resulting from the connection of a pair of routes that serve logical origins and destinations is between 4 and 7 percent of the original ridership on the two routes. Conversely, pairing routes that do not connect logical origins and destinations may not increase ridership at all. Overall, through-routing for ridership is not necessarily incompatible with through-routing for operational reasons, although passengers who transfer between the two routes may be only a small portion of the total ridership on the routes. However, route pairings for maximum user satisfaction may not be the same as route pairings for maximum cost savings and operations benefit.

Schedule Coordination

Schedule coordination involves the adjustment of schedules on routes to change the relative times of arrival of vehicles at transfer points to reduce average transfer wait time. Schedule coordination used alone benefits passengers who transfer in one direction more than passengers who transfer in the other, since to ensure that one vehicle arrives before another without disruption of the regular schedule, an offset must be used (rather than attempting to have two vehicles meet at exactly the same time). Hence, schedule coordination used alone is applied most beneficially to route pairs where the majority of transfers are in a single direction at any one time.

Schedule coordination generally takes one of three forms:

1. CBD schedule coordination is used in situations where there is a strong directional flow of transfers through the CBD during peak hours. This option can improve the level of service for transferring passengers, although it requires some scheduling effort and sometimes changes in headways.
2. Trunk-crosstown coordination involves transfers between trunk lines and crosstowns where the schedule coordination is imposed outbound on the low-frequency crosstown lines. This option generally costs little and has minor negative effects on people who transfer in the opposite direction due to the generally high level of trunk-line frequencies. Trunk-crosstown coordination is more widely applicable than CBD schedule coordination, since its effects are less sensitive to the directionality of the transfer flow.
3. Minor schedule coordination characteristically is implemented in response to the complaints of passengers on a particular run who are unable to make a connection to another route. This option is easy to implement, although it typically does not lead to large ridership gains. It can be implemented on any transit system, even the largest and most complex, at any time of day.

If transfers are strongly directional between two lines at any time of the day and if a reasonable degree of schedule reliability exists, schedule coordination may be a very productive action for the operator to undertake. It typically costs little, can involve only minor headway changes, and demands almost no real-time operator attention. Because user satisfaction for the (assumed) large proportion of people who transfer in the correct direction is increased as their average transfer time is decreased, ridership will be induced in most cases. (Overall ridership gains are likely to be on the order of 3-4 percent for CBD schedule coordination, 1-2 percent for trunk-crosstown coordination, and minimal for minor schedule coordination.) Schedule coordination therefore may be a very cost-effective way to improve service.

However, there are definite limitations on the opportunities for application of most types of schedule coordination. The major restriction is the need for strong directionality of transfers. People who transfer in the "wrong" direction will have a transfer wait time equal to the entire headway (minus the advance) of their connecting bus. From the point of view of ridership, equity, and public relations, this may be unacceptable if a sizable number of people are affected. The result is that schedule coordination is inapplicable in many situations. For instance, it is largely inappropriate for off-peak use, since shopping traffic is inherently two-way. More important, it cannot be used in the CBD unless there is a skewed distribution of origins and destinations by time of day. Because this is a condition that is much more likely to occur in small cities than large ones, city size is a determinant of the applicability of at least CBD schedule coordination.

Dynamic Control of Departure Times at Transfer Points

Dynamic control involves holding a vehicle beyond its scheduled departure time from a transfer point if it is known that a vehicle on another route is approaching that is likely to have transferring passengers on board. Such information can be conveyed by radio or by some other signaling device (e.g., headlights). Dynamic control of bus transfers is found in many different settings. Several small properties use it extensively, either to control meetings between trunk and crosstown routes or to facilitate transfers in CBDs where the schedule permits. Some larger properties use dynamic control marginally, on only a few routes or only in the evening. However, on smaller properties that use timed transfers, dynamic control is regularly used to ensure the meeting of buses at the transfer point.

By definition, dynamic control perturbs the schedule. On a simple system this disturbance may not have widespread effects. On a more complex network of routes, use of extensive dynamic control may produce harmful schedule disruptions. There is also a limit on the number of dynamic-control messages that a radio system can handle. The major constraints on the use of dynamic control thus tend to be the size and complexity of the system.

This does not mean that larger properties cannot use dynamic control. It does mean, however, that it should be used sparingly and substituted for as appropriate, particularly on larger properties. For example, if dynamic control is used regularly at a particular transfer point, then an option such as schedule coordination might be more appropriate than a regular real-time adjustment in operations.

There are many situations where dynamic control is a low-cost method for obtaining large gains in user satisfaction for some riders and for improving overall public relations. Dynamic control is applicable whenever two low-frequency routes intersect, and it is productive to guarantee that transferring passengers will make their bus. Dynamic control is also applicable in cases where a low-frequency route receives a significant volume of transferring passengers from a higher-frequency route. By holding the vehicle on the low-frequency route to ensure that it meets an approaching vehicle on the other route, wait time is reduced.

More generally, dynamic control as a separate option is appropriate either when transfer flows are intermittent or when schedule unreliability is common. In the first case, dynamic control provides a way of making adjustments in operations only when they are needed to accommodate transferring passengers and is thus a substitute for schedule coordination. In the second case, dynamic control can cause buses that would not have met otherwise to meet, thus mitigating the effects of schedule unreliability on transferring passengers. This is particularly important on timed-transfer properties, where a guarantee that buses that will meet is necessary to attract new riders and ensure the satisfaction of old riders. Dynamic control with timed transfers may require some additional layover time, although not as much as if layovers alone were used to overcome reliability problems. In this situation, dynamic control is generally a workable compromise on both cost and user-satisfaction grounds between no alleviation of schedule uncertainty and the addition of layover time sufficient to absorb all schedule variance.

Timed Transfers

A timed transfer is defined as a set of operator actions that provides some degree of certainty that vehicles on different routes will meet at regular intervals to exchange transferring passengers. Timed transfers for buses can be divided into four distinct types:

1. Simple timed transfers: Two routes are scheduled and operated to guarantee that some or all buses on the routes will meet at the transfer point;
2. Pulse scheduling: Buses on all (or most) routes that meet at the major transfer point are scheduled to arrive nearly simultaneously, hold until all buses have come in, and then leave together;
3. Line-ups: In larger cities, buses on all (or most) routes that meet at the major transfer point are scheduled to allow a pulse-type exchange of passengers, typically in the context of low service frequencies and possibly long layovers at the transfer point, and most often in the evening; and
4. Neighborhood pulse: The schedules of neighborhood circulator routes are coordinated to make travel within a sector of a city easier.

Simple timed transfers are used on many properties, from the smallest to the largest, and are most commonly employed in the evening when both routes have low frequencies. Pulse scheduling is currently found in smaller cities (service-area population up to 300 000) and is used all day at central transfer points with a normal pulse frequency of approximately 30 min. Line-ups are used by many larger, nonpulse properties (service-area population of 500 000 or more), typically with headways of 1 h. Neighborhood pulse is currently being implemented on at least two large properties (Portland, Oregon, and Denver, Colorado) as part of their conversion to the bus transit-center concept.

In general, implementation of timed transfers requires several operator actions. Headways on different routes must be synchronized by altering route length and/or modifying layovers. Extra layover times may be needed to improve schedule adherence. This may also be accomplished by providing dynamic control of buses at the transfer point in case any are late. The operator must provide suitable space and facilities to permit easy simultaneous interchange of passengers between buses and must make important decisions concerning user information.

Property size is the principal criterion for timed-transfer applicability, serving as a proxy for headway reliability, service frequency, and the number of buses meeting at one time. Properties with less than 400 000 people in their service area are generally able to use pulse scheduling at their main transfer point. Larger properties often have line-ups at night. Simple timed transfers are usable on any property, while neighborhood pulse is applicable on any system with subcenters that serve as logical pulse points.

Service frequency and reliability appear to be the major determinants of whether user satisfaction is greatly increased by timed transfers. Ridership gains on the order of 5-12 percent appear reasonable with a highly reliable pulse-type timed transfer. Because increasing reliability generally costs money, the operator can implement timed transfers in different ways, depending on local objectives (e.g., increasing layover times and shortening routes versus adding equipment). Overall, however, timed transfers appear to be a cost-effective way of increasing service and ridership under certain circumstances without necessarily increasing costs (3).

Schedule Adherence on Connecting Routes

Schedule adherence is an important aspect of the overall level of service on transit properties that affects all (transferring and nontransferring) riders on the system. Major causes of bus schedule-adherence problems include traffic congestion, bunching, and, somewhat surprisingly, interference from trains and breakdowns of new buses. Remedies include skip-stopping, use of electronic and manual monitoring systems to control bus bunching, passing the first bus, and insertion of extra buses.

The principal consequences of increased schedule unreliability are an increase in the variance in transfer wait time and the expected (average) transfer wait time. These, of course, lead to decreases in user satisfaction, ridership, and revenue. However, the direct tranfer-related consequences of schedule unreliability are usually dominated by the nontransfer effects. Nevertheless, there are indirect transfer-related benefits that are nontrivial. If unreliability is too high, other operator actions regarding transfers (e.g., pulse scheduling or schedule coordination) are likely to meet limited success. Because the consequences of these other transfer-policy options can be very significant, minimizing unreliability for the purpose of aiding transfers can be an important objective.

Service Frequency on Connecting Routes

Service frequency, like schedule adherence, is an important component of transit level of service that has consequences beyond its impact on transfers. Given good schedule adherence, increasing the frequency of service on a connecting route should decrease the transfer wait time. Typically, however, operators raise or lower service frequency in response to non-transfer-related factors. The exceptions to this rule arise when other transfer components such as timed transfers, through-routing, and schedule coordination are implemented, since headways must be synchronized between routes. Even in these cases, however, the headway adjustments currently made are usually not large.

In general, user satisfaction, ridership, and revenue will rise due to the reduction in transfer wait time associated with service-frequency increases. Furthermore, there is a threshold headway of 10-15 min below which other transfer-related actions regarding scheduling may not be worthwhile due to limits on how great a reduction in transfer wait time they can produce. However, because of the significant cost of increasing service frequency, large changes in service frequency are typically only made in response to changes in overall demand or other factors and not simply transfer demand.

Transfer Charge

The transfer charge is the amount of money, over and above the basic fare, that a passenger pays to transfer to a second bus. Most bus properties currently charge nothing or $0.05 for a transfer. Other transfer charge levels are comparatively rare. Very few properties have full-fare transfers.

There is no consistent trend in transfer charges over recent years. Some properties have raised the charge slightly, e.g., from $0.02 to $0.05. Other properties have made transfers free. On average, nominal (and certainly real) transfer charges have tended to drift downward, although this tendency is neither pronounced nor universal.

A variety of reasons exists for setting the transfer charge at a particular level. In approximate order of importance, these include the following:

1. Historical precedent: Many properties have not recently given serious consideration to the level of their transfer charge;
2. Transfer abuse: A nonzero transfer charge may reduce the resale of transfers at a price below that of a full fare or the giving away of the transfers to friends and relatives, since fewer people would take transfers with the intent of later distribution if they had to pay something for them; and
3. Political or equity considerations: A particular transfer charge may be justified on the basis of a desire not to penalize transfers, not to subsidize long trips, etc.

Revenue, public relations, bus running times, and other considerations may also affect the selection of the transfer charge.

The cost consequences of a particular transfer charge result from the possible slowdown in bus passenger entrance and the minor cost of counting and handling additional revenue. The cost of slowing down the bus to process the transfer charge may be significant and results from both the need to pay a charge and from disputes that may develop between drivers and passengers over transfer abuse.

User satisfaction is decreased as the transfer charge goes up. However, the magnitude of this effect is determined by the disutility associated with charges by different user groups and the justifiability of the charge (i.e., the feeling among riders that the charge is fair and has a purpose, such as to make longer trips cost more). Both total bus ridership and the bus-to-bus transfer rate are sensitive to the level of transfer charge, although different types of trips will be affected differently by a given change in transfer charge. Captive riders have their riding patterns altered least by an increase in transfer charge, while shopping and other discretionary trips would be most discouraged. These ridership changes translate directly into effects on revenue. Because of the generally inelastic demand for transit, total revenue will typically increase (sometimes by a substantial amount) as the transfer charge goes up.

Clearly, operator goals and policies play a major role in determining the best transfer charge in a particular setting. For example, maintaining a low base fare to encourage total ridership may call for relatively high transfer charges for revenue reasons. A large deficit may also necessitate raising transfer charges to raise more revenue. Of course, transferring and overall ridership may be discouraged by a high transfer charge. Furthermore, it seems unlikely that a full transfer charge could be imposed on a system that currently has a high transfer rate without substantial political or equity problems. Overall, each of the three levels of transfer charge--zero, small but not nonzero, and full--seems to be stable and viable. The selection of one over the others is based on the operator's priorities and various other site-specific factors.

Transfer Slips

Transfer slips are the principal method for offering reduced-fare bus transfers that entitle riders to board subsequent vehicles at a reduced fare. Most properties use conventional transfer slips, but some use daily (or longer-term) passes or even no transfer slips at all to grant reduced-fare transfers. (No transfer slips at all is relatively rare but can be used when a small number of buses meet on a pulse schedule or when there is a restricted-access facility at the single transfer point on a system.)

The cost of administration of transfer slips is low. User satisfaction, ridership, and revenue consequences of transfer slips follow primarily from their use in setting fare policy and not from any characteristic intrinsic to the transfer slips themselves.

Schedule Information

Schedule information useful for transferring can be provided either at the transfer point or prior to the start of the transit trip. At the transfer point, transit properties can supply or post printed schedules or maps and/or disseminate information about whether the connecting vehicle is late. Prior to the trip, sources of information include printed schedules (which may include information on transfer points, time points, and best connecting vehicles) and telephone information systems. Schedules can also provide information on the other components of the transfer policy (e.g., through-routing, dynamic control, schedule coordination, timed transfers). In general, most properties only indicate the transfer charge and procedure for transferring on their schedule and almost never indicate the use of dynamic control or schedule coordination.

The direct costs of providing schedule information include printing schedules, manning phone banks, etc. The indirect costs are a type of opportunity cost, which occurs when a operator publicly states a transfer policy and then feels committed to it even when it becomes unproductive to do so. Provision of schedule information at the transfer point may raise the satisfaction of the rider by reducing uncertainty and by creating opportunities for other productive activities (which may be equivalent to a significant reduction in transfer wait time). Awareness of schedule information prior to the start of a trip will raise user satisfaction by enabling the passenger to make beneficial changes in trip-making behavior.

If the schedule and routes of a transit system never changed, provision of schedule information of all types would almost universally be the preferred action. However, transit routes and schedules frequently change, typically requiring an information or cost trade-off to be made. Each operator must determine whether the benefits produced by providing information are offset by direct costs and the need to make periodic adjustments in schedules and routes. This trade-off may be considered separately for each aspect of the transfer system that might be publicized. Route structure, for instance, is usually more stable than the schedule, so listing transfer points in a schedule poses fewer potential problems than listing the schedules of connecting routes. Also, the best connecting vehicle may change as a result of a small change in the schedule. Therefore, few properties put best-connecting-vehicle information in their printed schedules. Each component of the transfer policy may or may not be ruled out by the need to make periodic adjustments in schedules and routes on any particular property.

Marketing

Transfer-related marketing initiatives by the transit operator can focus completely on transfers, be part of a broader marketing effort, or use transfers to market other aspects of the transit system. When the transfer policy has some important distinguishing feature that can potentially affect a significant number of riders (e.g., pulse scheduling or universal transfer valid between carriers), it can be the focal point of a marketing effort. In addition, it is possible to market responsiveness by such means as a limited use of schedule coordination in which individual runs are adjusted to promote user satisfaction in response to user complaints.

Many properties promote transfers as part of broader marketing efforts. For instance, properties often produce brochures that describe their special services, including brochures on how to transfer. Transit fare-prepayment plans are another important example of the marketing of transfers as part of a larger effort, since the transferring rider usually receives free transfers. Transfer slips can also be used as part of marketing efforts that have nothing to do with transfers, such as the use of transfer slips as daily passes or special promotions in which retail establishments offer their customers return fares in exchange for transfer slips.

The cost of transfer-related marketing can be low, especially if there are few transfer points. The user-satisfaction consequences of marketing are related to the changes it can cause in awareness of and attitudes toward transit. Transfer-related marketing can change people's perceptions that transfers are onerous by promoting aspects of the transfer policy that make transfers easier. Marketing has been used to raise the awareness of different market segments concerning the existence of various transfer-policy components, as well as coverage and services provided by the system as a whole. It remains uncertain, however, whether marketing directed specifically toward transfers on properties whose transfer system has no special attributes is appropriate or productive.

CONCLUSIONS

Each of the 11 transfer-policy components described can be cost effective in various situations. The decision to use a particular policy component must address the trade-offs among that component's various consequences. For instance, the introduction of a transfer fee on a property where there were previously free transfers involves balancing equity, revenue, and user-satisfaction considerations. Trade-offs of this type are important from the point of view of the operator in determining how well a particular transfer policy meets the goals and objectives of the system.

It is important to reiterate that combinations of transfer-policy components may produce consequences that are not simply additive. For example, instituting timed transfers will, in general, have a positive effect on user satisfaction, as will increasing schedule reliability. However, the magnitude of the consequences of timed transfers will usually depend on the reliability of the connection, which would be enhanced by increasing schedule reliability. Hence, the increases in user satisfaction caused by implementing timed transfers and increasing schedule reliability may exceed the sum of the benefits derived from using those two components individually. Furthermore, some options have more widespread applicability than others; through-routing, for instance, can probably be implemented on a wider range of property types than pulse scheduling, although pulse scheduling has more far-reaching effects. Each operator must evaluate the service, cost, and demand conditions on the property and the consequences of alternative policies to determine which actions would be the most productive.

ACKNOWLEDGMENT

The material presented in this paper is drawn from the results of a recently completed study funded by the Transportation Systems Center (TSC) under the Service and Methods Demonstration program, UMTA. We would like to thank Robert Casey of TSC and Thomas Parody of Charles River Associates for their many helpful comments and observations. Also, the contributions of the transit professionals who participated in this study are gratefully acknowledged. However, the opinions and conclusions expressed or implied in this paper are ours, and we are solely responsible for any errors of content or ommission that remain.

REFERENCES

1. Charles River Associates. State of the Art of Current Practices for Transit Transfers. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, June 1980.
2. Charles River Associates. Transit Operator Guidelines for Transfer Policy Design. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, June 1980.
3. M. Nelson, D. Brand, and M. Mandel. Use and Consequences of Timed Transfers on U.S. Transit Properties. TRB, Transportation Research Record 798, 1981, pp. 50-55.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Short-Term Ridership-Projection Model

CY ULBERG

A statistical model has been developed for use by a transit agency in making short-term forecasts of transit ridership. These factors have been used successfully to plan service changes and to forecast revenues. A by-product of the use of the model is an increasing understanding by staff members of determinants of ridership changes and a corresponding reduction in the emphasis on ridership as a performance indicator by the agency. The model uses a combination of multiple-regression and time-series analyses to produce monthly projections of ridership. The variables included in the model were chosen for simplicity, ease of collection, and explanatory power. The validity and reliability of the model are quite strong, given its simplicity. During a two-year validation period, the average monthly error was 2 percent. Errors in annual totals were 0.9 and 1.7 percent, respectively. One objective in the development of the model was to make it a useful tool for planners and managers within the agency. A monthly report has been developed that has become a part of the decisionmaking process in the agency. Even though experience with a model has been limited, it has been demonstrated that a transit agency can make use of a relatively sophisticated (although simple) statistical technique to develop ridership forecasts.

In the past several decades, transit ridership has varied dramatically. Long-term trends have been influenced by phenomena such as the rising popularity of the automobile, world wars, and population shifts from farms into cities and suburbs. In contrast to these long-term trends, short-term ridership gains and losses occur due to more rapidly varying factors such as seasonal effects, service levels and quality, fares, gasoline prices and supply, parking rates, employment, and population. This paper describes one transit agency's experience with producing useful short-term forecasts.

Transit agencies use a variety of nonstatistical and quasi-statistical methods to produce forecasts of ridership. Generally, these methods use interpretations of past trends modified by management objectives for increasing ridership. Most agencies try to predict the impact of fare changes and service changes on ridership. In the Seattle metropolitan area, Metro Transit traditionally has projected ridership by using a modified Delphi technique. Objectives for productivity (passengers per hour) were set by using qualitative assessments of the environment, particularly the impact of fare and service changes. Service hours were projected by using budget constraints and perceived ridership demand. Total ridership projections were determined by multiplying productivity and service hours.

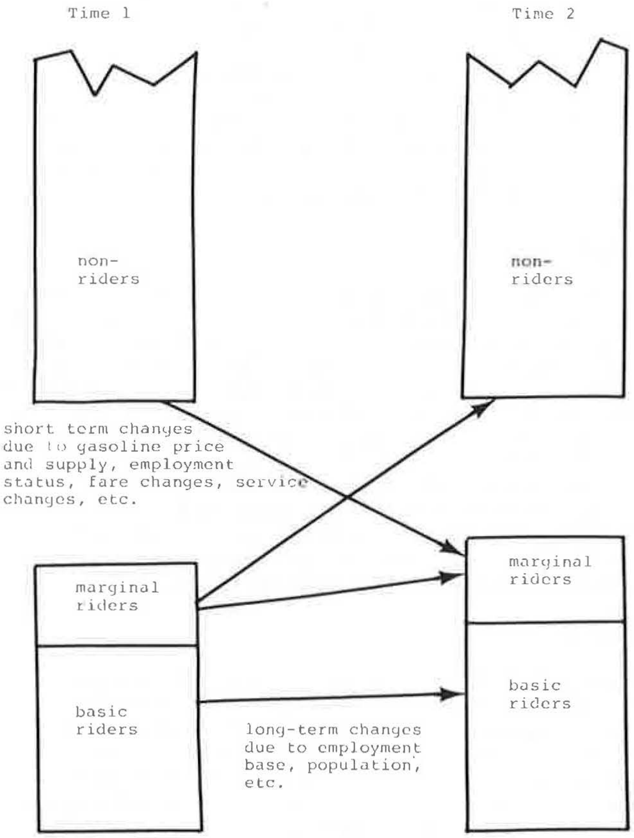
When ridership changes were relatively stable (such as between 1975 and 1979), these methods worked fairly well. However, in 1980 ridership trends changed abruptly. A gasoline crisis and rising employment were followed by a drop in gasoline price and declining employment trends. A major fare increase was implemented. Rapid increases in ridership changed to a leveling-off period. The extent of the change was unanticipated and resulted in major adjustments in service planning and budgeting.

In order to anticipate similar short-term changes in the future, Seattle's Metro Transit has developed a short-term ridership-projection model. It has been used during the past year to assist in the preparation of revenue projections and in planning service changes. It has also been used to anticipate the impact of a fare increase implemented in February 1982. Because the model uses variables extraneous to Metro's control, such as gasoline price and supply and employment, it has helped develop a new perspective on the use of ridership data for evaluating the effectiveness of the transit agency and its components.

BASIC STRUCTURE OF MODEL

A major objective in the development of the model

Figure 1. Model of behavioral assumptions that explain ridership.



was simplicity. Only five variables were used in projecting ridership:

1. Gasoline price,
2. Gasoline supply,
3. Service changes,
4. Fare changes, and
5. Employment.

Several other variables were considered for use but were rejected either because the data concerning them are inadequate or because they do not add significantly to the predictive ability of the model. Discarded variables include quality of service, population, employment in the central business district (CBD), parking prices, fuel efficiency of cars, and data disaggregated by time of day, route, or region served.

The model basically uses a multiple-regression approach. However, the dependent variable is the monthly change in ridership rather than the nominal value for ridership. Thus, the model is a time-series analysis. In addition, model development included investigation of the lagged effects of the independent variables. The assumption is that people do not respond immediately to changes in the economic environment.

The development of the model was guided by the constraint that data collection and analysis should be simple and straightforward. Statistical sophistication and rigor were sacrificed sometimes for simplicity of data collection and explanation of the model. A balance between rigor and simplicity was the goal. The development of the model was based on assumptions about aggregate responses to changes in the environment. The underlying behavioral model assumes that there are two types of transit riders: basic and marginal. Basic riders consist of transit dependents and those who ride for other reasons that do not fluctuate in the short run. This group changes with population and employment base. Marginal riders, or more aptly rides, are influenced by short-term phenomena such as gasoline price, gasoline supply, fares, employment status, and other economic factors. Figure 1 illustrates this model.

By using this model for underlying aggregate behavior, one would expect that the form of a model varies from agency to agency, depending to a great extent on the ratio of basic to marginal riders. Where basic riders compose the bulk of transit patronage, ridership would vary slowly except in response to changes in geographic coverage of the routes. Where marginal riders predominate, large fluctuations would be expected in response to changes in economic variables.

REGRESSION VARIABLES

In the process of choosing variables for the regression, several options were considered, including seasonal adjustments, lag times, and sources of data. This section details these choices and the development of data for the regression.

Ridership

Ridership is estimated monthly at Metro by using revenue data, including farebox collections, pass sales, and other estimates for special types of services. Periodic surveys are used to determine the proportion of farebox passengers paying special fares, transferring, or taking more than one zone trip. Recently, the introduction of automatic passenger counters on some of the buses has afforded the possibility to conduct reliability checks on ridership estimates. These tests show that estimates based on revenue data are fairly close to estimates based on actual counts.

Metro historically has published ridership data unadjusted for seasonal or calendar effects. Variations in average weekly ridership occur for different parts of the year (on the order of 10 percent). The number of working days, compared with holidays and weekends, can have a great effect on monthly totals (March generally has 15 percent more riders than February simply due to more weekdays and fewer holidays). In order to eliminate confounding variables in the regression analysis, ridership data are converted to seasonally adjusted average weekly ridership for each month.

Before the seasonal adjustment is made, an adjustment must be made for school services. Since fall of 1979, Metro has provided special service for school children, who account for about 3.5 percent of the total ridership. Because this service was not provided over the entire period of the data base, it is excluded from the historical data, and forecasts are adjusted with a separate prediction of this school service.

The first step in calculating the seasonally adjusted average weekly ridership is to add daily ridership figures together to produce weekly figures for each week of the month. The effect of holidays is eliminated by normalizing. For instance, if a holiday occurs on a Friday, the average Friday ridership for the rest of the month is computed and substituted for that day's data. Each week is standardized (i.e., the first week ends on January 7, the second on January 14, etc.).

In the second step, these weekly figures are converted to monthly averages. Each month is standardized (January has four weeks, February has four weeks, March has five weeks, etc.) so that there is an integral number of weeks in each month. The month's average is simply the total for that month divided by the number of weeks in the month.

The third step is to apply the seasonal variations to the months. Multiplying each month's average weekly ridership by a monthly factor gives the seasonally adjusted average weekly ridership used as the basis for the independent variable in the regression. Monthly factors are computed by averaging the deviation of actual ridership figures each month from the ridership figures predicted by the regression equation.

The fourth step is to compute the percentage change in seasonally adjusted average weekly ridership for each month. Ridership tends to have a wide variation from month to month. Therefore, a smoothing technique was added to produce stability in the data. The monthly change is taken to be the change over the average of the previous three months' figures. This change is converted to the equivalent of one month's change.

Gasoline Price

The variable that represents the price of gasoline is based on the price of no-lead regular gasoline. The no-lead price is used because it is the largest volume type of gasoline consumed. The data come from the Lundberg Survey, Inc., Recap of Wholesale Prices--Seattle. The survey shows the average price from each major oil company and the independents. The model uses the average of those figures.

The gasoline price is divided by the consumer price index for all urban residents of the Seattle-Everett standard metropolitan statistical area (SMSA). Using the real gasoline price was found to improve the multiple correlation coefficient by 10 percent compared with using the nominal gasoline price. Because it was expected that changes in the gasoline price would not affect ridership behavior immediately, a study was made to determine the lag time that best predicted ridership changes. The lag time that resulted in the lowest average residual was two months. The two-month change was converted to the equivalent monthly change by taking compounding into account.

Gasoline Supply

Gasoline price alone does not explain all changes in ridership, particularly during the gasoline crises of 1974 and 1979. The supply of gasoline was shown to be another separate important independent variable. The 1974 crisis can be characterized as one that had a moderately high growth in real gasoline price coupled with an extremely short supply of gasoline, while the 1979 crisis had an extreme increase in the price of gasoline and a moderate shortage in supply.

An attempt was made to use percentage shortfall in the state's allocation of gasoline as the variable to represent the supply problem. However, this variable was confounded so much by the cutback in use in response to the supply problem that it added little to the explanatory power of the model. The real problem that made people choose the bus was the inconvenience, as they saw it, in obtaining gasoline. A good quantifiable measure of this would be the average length of time waiting in line to get gasoline. Unfortunately, such data do not exist. As a surrogate for this information, newspaper articles written during the crisis were used as a basis for estimation. An energy specialist and I independently developed summaries of the problems on a month-to-month basis. These summaries were used to rate each month on a scale of 0-10 for severity of difficulty in obtaining gasoline. Adding these data to the regression served to reduce residuals during the 1974 and 1979 gasoline crises.

No gasoline shortfall is expected in the near future, so estimates of the variable that represent gasoline supply are not needed. If, however, one wished to assess the potential effect of a gasoline crisis, data from the 1974 crisis or the 1979 crisis could be introduced to represent the severity of a crisis at either of those two levels.

Fares

Four fare increases have occurred in the last eight years. In January 1977, fares were raised 10.5 percent on average. In January 1979, they went up 19 percent; in May 1980, 31 percent; and in February 1982, 10 percent. A trial regression was performed that included the effects of inflation on the fare price, but this added no explanatory power to the model; thus, in the interest of simplicity, it is not included in the data.

An investigation of the effect of lag in the fare variable revealed a relation similar to that with gasoline prices. The best predictor was the average monthly change over the previous two months.

Service

Each month the hours of service are computed by multiplying the number of hours of service on weekdays, Saturdays, and Sundays by the appropriate number of days in a standard month (i.e., 21.1 weekdays, 4.3 Saturdays, and 5.0 Sundays and holidays). The standard month is used because the dependent variable is weekly ridership adjusted for seasonal and calendar variations. Again, the effects of lag on the service-hours variable were investigated. No lag was found that resulted in a significant relation between service hours and ridership, except when service hours were lagged after ridership changes.

Employment

Raw employment estimates were computed monthly by the Research and Statistics Branch of the Washington State Employment Security Department. The model uses the change in employment for nonagricultural workers for the entire King County area, the service area for Metro Transit. Employment data were seasonally adjusted and the best lag was determined to be three months.

Calendar Variations

The level of ridership is highly influenced by the number of working days, Saturdays, Sundays, and holidays in a month. In transit agencies where ridership is recorded on a weekly or four-week basis, the only consideration is the number of holidays. However, at Metro, ridership data have traditionally been reported by calendar month, so an adjustment is necessary.

The dependent variable in the regression equation is a monthly percentage change in the seasonally adjusted average weekly ridership. By applying this percentage to the average of the previous three months' adjusted ridership, the average weekly ridership (adjusted for season) can be projected. The seasonal factor is applied to this figure to give average weekly ridership during a month. The next step in making the actual projection is to take into account the number of weekdays, Saturdays, Sundays, and holidays.

In order to compute the effect of the composition of a month, the total monthly raw ridership was divided by the average weekly ridership developed from revenue-based ridership estimates. This factor is generally slightly above 4. For each of the months during which ridership data exist, the number of weekdays, Saturdays, Sundays, and holidays was determined. By using multiple regression, a coefficient was determined for each type of day in a month where the dependent variable is the factor described above. Based on data through February 1982, the coefficients are as follows:

Type of Day	Coefficient
Weekday	0.1728
Saturday	0.0828
Sunday	0.0540
One-day holiday	0.0376
Two-day holiday	0.1773

Two-day holidays occur when a holiday falls on a Tuesday or Thursday, making either the Monday or Friday into a vacation day for many people. By applying these coefficients to the number of weekdays, Saturdays, and Sundays in future months, an estimate can be made of the calendar factor that must be applied to the weekly average ridership figure to compute the raw monthly ridership forecast.

REGRESSION ANALYSIS

The data were reduced to one dependent variable (the change in seasonally adjusted average weekly ridership) and five independent variables--change in real gasoline price, change in average fare, change in monthly service hours, change in perceived waiting time for gasoline, and change in employment. All of the variables used in the regression were converted to monthly percentage changes except the variable that represents gasoline supply. This allows regression coefficients to be interpreted as elasticities. It also allows a comparison of the strength of influence of each of the independent variables on changes in ridership. Each month, as new information is added to the data base, a new regression is performed.

The table below gives the regression coefficients for the model as of February 1982 (note that $R^2 = 0.694$, $F = 45.01$ (df = 99), and Durbin-Watson statistic = 2.07):

Variable	Coefficient	t-Value
Gasoline price	0.29	5.44
Gasoline supply	1.13	11.13
Fare	-0.14	-4.45
Service hours	-0.01	-0.16
Employment	0.80	4.01
Constant	0.38	3.29

The coefficients for all of the variables are significantly different from zero except for service hours. With only 5 variables, the regression explains about 70 percent of the variance in monthly changes in ridership. If nominal values rather than changes were used in the regression, R^2 would be much higher but the variables would be serially correlated. By using change rates rather than nominal values, the Durbin-Watson statistic is well within allowable limits.

The lack of relation between service hours and ridership in the regression deserves special comment. At least three factors explain this phenomenon. First, total service hours, as an aggregate, is too gross a measure of quantity of service. Some service-hour additions immediately attract new ridership while others may take a couple of years. If the data were available, the model could be improved by disaggregating service hours. Second, in a service area like Metro's, marginal riders predominate. Thus, variations in economic factors obscure the effect of variations in service. Third, new service hours have been implemented in response to changes in demand rather than preceding demand. Historically, changes in service can be predicted by changes in ridership rather than the other way around.

This regression is based on eight years of historical data. The more data the regression is based on, the more confident one can be in the regression coefficients. However, there is no reason to rule out the possibility that the relations between the independent variables and ridership change over time.

In order to test for this possibility, regressions were performed on subsets of the data for the time intervals shown in Table 1. The regression coefficients are fairly constant except for employment. An explanation for this phenomenon is in order.

In the early years of Metro, the relation between employment levels and ridership was low. One interpretation of this finding is that there was a large untapped market of commuters in the early years, so that the level of employment had little effect on ridership. In the past four years, however, as the system has become more strongly oriented toward serving the commuter and as that market is approaching saturation, new riders must come from new employment rather than an increased share of the employed. Hence, the coefficient for employment growth in the regression is higher than it used to be.

VALIDATION OF MODEL

During 1980 and 1981, Metro ridership underwent dramatic changes. These changes were due to a sudden end to large increases in the price of gasoline, rapidly declining rates of employment growth, and a substantial fare increase in May 1980. The exact behavior of the independent variables could not have been known at the end of 1979. However, there were signs that gasoline prices would decline and that employment rates would go down. The fare increase was already planned. Prepared early in 1979, the budget document for FY 1980 predicted a ridership of 62.2 million. Large ridership increases continued during the first few months of 1980 and, as a result, in the spring the ridership estimate for the year was raised to 68.3 million. The projection for 1981 was increased to 77 million. However, by mid-year ridership growth began to decline. Actual 1980 ridership was 66.1 million; in 1981 it was 66.0 million.

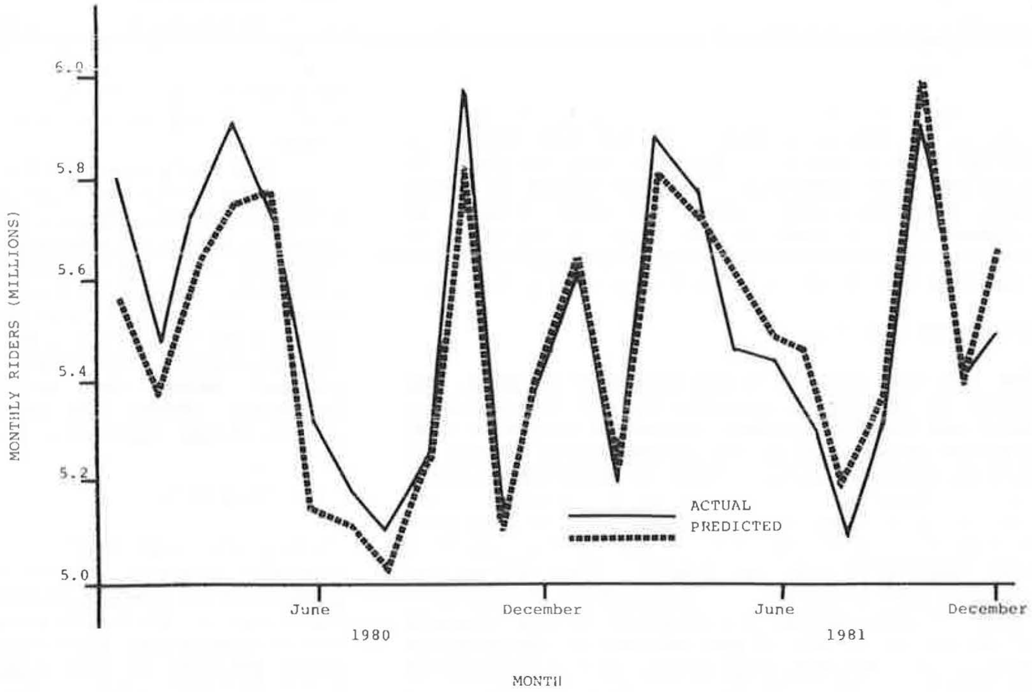
Figure 2 shows projections for ridership that the model would have produced at the end of 1979. These projections use data only up to that time as a basis for the regression. The projections for annual ridership in 1980 and 1981 would have been 65.0 and 66.6 million compared with the actual 66.1 and 66.0 million. The percentage of annual errors would have been 1.7 and 0.9 percent, respectively. The maximum monthly error would have been 4.1 percent, and the root mean square error over the 24 months, 2.0 percent. The correlation between actual and predicted ridership would have been 0.93. With seasonal and calendar adjustments taken out, the correlation would have been 0.75. All in all, use of the model would have allowed quite an accurate anticipation of the shifts in ridership patterns.

These years were very unusual, since there were wide variations in all the independent variables and the dependent variable. One should expect errors in

Table 1. Trends in regression coefficients.

Time Interval	Gasoline Price	Gasoline Supply	Fare	Service Hours	Employment	Constant Term
1/73-1/78	0.28	1.08	-0.14	-0.14	0.47	0.53
7/73-7/78	0.27	1.09	-0.14	-0.13	0.56	0.50
1/74-1/79	0.26	1.13	-0.15	0.09	0.61	0.48
7/74-7/79	0.34	1.15	-0.18	0.11	0.63	0.55
1/75-1/80	0.23	1.34	-0.19	0.19	0.42	0.61
7/75-7/80	0.23	1.34	-0.16	0.24	0.43	0.60
1/76-1/81	0.30	1.49	-0.13	0.09	1.17	0.09
7/76-7/81	0.32	1.50	-0.12	0.12	1.19	0.04
1/77-1/82	0.30	1.46	-0.14	0.08	0.97	0.23

Figure 2. Validation of 1980 and 1981 data.



monthly projections not to be so great during years when variables do not fluctuate so widely.

MAKING RIDERSHIP PROJECTIONS

One of the major objectives in the development of this short-term ridership-projection model was to produce information useful in service planning and budgeting. To this end, a monthly report was developed to present relevant information for wide distribution within the agency.

The first section of the report contains monthly projections for 24 months of ridership. Annual totals are included. These are monitored closely by several staff members and influence judgments about service increases and revenue estimates. Following these forecasts is a history of past projections. These are included primarily to give people a feeling for the accuracy of projections. It is more meaningful for people who have little statistical experience to see how the projections vary over time than it is to show the standard error or multiple correlation coefficient.

Most readers of the report are not interested in more than the first section. However, many readers have taken an interest in the data on the performance of the regression model. The regression coefficients are shown to give people a feeling for the importance of each variable. By showing the values of the coefficients for six months before, trends in the influence of the variables can be traced.

The last section of the report contains information about assumptions concerning independent variables. Historical information on real gasoline price and employment is included to give some basis for understanding the projections for these variables.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Connectivity Index for Systemwide Transit Route and Schedule Performance

TENNY N. LAM AND HARRY J. SCHULER

The services of a public transportation system are represented by its routes and schedules. The level and quality of service are, in part, determined by the ability of the route and schedule structure to serve the transportation needs of a service area. The concept of connectivity has been proposed for measuring how well the routes and schedules are integrated with respect to various transportation objectives within a framework of spatial-activity patterns. This study was undertaken to develop a methodology for determining the connectivity of the routes and schedules of an entire public transit system that serves a part or the whole of the service area. The objective is to use connectivity indicators as quantitative tools in the evaluation of service-delivery strategies. An investigation of the graph-theoretical connectivity by computer simulation found the mean of the reciprocals of the trip lengths of a representative sample of trips to be a good connectivity indicator. This indicator ties together the degree of connectiveness with the level of network development. It also offers a consistent picture of the level, as well as the quality, of connections offered by the route and schedule network structure.

The objective of the study is to introduce a methodology for defining and measuring network connectivity (1) for the purpose of evaluating transit system design and transit performance. The formulation provides a standard framework for evaluating problems of transit system performance on the basis of connectivity. The relevance of this line of inquiry lies in the overall importance of quality-of-service measures in assessing both the efficiency and effectiveness (2) of resource deployment and system design from a management or planning perspective.

In transit management practices, level-of-service measures are difficult to develop. First, there are many factors related to service quality (e.g., walking distance to transit stops, waiting time, travel time, and number of destinations served) that make it a multidimensional construct. Second, a transit system consists of many different routes. The extent to which the routes are integrated (to accommodate for transfers with a minimum of inconvenience) determines many of the service qualities of the transit system. Third, public transportation must be designed to serve peoples' needs. Perceptions of a transit system are influenced directly by individuals' use of the system and indirectly by the marketing program of the system operator. Because the perceived quality of service would likely not be the same to each individual engaged in travel, an analysis of the degree of network connectivity associated with fixed-route transit is a challenging research problem.

PROBLEM DEFINITION

Transit connectivity, defined as "the ability of a transportation network to provide the maximum number of origin-and-destination trip pairs through the optimal integration of routes, schedules, fare structures, information systems, and modal transfer facilities" (3), may serve as a valuable framework. This basic definition encompasses many of the considerations associated with the traveler's decision to use or not to use public transportation.

In modern transit management practices, quantitative information is required to develop evaluation tools that are necessary to assess system performance (4). In this context, connectivity is one of several measures useful for this purpose. However, it may be unrealistic and impractical to evaluate system performance on the basis of connectivity on any scale other than a relative one. As a relative measure of system performance, connectivity can provide useful information on alternative resource deployment and alternative system design. On an absolute basis, system comparisons would be controversial because of the uniqueness of environmental factors, route conditions, and operating characteristics that are uniquely associated with each transit system. Transit operators should realize these problems in order to render connectivity a meaningful role in management applications.

The basic managerial use of connectivity is as a standard against which a transit system might be compared. Deviant properties of the system would then be likely candidates for detailed examination. Connectivity measures may also be developed for performance characteristics. For example, one may consider connectivity as the percentage of potential or targeted trips that are served by the transit system with specific trip-time and distance performance measures. As one varies the performance specification, the connectivity indicators would also change. Because the performance level to be used is purely subjective, a continuous depiction of system performance with respect to the specification would allow the application of the base information to a number of potential situations. The approach eliminates the use of subjective value constraints in the application phase.

Other areas of potential use include evaluating route performance (both existing and proposed new routes) and periodic monitoring of transit operations on the basis of public investment in transit and derived benefits. Minimum standards for the continuation of service can be established and linked to route performance. In this manner, changing demand for transit service could be evaluated and the decision to continue or discontinue service could be internalized in the evaluation process.

In a public or quasi-public organization such as transit, evaluating public spending or investment is becoming increasingly important (5). Frugality is rapidly becoming a watchword in government appropriations. This consciousness with regards to the investment of public funds gave rise to the transportation system management (TSM) philosophy introduced originally in the 1970s. Consistent with the objectives of TSM, periodic monitoring of transit operations on the basis of connectivity can be linked to public spending in order to establish a return on investment in terms of overall performance. The potential exists to develop an approach that can make the evaluation of system performance more effective.

STUDY APPROACH

Two levels of transit connectivity are addressed. The first is the degree of connections between urban spatial locations provided by the transit network. For example, the degree of connections can be measured by the number of employment opportunities or locations accessible to transit from a particular residential address or by the number of homes that are within reach of a major shopping center by public transportation. This may be looked at from the point of view of mathematical graph theory. The connections between the spatial points may represent the idealized transit lines or they may represent point-to-point accessibility via the transit system. Transit connectivity expressed within a graph or network context is traditionally known as accessibility. Usually, accessibility means the ability provided by a transportation system to a person at a particular place such that he or she may go to other places that serve his or her needs. To what extent a transit system provides members of the community with accessibility would be one basis of evaluating transit system effectiveness. A transit network with many well-planned and coordinated lines would enable individuals within the service area to use public transportation to satisfy most of their mobility needs; therefore, such a transit system is well connected. On the other hand, a transit operation that has only a small number of disjointed lines would be poorly connected. An evaluation of the spatial configuration of transit service to different places within the service region can hence be made on the basis of some aggregated and weighted accessibility measures. One may expect that system accessibility measures would be strongly related to graph-theory-type connectivity indicators. A discussion of the graph-theory approach can be found in the literature (6-16).

A network of links and nodes, however, does not reflect the quality of service offered to users and potential users. It does not show the travel time, waiting time, walking distance, number of transfers, and transfer time. These user-oriented attributes can greatly affect user and community perceptions of the system's performance, ridership, and management policies. More significantly, a network representation fails to recognize the importance of system planning and the design of the route and schedule structure that are so essential for the efficient deployment of the often limited transit resources in order to maximize the system objectives. To take into consideration the route and schedule influence on the level of service offered to the users, a second level of transit connectivity indices needs to be introduced.

In order to look at transit connectivity in the proper perspective, it is necessary to consider the total system and problem setting and the different levels of factors that influence the development of the transit network, routes, and schedules and the quality of the services rendered. The system and problem setting can be viewed as a set of overlaying strata. They are the geographical terrain; the spatial-activity structure; the transit network, routes, and schedule; and the trip characteristics.

Each stratum represents temporal and/or spatial characteristics. By superimposing one on another, the effect that each level has on connectivity can be envisioned. For example, an area with natural barriers that channels urban development along narrow corridors has a positive influence on transit network, routes, and schedules. Relatively frequent transit services are provided and good connectivity is obtained because of the linear nature of the development. A simple route would provide connections between all places. A small number of vehicles on this single route would provide a high level of service without any need for transfers. On the other hand, for an area where the terrain is flat and urban development spreads in all directions, an extensive network, route, and schedule structure is necessary in order to achieve good spatial connectivity between all places. An extensive network with complicated routes and schedules is required to provide connections between all places. Even with greater expenditures, the services may not be as good as that for the linear city because the trips are so diverse that few trips may share the same direct routes.

A comparison of the performance of the transit network, route, and schedule systems cannot be made without recognizing the differences in the underlying geographical and spatial-activity distributions. For example, implicit in a bus transit network is an underlying highway network. A bus transit network, for all practical purposes, is a subset superimposed on the highway network. The urban structure also mirrors the shape, form, and function of the highways, which blend together the collective effects of geographical terrain, spatial distribution of activities, and urban development forces. There is an intricate interrelationship between the performance of a transit system and the dominating influence of highways. It would be difficult, if not impossible, to isolate the effects of the highway system on the transit network and the performance of the transit system. The influence of other transportation modes is also embedded in the connectivity and the service quality of any one mode of interest.

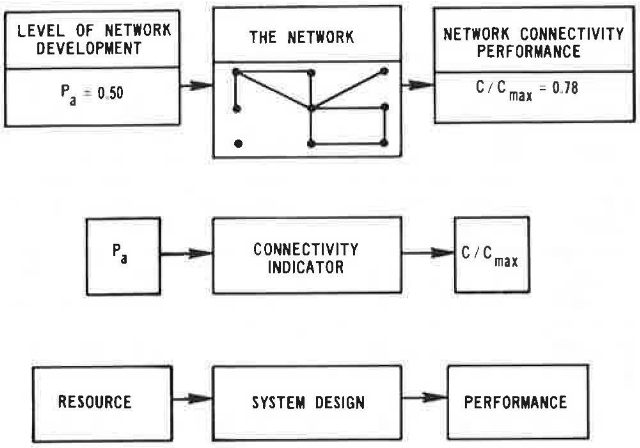
Transit connectivity must therefore be weighted by the trip patterns, although total ridership does not have a direct role on connectivity measures. Because transit demand is a function of many variables, including connectivity, the part of the total system demand that is subjected to the influence of system connectivity can serve as a basis for evaluating the performance of the transit network, routes, and schedule. Theoretically, connectivity, as a measure of how well the transit services are integrated through the coordination of routes, schedules, fare, etc., does not necessarily depend on demand. In reality, most transit operators cater to where existing or potential demand is the highest. The resulting transit system configuration reflects a great deal of the demand characteristics.

Transit operators have employed many different strategies in planning routes and schedules to effect the best services to the public. For example, take the use of timed transfer points. Timed transfer points are used to organize multiple nucleus radial routes such that maximum transit accessibility is provided for local collection and distribution and regional coverage with a relatively small number of simple routes. Although transfers may be required in such a network for many trips, there is a trade-off between short and coordinated transfers and areawide transit accessibility under the constraint of limited resources. Another example would be a highly directional commuter bus system that only provides accessibility between a limited number of spatial points but at a very good level of service for those who can use the system.

In addition to the four strata of system setting discussed previously, the key issues of community support, resource allocation, and management philosophy cannot be overlooked, especially in interpreting differences between systems. These issues, by themselves, do not enter into the measurement problem of connectivity, but the specification of the measures, the aggregation and sampling procedures for the data, and the interpretation of the indices depend on the policy viewpoints and the problem contexts.

The connectivity methodology developed here first ties resource input directly to system structure and system structure is then tied to performance outputs. In the form of a one-dimensional indicator, connectivity becomes a surrogate of resource input

Figure 1. Conceptual framework of network input and output and their relations with connectivity indicator.



and a surrogate of performance. In other words, the connectivity indicator is a reflection of both the level of input and the level of performance. In the sense of being a gross approximation of the characteristics of the overall system, connectivity is useful for comparing system alternatives between vastly different systems with respect to the quality of the route and schedule structure. Implicity embedded in the connectivity indicator should be a qualitative reflection of the level of system input and the level of system output. This concept is illustrated in Figure 1.

There have been many connectivity indicators proposed from graph and network theory points of view. Unfortunately, these indicators were found to be inadequate for the objective discussed here. Attention was given to the trip times as the focus of service quality. The travel time from origin to destination should be as short as practicable in the most extensively developed system with well-designed routes and schedules. Therefore, trip time is a good measure of the quality of the service in terms of mobility. However, the accessibility question must also be addressed. In other words, one cannot overlook the question of how well the transit services serve places--namely origins and destinations. The extent that places are connected by transit may be expressed in the form of the percentage of potential trip origin-destination pairs serviceable by transit. The primary focus of the study was on how to integrate these two level-of-service qualities that are determined by the route and schedule structure.

RESEARCH FINDINGS

Many network-connectivity indicators previously introduced have been examined with respect to their ability to represent the level of network development and system performance. System performance can be measured by both the directness of the route between an origin and a destination and the level of connectedness between all origins and destinations. None of the existing network-connectivity indicators offers a consistent picture among the level of resource input, number of links in the network, and output performance. Details of the investigation on graph-theory-related connectivity measures can be found elsewhere (17).

A new indicator was developed in this study. This indicator is the harmonic mean trip time for a representative sample of trips. When different experimental networks were examined under this indicator, it showed the expected correlations between the input and output measures. Suppose there are n trips that are representative of the travel within the service region of the transit system. Each trip is identifiable by its origin O_i and its destination D_i. In a fully developed network it can be assumed that potentially a large number of routes should be available, such that every origin-destination pair in the service region is served by the system at some standard rate of service in terms of frequency and overall trip speed. For the n trips in the representative sample, the trip times can be determined under this hypothetical fully developed system. If T_i is the trip time (weighted or unweighted for access time, waiting time, on-board time, and transfer time) of the ith trip in the sample between O_i and D_i in the fully developed system, the harmonic mean is given by the following:

$$\bar{T} = 1/(1/n)[(1/T_1) + (1/T_2) + \ldots + (1/T_n)] \tag{1}$$

Most networks, however, are much less developed than the hypothetical fully developed case. Therefore, for the same n trips in the sample, the actual travel time for trips i will be much longer than T_i. Let the actual trip time for trip i between O_i and D_i be denoted by t_i. A harmonic mean $\bar{t}$ of the actual trip time can be calculated as follows:

$$\bar{t} = 1/(1/n)[(1/t_1) + (1/t_2) + \ldots + (1/t_n)] \tag{2}$$

The connectivity indicator, which is the normalized reciprocal harmonic mean trip length, or R, is given by $R = \bar{T}/\bar{t}$. Because $\bar{T}$ is for the ideal case of full development, the actual trip time t_i is at best equal to the ideal trip time T_i and would be longer for most trip, i.e., $t_i \geqslant T_i$. As a result, $\bar{t} \geqslant \bar{T}$ and $0 \leqslant R \leqslant 1$. If R is equal to 1, the system is ideal. However, if trip i in the sample cannot be served by the system, we assume the actual trip time to be infinite, i.e., $t_i = \infty$. This also reflects the quality level of the transit service. When the travel time is long (in some poorly connected case this can be many hours or days), the reciprocal $1/t_i$ is small and contributes little to the harmonic mean. In the extreme case, when $t_i = \infty$, $1/t_i = 0$. For example, if there are five trips in the sample, let their ideal trip times be 25, 5, 16, 8, and 35 min. The harmonic mean $\bar{T}$ of the ideal trip times is as follows:

$$\bar{T} = 1/(1/5)[(1/25) + (1/5) + (1/16) + (1/8) + (1/35)] = 10.96 \text{ min} \tag{3}$$

Suppose for the actual network the second and fourth trips are not connected and the actual trip times are 25, ∞, 20, ∞, and 40 min. The harmonic mean $\bar{t}$ of the actual system is as follows:

$$\bar{t} = 1/(1/5)[(1/25) + (1/\infty) + (1/20) + (1/\infty) + (1/40)] = 43.48 \text{ min} \tag{4}$$

The resulting connectivity indicator R is then

$$R = \bar{T}/\bar{t} = 10.96 \text{ min}/43.48 \text{ min} = 0.25 \tag{5}$$

If the actual trip times for the five trips are 26, 60, 20, 45, and 35 min instead,

$$\bar{t} = 1/(1/5)[(1/26) + (1/60) + (1/20) + (1/45) + (1/35)] = 32.07 \text{ min} \tag{6}$$

and

$$R = 10.96 \text{ min}/32.07 \text{ min} = 0.34 \tag{7}$$

In the extreme case when no transit service is

Figure 2. Network connectivity performance C/C_{max} plotted versus level of network development P_a.

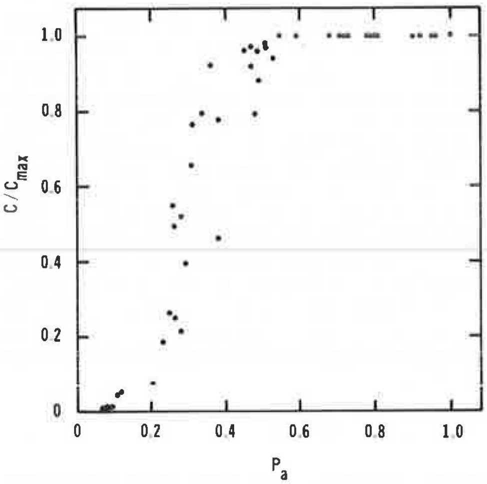


Figure 3. Relation between reciprocal normalized harmonic mean of trip lengths and level of network development P_a.

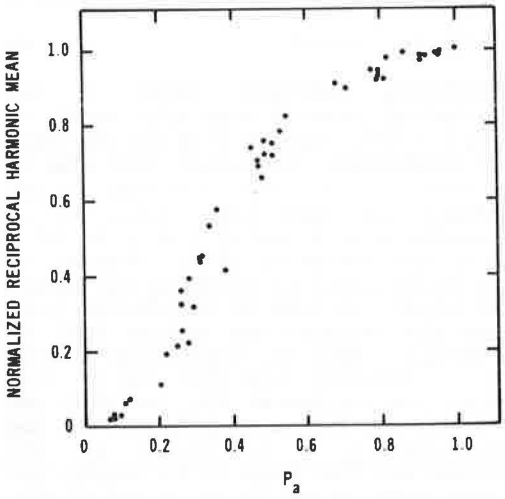
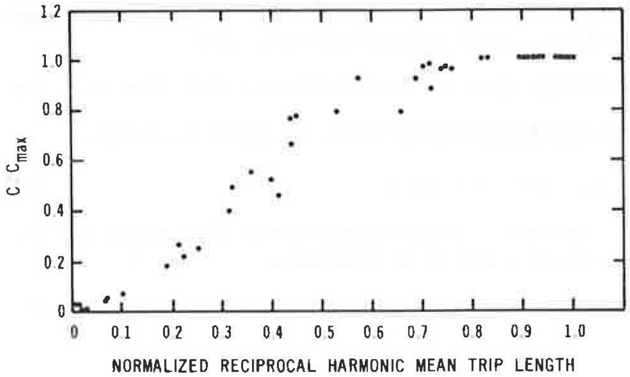


Figure 4. Correlation between network connectivity performance index C/C_{max} and reciprocal normalized harmonic mean trip length.



provided, the connectivity indicator R, so defined, is zero.

The results from the research experiments are shown in Figures 2, 3, and 4. Figure 2 shows the usual quantitative connectedness measure C/C_{max}, where C is the actual number of origin-destination pairs connected by the network, while C_{max} is the total number of origin-destination pairs in the sample. This indicator is related in Figure 2 to the level of network development P_a. The indicator P_a is the ratio of links between the actual system and the hypothetical fully developed system. One can see that as more and more links are provided, P_a increases. As P_a increases, the connectedness offered by the system also increases. However, the index C/C_{max} does not provide any insight into the quality of the connections. The relation between C/C_{max} and P_a is also undesirable because of its abrupt change in the middle range of network development, namely when $0.2 \leq P_a \leq 0.5$.

However, the normalized inverse harmonic mean of trip length R offers a much more smooth relation with P_a. As a result, a better differentiation between system performances is possible, as shown in Figure 3. Moreover, Figure 4 shows a very good relation between C/C_{max} and R. The result suggests that R is a very satisfactory indicator. Because trip length is used in this indicator instead of some abstract mathematical notions, the usual representation of system service by trip length (i.e., time) and the weights associated with the different components in trip length (or time) can be maintained. The weights developed from attitudinal and behavioral studies are useful to reflect the human perception of the quality of the transit services.

METHODOLOGY

There are two remaining problems that must be resolved in order to make the connectivity indicator--the normalized reciprocal harmonic mean trip time R--operational. One is the strategy for statistical sampling and another is on establishing some reference of the indicator to local geographical, highway system, and transit system conditions that are unrelated to the quality and performance of routes and schedules.

Definition of Study Area

The definition of the study area is a policy-oriented issue and is beyond the realm of the present research. However, some general discussion can be offered here as to how the definition of the study area may be addressed. If the policy question is on the service quality within the transit district at large, the study area should be the entire district. A study area so defined will yield a connectivity indicator that is broadly based from the point of view of the total community, independent of the marketing and operational strategies. Another definition of the study area is the effective service area, which may be the area covered within 0.25 mile on both sides of all transit routes. However, express routes should be defined with service areas that are within the actual or expected catchment basin of each station or terminal.

The distinction between total district area and effective service area does not pose any difficulty for the connectivity-measurement procedure developed here. Both spatial connectivity and within-system connectivity are measured. In general, system connectivity should integrate both. Therefore, the definition of the study area is not too critical.

More significant is the definition of the perimeter and the size of the total area. The exact definition is an administrative and policy matter and is not a technical issue.

Procedure for Developing Trip Samples

The basis of the connectivity indicator is the travel characteristics of the target population of the transit operation. The target population may consist of everyone in the metropolitan area or may consist of the potential and present users of the transit services or special subgroups. The special subgroups may be the patrons of particular land use types or particular social services or certain socioeconomic groups.

It is neither practical nor necessary to calculate the connectivity of the service offered to each and every trip of the entire target population. All of the required information can be obtained by evaluating the service to a small sample that represents the trip characteristics of the entire population. Statistical sampling is widely used in all kinds of surveys, in engineering and scientific studies, and in management practices. In transportation, almost all the information used in planning and analysis comes from samples of a very small number of individuals and trips. In home-interview surveys, the percentage of households included in the studies varies from 20 percent in small cities to about 2-3 percent in large metropolitan areas. Most transit surveys usually involve samples of less than 1000 individuals.

Within the context of the present study, two strategies may be used to develop the trip sample. Where there already exists an extensive travel survey conducted recently, the survey may be used. Depending on the connectivity indicator to be developed, the entire trip sample may be used if the travel time and travel distance are included in the survey. If only origins and destinations are available from the survey, travel times and distances may have to estimated. The estimation of travel times and distances is costly and time consuming and, therefore, only a small sample is practical.

Estimation of System Performance

Travel time on the transit system is used as the basic data for determining the connectivity indicator. For each of the trips in the sample it is necessary to measure the transit time, the distance between the origin and destination, and, if access time and waiting time are included, the estimation of the access and waiting time. The transit time should include all transfer times and number of transfers as well as walking time between transfer points. Previous studies have indicated that transit users place more weights on access times, waiting times, and transfer times than on the on-board times. By determining these separate time elements, proper weights can be assigned to them and a weighted total transit travel time may be determined.

For the connectivity measuring concept developed here, there is no need to set arbitrary cut-off criteria on whether a trip is effectively connected. The contribution of a long trip (even unrealistically long) can be readily incorporated. The longer the trip time is relative to average transit system performance, the less its value is in terms of spatial connectivity.

The determination of the travel-time elements is based on the origin, destination, and starting time of the trip. Knowing the input information, the travel time can be determined from transit system route maps and timetables. If more accurate information is required, the transit travel time, etc., can be actually measured by taking the actual ride. However, it is inconceivable that such a procedure is necessary unless the timetable information is very inaccurate. Occasionally, in the absence of trip-time information, the connectivity indicator can be measured in terms of route distance between the origin and destination of the trip. The distance information is useful to complement the time information, rather than in lieu of the time information.

The purpose of determining the transit route distance and the straight-line distance is to facilitate the development of the reference base necessary for making connectivity indices comparable for different transit operations. In the connectivity indicator, the reference base is the travel time on a hypothetical transit system that is fully developed. By fully developed, the average speed on the transit system without transfers is applied to the most direct highway route that connects the origin and destination of a trip in the sample. Therefore, it is necessary to determine the average transit system speed and the average highway speed. In addition to the ratio between transit route distance and straight-line distance, the ratio between automobile-route distance and the straight-line distance is also useful. This information may be obtained for the trips in the trip sample or independently. Actual field measurements may be used from standard transit and traffic travel-time studies. Or, where there exists an updated urban transportation planning analysis network, the information may be obtained from computer network analysis.

Calculation of Connectivity Index

Table 1 gives an example of the type of information for a sample of 30 trips. Of the 30 trips in the sample, 10 are not served by transit. For these trips, the transit travel time is infinite. The first step in the calculation is to compute the reciprocal of the harmonic mean by the formula

$$(\bar{t})^{-1} = (1/n) \sum_{i=1}^{n} (1/t_i) \tag{8}$$

where t_i is the total transit time in column 4 of the table.

For the example in Table 1,

$$\begin{aligned}(\bar{t})^{-1} &= (1/30)(0.74)\\ &= 0.024\,58\ (\text{min})^{-1}\end{aligned} \tag{9}$$

and $(\bar{t})$ = 40.68 min. In order to determine the reference base, the travel times of all the trips in a fully developed transit network are estimated. For the fully developed transit network, the direct route is assumed to be the shortest highway route. On this fully developed network, transit speed is assumed to be the route speed for those transit trips that are served. The route speed is given in column 9, which is determined from the on-board time and the transit-route distance. The on-board time is the total transit time minus the transfer time. Multiplying the average of column 9 to the highway-route distance in column 8, the value t_i is given in column 10 to represent the equivalent transit travel time over direct routes between the origin and destination over the fully developed transit network without transfers. The average route speed for the example is 17.35 mph. The reciprocal har-

Table 1. Example of transit performance data.

Trip No.	Origin	Destination	Total Transit Time (min)	Transit-Route Distance (miles)	Transfers No.	Transfers Time (min)	On-Board Time (min)	Highway Distance (miles)	Transit-Route Speed (mph)	Transit Time, Fully Developed Network (min)
1	2804	4710	-	-	-	-	-	16.8	-	58
2	3403	3401	-	-	-	-	-	3.2	-	11
3	1802	4714	103	18.2	2	35	68	13.0	16.1	45
4	3004	1501	5	2.6	0	0	5	2.8	31.2	10
5	2702	1803	5	1.45	0	0	5	1.5	17.4	5
6	1802	1508	70	4.8	1	55	15	3.5	19.2	12
7	2902	4401	48	10.7	1	5	43	5.9	14.9	20
8	3607	1303	99	22.45	2	20	79	11.1	17.1	38
9	2702	4302	57	13.7	2	8	49	9.4	16.8	33
10	1501	2102	35	6.65	1	10	25	6.7	16.0	23
11	4399	4705	56	10.0	1	6	50	5.1	12.0	18
12	1206	4711	-	-	-	-	-	4.9	-	17
13	1402	2501	49	11.7	1	13	36	8.0	19.5	28
14	2101	3501	51	12.8	1	10	41	11.4	18.7	39
15	3611	4702	25	5.0	1	1	24	4.7	12.5	16
16	4710	1207	-	-	-	-	-	6.7	-	23
17	1901	2908	-	-	-	-	-	2.3	-	8
18	1602	1706	-	-	-	-	-	5.6	-	19
19	1803	3601	-	-	-	-	-	18.0	-	62
20	2102	2903	-	-	-	-	-	13.2	-	46
21	4711	1402	101	21.2	2	30	71	14.5	17.9	50
22	5004	4803	-	-	-	-	-	7.9	-	27
23	1203	3901	73	17.5	2	13	60	10.9	17.5	38
24	3302	1603	73	14.4	2	20	53	6.1	16.3	21
25	4301	1801	103	19.25	2	35	68	12.1	17.0	42
26	3606	2807	49	11.4	1	5	44	7.5	15.6	26
27	4704	3403	39	7.45	1	10	29	3.0	15.4	10
28	2903	4711	46	8.3	1	10	36	8.2	13.8	28
29	2908	4706	-	-	-	-	-	2.7	-	9
30	5002	1302	42	11.8	1	10	32	6.4	22.1	22

monic mean of the t_i's in column 7 is as follows:

$$(\bar{T})^{-1} = (1/n)\,[\sum_{i=1}^{n} (1/t_i')]$$
$$= (1/30)(1.65)$$
$$= 0.054\,89\ (\text{min})^{-1} \quad (10)$$

and $\bar{T}$ = 18.22 min. By using the concept that the connectivity indicator is the ratio between the actual reciprocal harmonic mean transit time and the reciprocal harmonic mean transit time on a hypothetical fully developed network, the connectivity indicator R is given by R = $\bar{T}/\bar{t}$. For example, the connectivity index of the transit network that serves the 30 trips in the sample is as follows:

$$R = 18.22\ \text{min}/40.68\ \text{min} = 0.45 \quad (11)$$

CONCLUSIONS

The objective of this study is to develop operational indices to represent the ability of a transit system to connect urban places and the quality of service provided on the connections. Connectivity is related first to the structure and the level of development of the transit network. Then the connection between two points on the transit network is influenced by the coordination of the routes and schedule. The routes and schedule, in turn, are influenced by management policies on resource allocation and deployment.

The difficulty for developing connectivity indices lies in the many complex interacting factors involved in transit service delivery. There are great differences among the geographical, land use, highway, and user characteristics between regions. The indicators developed must, hence, incorporate other measures that could be used as references from which the actual performance of the network, routes, and schedule of the transit system can be measured. The resulting measurement should be realistic in representing subjective evaluation of the quality of connectivity, flexible in allowing different data-collection procedures to be used, and robust in its applicability to all systems.

In this study, the main focus is on identifying the contribution of transit system connectivity to the overall performance of how well urban-activity connections are served by transit. The study approach involves looking at the problem from the perspectives of graph theory, urban transportation planning models, and statistical sampling. Attempts were made to develop sets of measures that would reflect, as much as possible, transit connectivity viewed from both accessibility and level-of-service points of view. The evaluative and performance measures such as accessibility and quality of service are commonly used in almost every aspect of transportation planning. They reflect many important planning and management factors. Connectivity of the transit system's network, routing, and scheduling is only one of the factors. Care must be used in not confusing the evaluative and performance measures with connectivity measures, despite the fact that connectivity reflects the level of transit service.

A number of remaining questions need to be addressed before full implementation and application should take place. One question is the sensitivity of the indicator to sample size. This question can be readily resolved with a sensitivity analysis of the results with samples of different sizes for the same area. The next question is on how trip samples should be drawn with respect to different types of issues. Should spatial area or traffic zone be used as the trip sample base? Should the sample be based directly on a surveyed sample of trips or a sample of trips from available planning model information? Should 24-h trips be used or trips within some specific period of time? Should the trips be sampled for weekdays as well as weekends? Should transit trips be used or should all personal trips be used?

The sampling questions cannot be answered except within the specific context of a problem or issue to be addressed. When application is to be made, it is necessary to first detail the objectives of the application. What exactly is of interest within the policy and issue context? What role does the route and schedule play within the context? How does connectivity enter into the consideration? What would the indicators mean with respect to the issues? How should the indicators and the results be interpreted in answering the questions being addressed?

With respect to the application to be made, an interview with managers of each of the transit operations to be involved should be made to qualitatively determine the subjective impressions of those intimately knowledgeable of the systems. The calculated indices must also be correlated with the subjective impressions. The purpose of the indices is to provide a systematic basis of estimating and quantifying subjective impressions. Therefore, the indicators should correspond to the collective wisdom of the experts. A good correlation between the quantitative and subjective evaluations should adequately validate the methodology and the procedure. As a result, the connectivity indicator would then have the necessary credibility and acceptability for full implementation.

ACKNOWLEDGMENT

The study benefited from the participation of Wilfred W. Recker. Michael J. Uyeno and Behrooz Mohammadi assisted in much of the technical work.

The study was supported, in part, by an Urban Mass Transportation Administration technical study grant awarded to the State of California. The Institute of Transportation Studies, University of California, Irvine, performed the research with the Division of Mass Transportation, California Department of Transportation. The views expressed here are ours and not necessarily those of the University of California, the State of California, or the U.S. government.

REFERENCES

1. P. Haggett and R. Charley. Network Analysis in Geography. St. Martin's Press, New York, 1969.
2. G.J. Fielding and R.E. Gauthier. Distribution and Allocation of Transit Subsidies in California. Institute of Transportation Studies, Univ. of California, Irvine, Rept. ITS-I-SR-76-1, 1976.
3. Glossary of Urban Public Transportation Terms. TRB, Special Rept. 179, 1978.
4. J.S. Dajani and G. Gilbert. Measuring Performance of Transit Systems. Transportation Planning and Technology, Vol. 4, 1978, pp. 97-108.
5. E. Fuller, Jr. Performance Measures for Public Transit Service. Division of Mass Transportation, California Department of Transportation, Sacramento, Rept. CA-09-8001, 1977.
6. T.N. Lam and M.J. Uyeno. A Bibliography on Network Connectivity as Related to Public Transit Systems. Department of Civil Engineering, Univ. of California, Davis, 1978.
7. L.W. Beineke and F. Harray. The Connectivity Function of a Graph. Mathematika, Vol. 14, 1967, pp. 197-202.
8. S. Even. An Algorithm for Determining Whether the Connectivity of a Graph Is at Least k. SIAM Journal of Computing, Vol. 4, 1975, pp. 393-396.
9. I.T. Frisch. An Algorithm for Vertex Pair Connectivity. International Journal of Control, Vol. 6, 1967, pp. 579-593.
10. W.L. Garrison. Connectivity of the Interstate Highway System. Proc., Regional Science Association, Vol. 6, 1960, pp. 121-137.
11. D.R. Ingram. The Concept of Accessibility: A Search for an Operational Form. Regional Studies, Vol. 5, 1971, pp. 101-107.
12. K.J. Kansky. Structure of Transportation Networks: Relationships Between Network Geometry and Regional Characteristics. Department of Geography, Univ. of Chicago, Res. Paper 24, 1963.
13. F.P. Stutz. Accessibility and the Effect of Scalar Variation on the Powered Transportation Connection Matrix. Geographical Analysis, Vol. 5, 1973, pp. 61-66.
14. E.J. Taffe and H.L. Gauthier. Geography of Transportation. Prentice-Hall, Englewood Cliffs, NJ, 1973.
15. M. Tainter. Statistical Theory of Connectivity I: Basic Definitions and Properties. Discrete Mathematics, Vol. 13, 1975, pp. 391-398.
16. C. Werner. A Research Seminar in Theoretical Geography: Networks and Their Service Areas. In Geographic Studies in Urban Transportation and Network Analysis (F. Horton, ed.), Northwestern Univ., Evanston, Studies in Geography No. 16, 1968, pp. 128-270.
17. T.N. Lam and H.J. Schuler. Public Transit Connectivity. Division of Mass Transportation, California Department of Transportation, Sacramento, Rept. UMTA/CA/MT-81/084, 1981.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Automation in Public Transit Operation and Management: Update

DONALD WARD, MICHAEL COUTURE, RICHARD ALBRIGHT, AND GRANVILLE E. PAULES

The concept of a system of operations, planning, and management tools for transit operators is discussed in light of recent developments in automation. Minicomputers and microcomputers have improved the ways in which computers are perceived, increased their acceptance, and vastly increased their potential uses for transit planning. This paper discusses the elimination of barriers to the use of automation, some events that create an ideal opportunity for development of an operations planning system, and a general concept for use of microcomputers and automated tools. A sample application that involves the analysis of route-level data for the purpose of route-performance evaluation is presented. Design issues for the development of transit operations tools are discussed.

The decade of the 1970s was a period of expansion for the transit industry. As the demand for transit rose, in part due to the rising costs and intermittent shortages of gasoline, coverage was extended to more outlying areas of urban regions and service hours were increased. Ridership and revenues increased but not as fast as operating costs. Deficits were covered by both the federal government, which initiated the Section 5 operating assistance program (Urban Mass Transportation Act of 1964, as amended), and by local and state sources, which appeared willing to increase their support to transit. In general, there was not a strong incentive for most operators to vigorously seek ways to improve productivity and keep costs to a minimum.

Now, a new conservatism, spearheaded by a new administration, has led to changes in transit operating philosophies and the need to review and revamp the service provided. Financially hard-pressed local areas can no longer support rapidly rising operating costs and must accept cutbacks in service and fare increases. The planned phase out of federal operating assistance over the next few years will also hurt, particularly small and medium-sized properties.

Although federal funding support is being reduced, the federal government recognizes the substantial payoff in supplying the transit industry with technical aids and information designed to improve service delivery. Transit operators have never had the appropriate tools to be able to plan and operate service at a high level of efficiency. Further, they have rarely had even the information required to make good decisions consistently. In the current planning and operating environment, the ability to forecast the results of possible service and operations changes may be essential to the survivability of many operators in the coming lean years. As an example, most operators cannot estimate the net revenue impact of an operating change or, even in some cases, whether the impact might be positive or negative.

This paper describes the potential development and use of a system of automated tools that will enable transit operators to plan, maintain, and operate service efficiently (i.e., at a level of efficiency that will be acceptable to those who use it and to those who pay for it).

BACKGROUND

The concept of a system of transit planning, management, and operations aids in the form of handbooks and especially automated tools is not new. An early comprehensive description of what such a system should entail was prepared by the Transportation Systems Center in 1978 (1). That paper discussed the requirements of an operations-planning system (OPS) based on a number of structured conversations with transit operators across the country. The paper presented several examples of applications and a possible system design. In addition, it described in some detail the performance requirements and preferences for computer hardware and software capabilities that were perceived to be appropriate at the time.

Another 1978 paper (2) prepared by the Urban Mass Transportation Administration (UMTA) outlined many of the potential applications and benefits of automated data processing in the transit industry. It also suggested exploitation of emerging low-cost minicomputer and microcomputer technologies and state-of-the-art communications methods. A 1980 paper (3) updated the OPS concept by describing potential uses of the still-evolving microcomputer and minicomputer technologies. It also raised a number of issues relating to how automated processes could be designed and developed.

Since these papers were written, a number of events have taken place that provide increased impetus for the development of an OPS and suggest slightly different frameworks for both its development and use. First, the transit industry is again experiencing severe financial difficulties. The past decade saw expanding service and steadily increasing operating assistance at local, state, and federal levels. Operations and planning needs were often focused in such areas as new route design or schedule improvement. Although these remain important topics for planning-method improvements, a shift in emphasis is necessitated by the new transit operating environment. Reductions in operating subsidies that are forcing higher fares and service contractions will require a redirection of operating policies toward improvements in productivity and efficiency. Thus, tools that will quickly evaluate the effects of alternative fare increases and/or service changes are needed. Capital-planning tools to achieve the proper balance between facilities and operating costs are of extreme importance. Performance evaluation, cost analysis, and management information reporting are areas where help is needed badly, and soon.

Another significant event is the release of the first year's worth of Section 15 operating data (Urban Mass Transportation Act of 1964, as amended) (4). Now that comparable operating and financial information from most of the country's transit properties is available for all to see, it is certain that operators will become much more aware of areas for improvement and will appreciate help in realizing potential gains.

A third major event of relevance to transit planning is the rapid growth in the availability and use of small computers, which now can be found in common use in schools, libraries, small businesses, hospitals, and homes. Standardized components, peripheral equipment, and a variety of useful and inexpensive software make small computers ideal for an OPS. Their features and advantages have significant benefits for both the development of an OPS and the nature and extent of its use.

This paper describes a development philosophy for the OPS, a concept for its use, and a scenario for a subset of planning modules that would be representative of the system.

ELIMINATION OF BARRIERS TO USE OF AUTOMATION

There are a number of reasons why automated planning tools have not been extensively developed for the transit industry and why those that are available have not been widely used. The most important of these is cost for both development and use. Although the costs of computer hardware and machine time have been decreasing steadily for many years, it is only recently that the price of using automation would not represent a significant item in the budgets of small and medium-sized transit properties.

Today, for example, a microcomputer and associated peripheral equipment with significant capabilities and capacity can be bought for less than $10 000, and this figure will likely fall further. Also, there are no computer use charges.

However, the cost and time required to develop software for larger machines are items that have not decreased substantially, if at all. These costs, combined with other problems discussed below, have inhibited the development of planning tools for the transit industry. The risk of an expensive effort that might only result in limited use was too great, in many cases, to allow development to take place.

It is likely that cost efficiencies associated with microcomputers and minicomputers can change the nature of software development. For example, a large and increasing volume of commercial software that is transferable across a variety of machines is available. These packages, which approach the power of software that costs tens of thousands of dollars and that can be bought for a few hundred dollars or less, can be easily modified and tailored to transit planning and operations needs. If applications software developed on and for small machines can be acquired or built quickly and inexpensively, and if there is confidence that it will receive wide acceptance and use, then the need for a slow, systematic, and totally integrated design approach is vastly reduced. Rather, software modules for a variety of planning and operations applications should proceed in a fast parallel-track effort. Several approaches to addressing the same problem can be developed concurrently. Tools of a more experimental nature can also be built and quickly distributed to a select group of operators for assessment. Software can be tested, modified, and tested again or later upgraded, replaced, or scrapped if appropriate--all of this at a cost much less than that of traditional software development. Ease of maintenance and modification will also allow planning tools to be tailored to the needs of individual properties that have operating and financial processes that are rarely identical.

Until recently, the skills required to use computers and terminals were significant; hardware and software design was not user-friendly. Now, since no special skills or training are necessary to use microcomputers and many minicomputers, no specialized personnel need be hired or regular personnel sent to training courses.

Another general reason for limited use of automation is the psychological barriers involved in using larger facilities, which traditionally were felt to be imposing and intimidating. Many subconscious fears have been identified in conjunction with computer use (5). Terminals connected to a remote site are associated with fear of the unknown. Users of terminals or batch-processing facilities experience a feeling of lack of control. They may worry about the possibility of causing program blow-ups or system failures. They may fear embarrassment at making errors or appearing ignorant. Many of these problems apply particularly to managers and executives.

The new small machines, particularly microcomputers, eliminate most of these difficulties. Judicious design of software and user interface relies on a comfortable dialogue, eliminates jargon and job-control language, and can result in a virtually error-proof system. The instant response, compactness, and proximity of microequipment lead to a user atmosphere of control and privacy.

Finally, the state of the art of computer hardware and software has been advancing so rapidly in recent years that it was difficult to plan and design a software system that would not be obsolete in some aspects when it was finished. Although computer advances are not likely to slow, the basic features of today's small machines (e.g., personal operation, range of purposes, layman skills, etc.) are not likely to change.

The general population is becoming more and more comfortable with the use of computers. Witness the tremendous growth in 24-h cash-dispensing machines. In part this is due to the increased exposure of students at all educational levels to computers and to the increased use of computers in diverse facets of our daily lives. These trends may be even more relevant to the transit industry, which has been experiencing a high rate of management turnover and looks toward a younger group of transit executives and line managers to take over.

WORK-STATION CONCEPT

The flexibility, convenience, and economy of microcomputers permit a variety of uses widely ranging in functional complexity, some of which would not be appropriate with larger machines. Simple although time-consuming tasks of a clerical or computational nature are not usually feasible to perform with mainframe computers even on an interactive basis, or with minicomputers of the desk-size variety, for several reasons. First, machines that have processing-time costs are perceived to be too expensive to use for simple tasks, even if it could be shown that the net costs of labor and machine are lower than that of a labor-only process. Machine costs are often considered out-of-pocket costs and allocated to different budgets than labor costs, which are usually considered to be sunk and to have little or no marginal value for a small task.

Second, start-up chores for minicomputers and larger machines in each and every application usually make it highly inefficient for simple needs. These chores often involve telephoning, log-on procedures, use of job-control language, etc. The time required to accomplish just the start-up procedures in many instances would exceed that for performing the work by hand. Third, prior to the advent of compact typewriter-size terminals, the inconvenient locations of access to large machines made their spontaneous use infeasible.

Aside from the more obvious applications of microcomputers for transit operations planning and policy needs [e.g., routing and scheduling aids, forecasting and cost-estimating tools, training aids, management information reports, etc. (which are discussed later in this paper)], it is other applications, perhaps considered somewhat mundane, that make the work-station concept so appealing. For example, microcomputers can be used as scratch pads and simple calculators, either in conjunction with more complex processing or in separate computations. Or they could be used to recall and review frequently needed documents, tables, or data such as work assignments or daily cash flows that would otherwise have to be stored and retrieved from hard-copy files. Notes from telephone calls can be typed in real-time or stored after the fact along with supplementary ideas. Draft memos and letters can be input to the machine and transmitted to a secretary's terminal for editing and final preparation on a word-processing package. Most of these applications can be initiated with the touch of a few buttons and completed even more simply.

The work-station concept involves the placement of a microcomputer or a microcomputer terminal on an employee's desk or proximate work table for routine use--sometimes continuous, sometimes intermittent--similar to the use pattern of a desk calculator. In many circumstances it will be appropriate for several staff members to have readily accessible terminals. For example, a junior clerk may be responsible for receiving, processing, and reporting ridership data by using a microcomputer data-base package. One of a service planner's major duties may be the design and evaluation of route-service changes by using stored information on schedules and patronage combined with a service-cost model. And a department head could access the latest budget projections from several divisions (e.g., maintenance, supplies, utilities), display the data in tabular or graphic form, and perform a variety of analyses including, for example, a comparison of the property's cost trends with those of other operations (by using stored Section 15 data) or perhaps forecasting future Section 15 cost and performance measures for planning purposes.

In fact, it can be seen that maximum benefits will accrue when several types of staff have access to and use microcomputers, thus enabling information and data at many levels of detail to be passed back and forth with little effort. Effects of changes in one area on another (e.g., revenue service on maintenance operations) can be observed--an important control capability for management. Data can be transferred either by linking the microcomputers to a larger central computer (which might be required to maintain a centralized data base) or by linking them directly to each other by using a communications network.

Further, the use of microcomputers expands, rather than precludes, the possibilities of using larger machines, data bases, and software applications. Microcomputers can be used as terminals to connect to a variety of computers that have features and capabilities far beyond those of the small machines. In particular, present users of the Urban Transportation Planning System and other existing automated planning tools will find that they can access and exercise these tools more easily with microcomputer terminals.

In summary, the benefits of a microcomputer work station are the ease of information organization and storage; the facility to communicate ideas, text, and data in a structured, yet effortless, manner; and, of course, the capabilities of an on-line computer always at one's fingertips.

OPS CONCEPT

The OPS concept is that of evolving systems of compatible automated tools for use directly by transit property staffs. Ideally, every staff member who would make substantial use of these tools would have a dedicated terminal, although this is not essential. The OPS will contain both management information functions (such as monitoring, data organizing, and reporting) and analytical techniques for accounting, planning, forecasting, and evaluation.

Although it is intended that the products be useful to properties of all sizes, the target user groups will be the small and medium-sized operators who perhaps have greater needs. Initially, some of these products may be automated versions of existing manual procedures, many of which are inefficient only because they are manual. Later research will result in new techniques that are designed to take maximum advantage of the computer and the integrated systems. As much as possible, OPS tools will be compatible with management information systems already in place. They will also take advantage of commercially available software.

Importantly, the OPS is also seen as a framework for the development of automated systems. Because the federal government does not want or intend to be the sole developers of transit operations tools, it will be highly supportive of private organizations that wish to invest their own capital to develop and sell their own related products and services. In order to encourage these efforts and to increase their overall usefulness, the OPS project will include the development of standards for creating automated tools (e.g., protocols, software languages, interface formats, data structures, recommendations for hardware and software configurations, and other guidance). These standards will improve the chances that independently developed products will be compatible with others and thus promote a wider market for their use.

SAMPLE APPLICATION

As discussed above, the work-station concept encompasses so many useful functions that it is hard to imagine a process or task that could not be carried out more efficiently with a microcomputer OPS. Here, however, we shall briefly describe an application that represents one of several prime uses of the system and would be performed on a fairly regular basis: Analysis of transit route data for purposes of route-performance evaluation and performance forecasting. The primary components of this application include (a) capture of route-level data, (b) data synthesis and route-performance reporting, and (c) forecasting of route performance under assumed conditions.

Data Capture

Data required for route analysis include those that describe both the demand for and supply of service on a route. Primary demand data include route ridership and revenue and rider characteristics. Major supply data include route operating cost and level of service (i.e., service frequency, accessibility, speed, and reliability). The focus in this example is on capture and use of route-level demand data.

Procedures for gathering and processing route ridership and revenue data vary across operators. Some record total farebox receipts on a daily basis and later allocate them to routes and derive passenger totals. At the other end of the spectrum, a few properties with automatic vehicle monitoring (AVM) can gather passenger and revenue data by location and time of day. Flexible data-entry capabilities can be provided at the work station to accommodate these variations in data-gathering procedures and data levels of detail.

For example, clerical or other staff will be permitted to enter via the work-station keyboard daily fare receipts on an hourly, per run, or other basis depending on typical practices or whether the data were to be used for a special study. Data-entry software can be structured to automatically accumulate receipts to a desired time period (e.g., weekly or monthly) to simplify routine updating and provide period-to-date information. Electronic worksheet methods are particularly well suited to this type of application. Additional features can be provided for entering other types of data (e.g., rider characteristics from surveys) that support related applications (e.g., revenue-to-ridership derivation).

Performance Evaluation

In this example, route revenue and ridership data sets that correspond to a particular time period would likely be stored in a microcomputer's secondary memory (i.e., on magnetic disc) or perhaps in the memory of a larger centralized computer that could be accessed by a microcomputer or a terminal. These data could usefully be synthesized with supply data (e.g., operating cost, vehicle miles) to produce route-performance indicators and reports. Capabilities could be available to the work-station user for defining or modifying performance indicators and report formats. In addition, various evaluation aids (e.g., flags for substandard performance) can be used to simplify information assimilation. Data-base management methods can be used for performing the data retrieval, manipulation, and storage operations concomitant to these applications.

For example, existing route ridership, revenue, and operating cost data sets can be synthesized to create the performance indicators revenue per dollar of operating cost and operating cost per passenger. Figure 1 shows a possible performance report as it would be displayed at a work station. The asterisks adjacent to certain numbers are flags for substandard performance (as defined by the user).

In addition to tabular reports, graphical-display capabilities will be provided as an integral part of the performance-evaluation subsystem. Standard bar chart, histogram, and data-plot displays can be used in performance comparison and statistical analysis (e.g., for market research purposes). Graphical displays can also be tailored to specific service-planning applications as, for example, the route passenger-loading profile shown in Figure 2. In addition, interactive graphics methods can be used to assist in more complex, spatially oriented design applications. Relevant service-planning applications include specification of zone fare structures and geographic placement of transit stops and routes.

Performance Forecasting

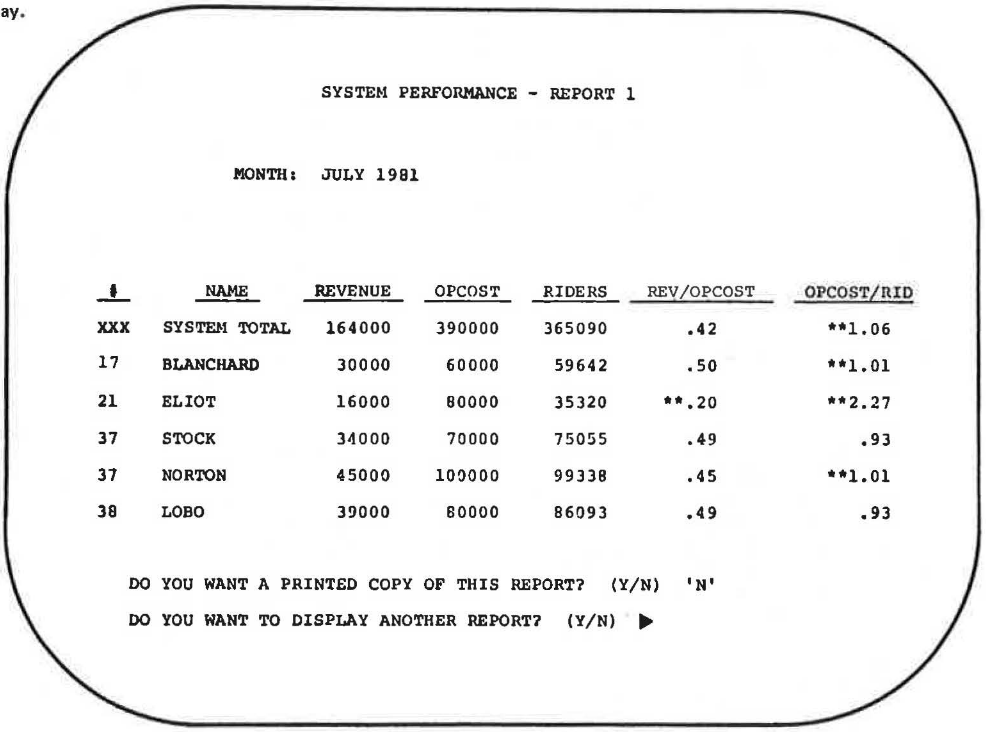
Computer-aided entry and analysis of data on existing service operations can help greatly in diagnosing problems and suggesting possible service improvements. However, to evaluate proposed service changes or answer "what if" questions requires forecasting of performance under various assumed conditions regarding level of service, service definition, and environmental factors.

Applications programs that use supply-and-demand models to forecast ridership, revenue, and cost performance are being developed for use on microcomputers (see paper by Turnquist, Meyburg, and Ritchie elsewhere in this Record; also see BUSMODEL from Colin Buchanan and Partners, 47 Prince Gate, London SW7 2QE, England). Such programs enable a transit analyst to specify (at the work station) changes in route or systemwide headways, fares, speeds, stop spacing, and alignment and to receive within seconds forecasts of resultant impacts on performance. Performance measures and reports used for evaluating existing service can be used for evaluating proposed new services or service alterations.

For example, service changes on a route with a substandard operating ratio can be tested with forecasted operating ratios reported for each attempted change to compare with the existing operating ratio. Other measures can also be reported to examine trade-offs in service performance along several dimensions (e.g., operating ratio versus ridership).

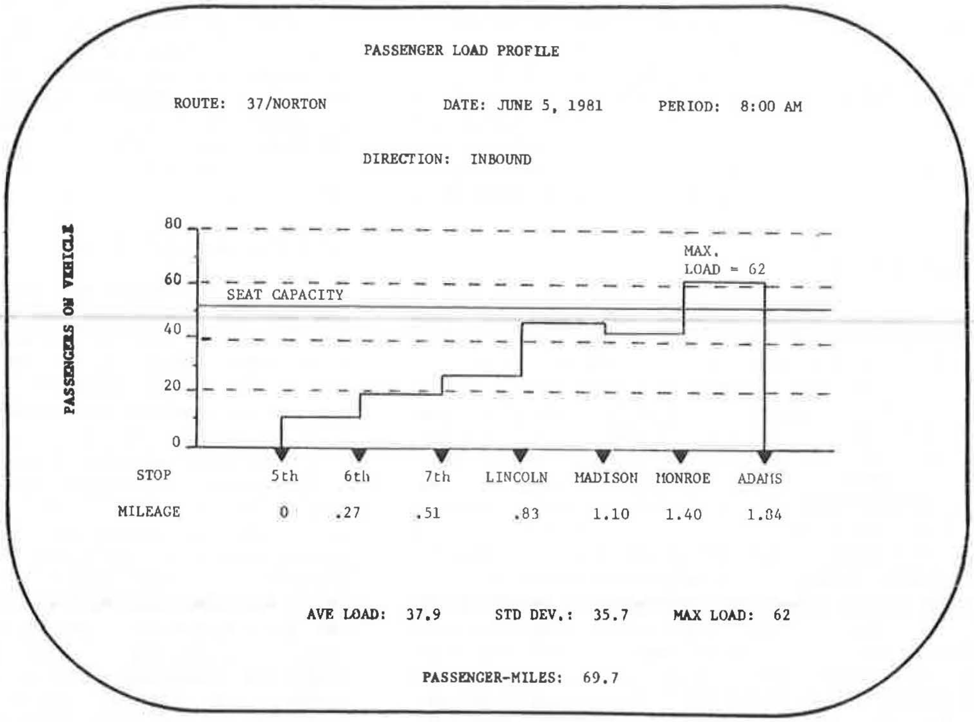
Numerous tests of alternative service or route

Figure 1. Performance report display.



```
                  SYSTEM PERFORMANCE - REPORT 1

          MONTH:  JULY 1981


 #       NAME       REVENUE   OPCOST   RIDERS   REV/OPCOST   OPCOST/RID

XXX   SYSTEM TOTAL   164000   390000   365090        .42        **1.06
17    BLANCHARD       30000    60000    59642        .50        **1.01
21    ELIOT           16000    80000    35320      **.20        **2.27
37    STOCK           34000    70000    75055        .49           .93
37    NORTON          45000   100000    99338        .45        **1.01
38    LOBO            39000    80000    86093        .49           .93


   DO YOU WANT A PRINTED COPY OF THIS REPORT?  (Y/N)  'N'
   DO YOU WANT TO DISPLAY ANOTHER REPORT?  (Y/N)  ▶
```

Figure 2. Load profile display.



specifications can be made quickly at the computer work station by using automated performance-forecasting methods, especially if the man-computer dialogues are well designed and simple. More thorough examination of potential service improvements by the transit planner or manager will be encouraged. Higher quality management decisions and greater service delivery efficiency will result.

OPS DESIGN ISSUES

The OPS will be, in essence, a set (or, more likely, sets) of tools that provides information to transit management. Like any tool, it must be used to be effective. Because its use will not be mandated, the OPS will be used only if the value of the information it provides exceeds the cost of obtaining it. The basic goal of the OPS design and development effort, then, is to produce tools that provide the best possible ratio of information quality to price, subject to the constraints of the development effort. This section investigates certain OPS design issues that will influence its ratio of quality to price.

Maximizing Information Quality

The quality of the information produced by an OPS is related to its usefulness to management and to its reliability. It must be recognized that different transit properties operate in different ways and information useful to one may be of little interest to another. For this reason it will be vital to design the OPS so as to permit customization. For example, while certain basic reports may be available to all transit properties (e.g., reports that support Section 15 reporting requirements), an OPS should permit each property to design its own ancillary reports. Similarly, different properties collect different operations data; thus, while certain data items may be required of all properties, an OPS should permit each property to add supplemental items as it sees fit. Each operator will have the opportunity to incorporate local innovations and tailor the system to his or her own capabilities. In general, an OPS should be designed to permit each property to mold the OPS information output to fit its particular information needs.

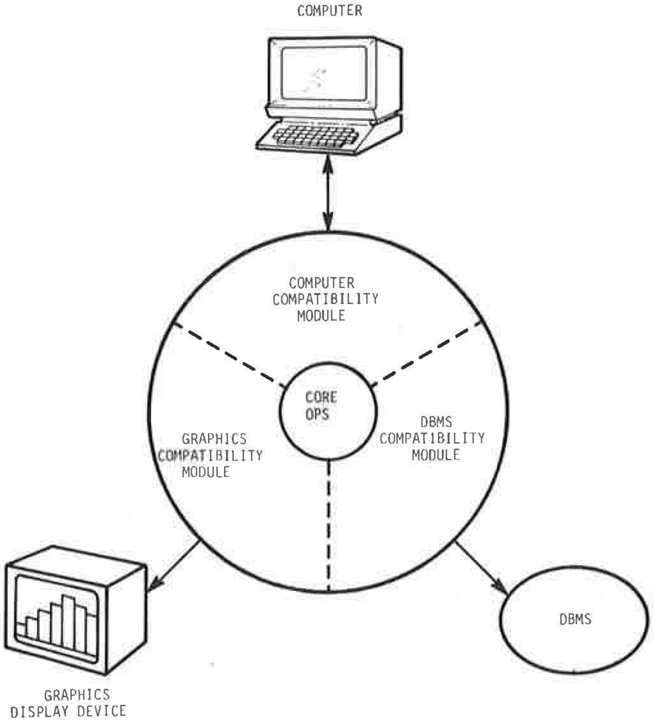
The other component of information quality--the reliability of the information produced--is related to both the quality of the methods that produce the information (e.g., the goodness of a particular forecasting technique) and the frequency of errors in the OPS. Although the OPS will be designed to minimize errors, not all errors will be found during development; thus, some mechanism must be developed to enable the OPS to be upgraded as necessary. In addition, since better forecasting techniques and new analytical capabilities will also be developed later, the OPS should be designed to incorporate such modifications gracefully by using procedures that require minimal specialized knowledge, and it should permit a typical user to accomplish any upgrade with no more than, say, an hour of labor.

Minimizing OPS Costs

The cost of creating and maintaining an OPS data base will be a major component of OPS costs. Many of the advances in data-processing technology in the past decade have been aimed at eliminating data redundancy (i.e., separate storage of the same data for different purposes). The OPS will of necessity use much information that has other uses. Revenue data and timetables are two simple examples. Entire software subsystems called data-base management systems (DBMS) have been developed to reduce data redundancy. The use of a DBMS can dramatically reduce total system data costs. It is likely that the OPS must use a DBMS if it is to be economically viable. The OPS will also require the use of a graphic-display device to produce visual displays of operational data.

The OPS is thus a system that will require the support of two pieces of specialized hardware (a computer and a graphic-display device) and one piece of specialized software (a DBMS). (Other pieces of

Figure 3. OPS compatibility modules.



hardware such as a printer, disk drives, and user terminals will be required as well. However, these items have achieved a degree of standardization that makes it likely that the OPS will be able to use whatever brands the operator selects. In short, compatibility with these items is not a major design issue.) All three items are available in numerous forms from numerous manufacturers, and the diversity is likely to increase in the future. Moreover, many transit properties already own some of these items and use them to perform tasks that complement the OPS functions (e.g., maintaining payroll and inventory records). The initial cost of an OPS can be minimized if it can use the support items already possessed. To permit this, however, the OPS will have to be designed to permit it to interface with a wide variety of computers, display devices, and DBMSs.

The OPS could best respond to this diversity if it were designed to employ compatibility modules. All of the analytical functions of the OPS would reside in a core module that would interact with the computer, the graphic-display device, and the DBMS only through the compatibility modules. This arrangement, depicted in Figure 3, would isolate the analytical capabilities of the OPS from the diversity of the environments in which it will operate. Every transit property would use the same core module, but each would use different compatibility modules to link the OPS to the particular computer-display device and DBMS in use on its particular computer system. Sets of modules could be merged together to form turnkey packages. The compatibility modules will also increase the ease of integrating software developed by private organizations and the transit properties themselves. This concept has been applied successfully to many other systems.

The operating cost of the OPS will be of little consequence to properties that can use the cheaper minicomputers and microcomputers. But even for those that choose to use OPS on large mainframe computers, careful design will minimize operating costs.

One cost factor that is normally a major concern for transit properties is personnel costs. These costs can be minimized by designing the OPS to be very easy to use. The design should strive to ensure that every minute of human interaction with the OPS will be spent productively. All clerical chores, such as keeping track of data, should be automated. Also, the OPS system must be easy to learn so as to minimize staff training time. Again, use of the compatibility-module concept will help reduce these costs by minimizing the new equipment with which the staff must become familiar.

One final cost is the hassle factor--the intangible cost of frustration. A system that frustrates its users, perhaps by surprising them with unexplained bugs or by requiring convoluted machinations to accomplish an essentially simple task, would not be worth using any more than would an expensive system. The hassle factor can be reduced only by emphasizing the human engineering aspects of the OPS design.

AN OPERATOR'S PLANNING SYSTEM

The U.S. Department of Transportation, through UMTA, is supportive of programs to supply the transit industry with technical aids and information to allow self-improvements in performance and efficiency. UMTA, in conjunction with the Transportation Systems Center, is generally following the concepts described in this paper. The development strategy includes a number of steps designed to provide and maintain consistency between operator requirements and capabilities and the technical design of an OPS. A primary element of the development strategy is transit operator involvement. This includes discussions at the outset with operators concerning their immediate needs, testing and experimentation by operators as each module is prepared, and continuous feedback from operators as the system begins to take shape. An advisory group of representatives from the industry has been formed to advise, review, and test through the development process. System-design standards related to technical design approach and system structure, coding conventions, formats, documentation, and hardware will be researched, including a study of where standards and design guidelines make sense. A clearinghouse for information on ongoing projects, experiments, and up-to-date planning methods and ideas will be set up.

It is hoped that these steps will result in a smooth and effective development process for a highly useful set of planning tools for the transit industry. These tools will include some developed by the federal government, but also those developed by private organizations and the transit properties themselves. When the system is complete (to the extent that most of the functional objectives have been met), it may exist in several variations that correspond to the idiosyncrasies of different transit operations. But each one will be the operator's system. It will need to be modified or upgraded only when the operator's future requirements change.

Today the transit industry is attempting to withstand severe shocks to its financial structure and its base of support. We hope that speculation that these changes will eventually bring the industry to a strengthened position prove to be true, and we trust that the OPS will greatly increase its chances.

ACKNOWLEDGMENT

This paper was written as part of the Planning Methodology Development Support Project of the Transportation System Center under sponsorship of the Office of Methods and Support, UMTA. The views expressed are ours and not necessarily those of the U.S. Department of Transportation.

REFERENCES

1. M. Couture, R. Waksman, and R. Albright. A Preliminary Analysis of the Requirements for a Transit Operations Planning System. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Staff Study SS-24-U.3-159, Oct. 1978.
2. G. Paules. The Potential for Automated Data Systems in Public Transportation. IEEE Conference Paper, Dec. 1978.
3. M. Couture and G. Paules. Automation in Public Transit Operation and Management: Trends and Prospects. TRB, Transportation Research Record 761, 1980, pp. 51-53.
4. National Urban Mass Transportation Statistics. U.S. Department of Transportation, Rept. UMTA-MA-06-0107-81-1, May 1981.
5. J. Martin. Design of Man-Computer Dialogues. Prentice-Hall, Englewood Cliffs, NJ, 1973.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Automatic Vehicle Monitoring: Effective Technique for Transit System Management and Control

RICHARD W. LYLES AND MAURICE H. LANMAN III

In the context of searching for new approaches for efficient transportation system management and utilization, the Urban Mass Transportation Administration (UMTA) has funded a comprehensive demonstration of an automatic vehicle monitoring (AVM) system in Los Angeles. AVM coverage includes approximately 10 percent of the Southern California Rapid Transit District's (SCRTD) route miles and buses. The system is now operational, the AVM capabilities having been phased in over a year's time. Although the evaluation program on the part of UMTA and SCRTD continues, analysis of the impacts to date shows that benefits have accrued in several areas of transit system operations, including route scheduling and information management, improvement of day-to-day system reliability, rendezvous of scheduled and nonscheduled vehicles, and response to emergency situations.

The cost of providing public transportation service continues to increase due to the high price of energy, other increasing operation and maintenance costs, and the rising costs of building new system elements and/or replacing rolling stock. Thus, operators of public transportation systems, as well as federal, state, and local officials, are looking in earnest at techniques that enable better use of existing systems, and especially the use of buses that have the flexibility to accommodate geographically shifting passenger demand. Questions arise as to how the nation's bus fleets can be used more efficiently and effectively. One approach that is receiving increasing attention is the use of automatic vehicle monitoring (AVM).

In the above context, the Urban Mass Transportation Administration (UMTA) funded an AVM demonstration project in Los Angeles with the cooperation and participation of the Southern California Rapid Transit District (SCRTD). The basic purpose of the demonstration was to enable the evaluation of the effectiveness and efficiency achieved in bus system operation through use of the real-time monitoring and control capabilities of a fully operational AVM system. The project represents the first such comprehensive AVM implementation in the United States. The system was developed and installed by AVM Systems, Inc. (formerly a division of Gould, Inc.) and was operational in spring 1980. The Transportation Systems Center, UMTA, served as system manager for the project.

HISTORICAL DEVELOPMENT OF AVM

AVM is not a completely new concept (1), having been used in one form or another as early as 1935 in Chicago to check on streetcars and for buses in the 1940s. Information on headways was being collected automatically by 1955 in Pittsburgh, St. Louis, and Philadelphia, and optical scanning was being used in London in 1958 (2). However, in these early attempts, vehicles were not necessarily explicitly identified nor was there any attempt at real-time control by using the information that was obtained.

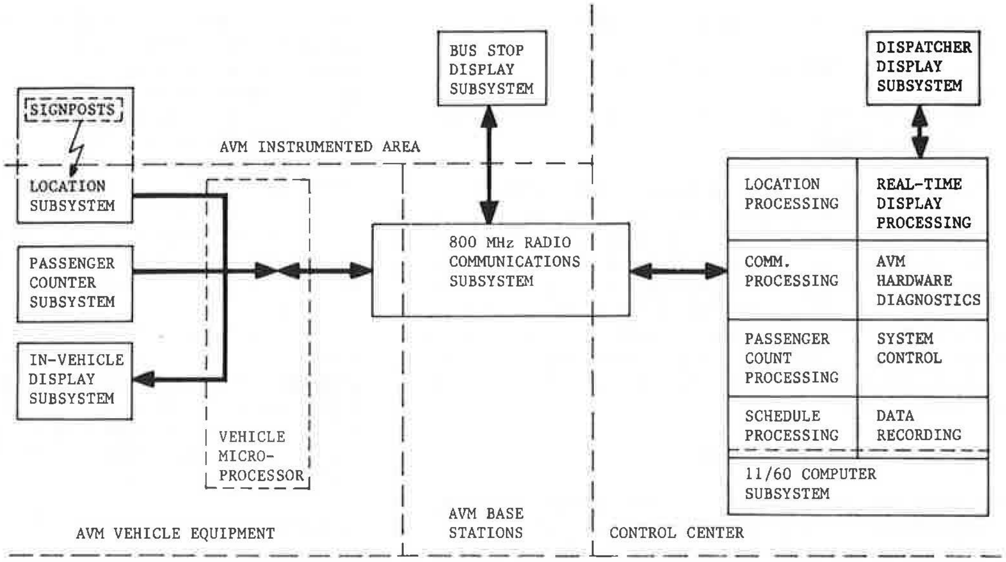
Currently, AVM is considered to be directed to real-time monitoring of vehicles (e.g., location) with the potential for exercising control as opposed to only identification. The basic components of the modern AVM system include the following (3, p. 202):

1. Vehicle locating and status monitoring devices,
2. Communications system, and
3. Central control facility.

Although AVM has demonstrated application in any operation that involves the coordination of fleets of vehicles (e.g., taxis, police cars), several recent experiments with transit operations are of immediate interest. Lukes and Shea (2) and Miller and Basham (4) describe an experiment in the early 1970s that focused on citywide AVM coverage for a small percentage of the Chicago Transit Authority's (CTA) rolling stock. The AVM-equipped buses were those providing "owl" service. Although problems with equipment and an apparent lack of execution of the control potential by CTA dispatchers hindered the experiment, conclusions were that use of AVM resulted in schedule and headway adherence that was at least as good as that achieved with manual control that, according to the authors, would have resulted in sufficient manpower savings to economically justify the system.

Bevilacqua and others (5) describe a more recent demonstration--the General Motors' transit information system (TIS) in cooperation with Cincinnati's Queen City Metro transit system. That system, while

Figure 1. System block diagram.



not an AVM system per se, produced data on passenger boardings, alightings, and bus travel times that were available on a bus trip or route-segment basis. Locations were established through the use of radio signposts and bus odometer readings. The principal difference between TIS and AVM is that information generated by the former is for off-line use (e.g., for trip scheduling and route planning) whereas AVM information is useful for real-time control as well. The authors concluded that TIS, if properly functioning, does appear to be economically viable for systemwide use in Cincinnati.

Another effort of note is being undertaken in Toronto, Ontario, Canada (6). The system installed on the Toronto Transit Commission property provides a continuous flow of information on vehicle location, schedule adherence, passenger loadings, and emergency status. In addition, options for further development include the capability to connect into the traffic-control system to provide priority signal phasing for transit vehicles and for voice response to passenger phone-in queries for current schedule and service information. The benefits realized in Toronto after a year of operation (1979-1980) included the identification of a 5-10 percent oversupply of buses on AVM-equipped routes and a 5-10 percent savings in operating costs. In addition, the off-line uses of data for scheduling and route planning were expected to make overall use of equipment more efficient as well as more responsive to the demand for service.

LOS ANGELES DEMONSTRATION PROJECT

The demonstration in Los Angeles is the second of a two-phase effort funded by UMTA. The first phase (7) was undertaken in Philadelphia during 1976-1977 and was concerned with identifying the subsystem for vehicle location to be used in phase 2. The original purposes of the second phase were identified by Symes (8, p. 236):

1. Conduct a thorough test and evaluation of a fully functional, area coverage, multiuser AVM system;
2. Quantify the benefits to transit and other users;
3. Advance the state of the art of AVM; and
4. Establish technical and economical bases for future deployments.

The demonstration is not, however, systemwide--4 of 214 routes were instrumented and about 200 of 2600 buses--overall about 10 percent of route miles and buses. The demonstration was, however, comprehensive as far as the four lines and an adjacent 54-mile^2 (random-route) area.

Key components of the Los Angeles system are the equipment on board each instrumented vehicle, the signposts for the location subsystem, and the central control and transmission facilities. The basic relations between all components of the system are shown in Figure 1.

Signposts

The signposts for the location system are small, low-power radio transmitters mounted on existing utility or lighting standards and located at intervals of approximately 900 ft along each of the four specified routes and at somewhat larger intervals over a 54-mile^2 area for control of random-route vehicles and ascertaining the locations of off-route vehicles. On-route buses can be typically located within 300 ft by extrapolation, based on the signal strength of two adjacent and sequential signposts, and within 500 ft in the random-route area.

In-Vehicle Equipment

The on-board equipment for the buses consists of the following:

1. Pressure-activated passenger-counter devices in the bus stairs,
2. Electronics for information storage and transmission,
3. An antenna mounted on the top of the bus, and
4. An in-vehicle display (IVD) for the driver.

Although the first three items are relatively self-explanatory as to purpose, the last requires further explanation. The main components of the IVD are a message status panel, a schedule indicator, and a system clock. The message status panel consists of 10 message lights that are lighted when appropriate and are otherwise blacked out. The message panel serves three functions:

1. To convey an automatic start message from the system computer to the driver when it is time for

him or her to pull out from a division garage or layover point,

2. To convey messages to the driver from the dispatcher, and
3. To inform the driver of the status of pending radio communications.

The dispatcher-to-driver messages indicated in 2 above are essentially responses to tactical (problem) situations identified by the dispatcher or automatically by the system. The available messages include such things as "observe schedule", which means observe the schedule indicator, and "schedule adjust", which indicates that a schedule adjustment, relative to the trip sheet, is in effect. The IVDs were only operational on buses on one of the lines as of summer 1981.

Central Control and Transmission

The central control element of the system is where the significant departures from typical transit operations occur. Although the existing SCRTD system was already equipped with two-way radios, had a covert crime alarm for the driver's use, and was computer assisted, location estimation for buses was still a manual procedure. With the availability of comprehensive real-time operational information, the control of daily operations is largely shifted to the dispatcher. The extent of the differences between operations with and without AVM is implicitly defined in the description of the dispatcher's control console and the AVM capabilities at his or her command. The dispatcher's control center has the following elements: two cathode-ray tube (CRT) displays for real-time monitoring of operations on individual transit routes and for calling up displays such as location and status for specified buses, listing of buses that share a specified problem (e.g., all those buses on a specified line that are operating behind schedule), roll-outs from a specified division, and so forth; a 45-function keyboard for controlling the displays; a standard typewriter keyboard; and the existing vehicle identification and voice communication equipment already being used by SCRTD.

Colorgraphics Display

The first of the two CRTs is the colorgraphics, and it has the following components: the display identification, an emergencies list, a tactical-situation list, and the graphics area.

The display identification specifies the transit-route or random-route map being displayed, the scale of the display (e.g., one of three levels of detail available for transit-route maps), and the system time. The emergencies list shows, by bus and run numbers, all AVM buses that currently have an active silent alarm.

The tactical-situation list consists of a set of 10 situations that are automatically identified by the system for any AVM bus on an AVM line without dispatcher intervention. If one or more buses fall into a tactical-situation category, then that category name is lighted on the display. Tactical situations include such things as very late, off-route, and not at layover, all arranged in a priority order.

In the graphics area, displays of various maps are actually seen. Each of the four instrumented routes can be displayed here at full, one-fifth, or one-tenth scale. Buses are indicated at each scale and each is identified and color coded. The identification includes type of bus (e.g., accessible), indicated by a symbol; branch and destination code; service class; run number; and passenger count or schedule deviation. The color codes are red, emergency; blue, early; green, on-time; yellow, late; purple, very late; and white, a non-AVM bus. Individual bus positions on the displays are updated approximately every 40 s by automatic radio polling of each bus or every 10 s when a emergency has been declared.

The transit-route maps are linearized versions of the actual routes with cross streets shown for reference. On these maps all buses on the route can be shown or only specified groups (e.g., only late buses).

In addition to the transit-route maps, random-route maps with actual street patterns can be displayed. On these maps random-route vehicles (e.g., supervisors' vehicles) can be located and a specified line can be overlaid or a specific bus located. The latter capability is especially useful in locating off-route buses.

Alphanumeric Display

The other CRT screen is the alphanumeric display, and it also has four sections: a 2-line work area, a 4-line bus data area, a 37-line general data area, and a 2-line communications area.

The work area is where the dispatcher "talks" to the system computer. For example, one of the function keys allows the dispatcher to get all currently available data on a specified bus. The dispatcher would press a function key labeled GET BUS DATA; the system would respond with a prompt (question) in the work area for the line, run, and bus numbers; the dispatcher would type in the appropriate response and transmit it to the system; and the system would display the data on the specified bus in the bus data area. All such information exchanges are carried out in the work area, although the actual result is displayed in one of several locations.

The bus data area is where all current information on a single specified bus can be displayed. This information includes line, run, and bus numbers; current operating status; passenger count; schedule deviation in minutes; schedule adjustment in effect, if any; schedule deviation at last time point; whether the schedule indicator in the bus is on or off; tactical information; whether a problem currently exists; the identification of the console that responded to the problem; and schedule information including current location and time, next layover location and estimated time of arrival, and scheduled departure time from the next layover. The fourth line of the bus data area can be accessed by the dispatcher and is for comments. Once the bus data is displayed, it can be updated by using another key that causes a special polling of that bus.

The general data area is the largest portion of the alphanumeric display and has several uses, such as listing schedule information about individual or specified groups of buses, listing all buses in a specified group (e.g., all late buses), listing schedule information for a bus or group of buses, and providing instructions or information about the use of the system or explicit function keys. Schedule information that can be called into the general data area includes a list of the next 20 starts from a specified location and on a specific line, a display of the complete schedule of a specified bus that shows all time points, and a schedule block for a given line, direction, and starting time.

Buses are grouped in numerous ways by the system and a listing of those buses can also be reviewed in the general data area. There are five different classes of groupings: service class (e.g., local), passenger load (e.g., overloaded), tactical situation, status (e.g., unassigned or out of division), and a miscellaneous class that includes such buses as those that are non-AVM but on AVM lines or those with active tactics in effect. In all, there are 31 different groups that can be specified. As indicated earlier, the system can also give the dispatcher assistance, e.g., the list of the 31 possible groups can be called up for review.

Dispatcher Capabilities and Other System Aspects

Although there are other functions that could be discussed, the basic types of AVM monitoring and display capabilities available for the dispatcher have been outlined. The dispatcher can (a) monitor the operation of all buses or a specified group on any given AVM line at one of several scales, (b) call up real-time information on any AVM-equipped bus and locate the bus, (c) review the available information on a specified group of buses, and (d) make out trouble reports. In addition, there is a specialized procedure for dealing more efficiently with active crime alarms (emergency situations).

Thus, the dispatcher has a comprehensive view of ongoing transit operations on the AVM lines, i.e., more comprehensive than could possibly be obtained in the past. Given that information, a significant opportunity exists for exercising real-time control over transit operations.

Before reviewing the effects of AVM to date, several other aspects of the system should be noted. First, there were additional capabilities in the system that were not fully operationalized during the demonstration. Foremost among these were the tactical keys on the dispatcher's control console. These keys would have afforded the dispatcher the opportunity to send specific drivers (or groups of drivers) nonverbal tactical messages via the message status panel on board each bus. As actually implemented, the dispatchers were encouraged to give such messages via the normal communications channel.

Another aspect of the system that was not actually implemented in the field was the bus stop display for use by waiting passengers. Although a small number of displays were operational, they were never actually used on any of the lines. Information from this display included time of arrival of the next bus, branch or destination of that bus, type of service offered, and time of day.

In addition to the real-time aspects of AVM that have just been reviewed, the system also provides data tapes of raw data for experimentation and extensive summary files for use in scheduling, route planning, and other information management functions. These latter products are not trivial and possibly provide enough benefits in and of themselves to justify system costs.

EXPERIMENTATION AND EVALUATION PROGRAM

Benefits to be derived from AVM capabilities fall into several categories and are accrued only over a period of time. Major categories include the immediate payoffs associated with the improved response to emergency situations, somewhat longer-term advantages in reworking schedules and other information management areas, and future beneficial shifts in demand for service in response to increased system operating efficiency, dependability, and reliability. Thus, the potential impacts and benefits of the AVM system being demonstrated in Los Angeles are not all measureable in the short term. For example, considering demand sensitivity, it is quite likely that if transit system operations (insofar as the AVM lines are concerned) improve in dependability and reliability there will be resultant changes in patronage. Such incremental changes will be difficult to isolate and measure on a systemwide basis, let alone for isolated lines, considering the impact of seasonal variations and the impacts of other exogenous factors (e.g., What is the impact of continuing energy shortages and/or high prices on transit patronage and how is it separated from the positive impacts of system improvements?).

The remainder of this discussion is concerned with the shorter-term effects of AVM in three principal areas: (a) scheduling changes as a result of AVM, (b) impacts on day-to-day system dependability and reliability in terms of schedule and running-time deviations, and (c) response to emergency situations that involve driver-activated crime alarms and the rendezvous of scheduled and nonscheduled vehicles.

Evaluation Approach and Problems

The basic approach to the evaluation, insofar as day-to-day system dependability and reliability were concerned, was to collect data on schedule deviation, running-time deviation, and several other variables (automatically) as AVM capabilities and control potential were phased in, which provided the basis for a rough before-and-after AVM comparison as well as a comparison with a control line that, although monitored by AVM, was not subject to real-time intervention. The data were typically aggregated into 31 two-week test periods in three different phases. Phase 1 data were collected by using AVM capabilities, although no system control was exercised. Phase 2 data were collected as the dispatchers were introduced to the system and its capabilities, although minimal real-time control was exercised. Phase 3 and 3A data were collected during the final test periods when the dispatchers had been trained and familiarized with the system and were being encouraged to exercise real-time control over system operations. As indicated, the phased-in AVM capability was undertaken on only three of four lines, the fourth being retained as an experimental control in an attempt to more accurately track seasonal and any other normal background variations in service levels.

In addition, the 62-week duration allowed for a moderate year-to-year comparison to be made (i.e, the first 10 weeks with the last 10). This sort of comparison was useful insofar as variations due to seasonal fluctuations in service and demand were minimized.

The ability to isolate and quantify specific improvements in ongoing system operations, as indicated above, proved to be more difficult than originally anticipated and stems from several sources. First, the impact of exogenous factors made it difficult to isolate the effects of real-time AVM system control; e.g., How are the AVM effects accurately separated from acknowledged seasonal fluctuations both in the demand for service and the ambient traffic congestion?

Second, and perhaps more importantly, there was some reluctance on the part of some SCRTD dispatchers to become actively involved in exercising the full extent of the available control capabilities; some were quite interested and adept at using the system and some were not. These differing attitudes led, in turn, to inconsistent application of AVM capabilities and little use of the more comprehensive forms of intervention (e.g., shifting buses to a problem area).

Third, during several test periods there were problems with system operation due either to system malfunction or other interruption in system communication. Even though such breakdowns were often not a problem with the system hardware or software per se (i.e., interruptions also came about as a result of experiments, software modifications, and other activities), they nonetheless reinforced any negative feelings dispatchers might have had about the system.

Problems of the last two types described above became less troublesome during the final weeks of data collection as the dispatchers became more accustomed to the system and other interruptions became less frequent. These improvements notwithstanding, the effects of AVM (either positive or negative) were difficult to isolate except when the effects of exogenous factors and the degree of dispatcher intervention could be accurately identified. (This was the case during the evaluation of the timeliness of the response to a crime alarm and in accomplishing the rendezvous of scheduled and nonscheduled vehicles.)

Scheduling Changes as Result of AVM

Given that only four SCRTD routes were subject to AVM control and that one of those was reserved for experimental control, the opportunities to make significant changes in schedules were somewhat limited. However, the information from the first two phases of AVM data collection did indicate that, for one line in particular, overall running times were consistently longer than scheduled. Thus, in the third phase running times were lengthened in one direction with a resultant improvement in the running-time deviation.

This sort of schedule improvement is important not only to the drivers who no longer have the frustration of always running late but also to passengers who can have more confidence in the schedule. The identification of such a problem in the schedule also highlights the comparative advantages of AVM data collection and information management. In order to review such data with the old system, considerable manual effort would have been required to collect it in the first place (e.g., by using mobile supervisors) and then additional effort for the processing. The AVM system, by way of contrast, produces such summary data as an element of normal system operation, thus making both the identification and solution of such problems straightforward and routine.

In late 1981, problems in transferring information from the AVM computer system to the machine used by SCRTD were overcome, and other summary information was being processed that had direct bearing on off-line scheduling and management of the system.

Increased System Reliability and Dependability

As indicated earlier, data were grouped into two-week test periods and by level of AVM control available. Other stratifications of the collected data included time period during the day (e.g., morning and evening rush hours), direction on the line, and type of service (local or limited). Furthermore, evaluation data were limited to those collected on weekdays, during good weather, and within identified segments of each line. Typically, measurements (e.g., of schedule deviation) were made at four time points along the segments (i.e., near the beginning, two in the middle of the segment, and the last near the end).

Principal dependent variables that were examined were schedule deviation and running-time variation. Total passenger loadings were also examined for correlation between service provided and demand and to check for seasonal fluctuations. Available statistics for data from a test period (for data from one segment in one direction and for one time period) included the distribution (11 cells) of all observations, the mean of those observations, the standard deviation, and the sample size.

The basic hypothesis that was tested was whether the availability (and presumed use) of AVM control capabilities had a positive effect on the service-related variables that were measured. Comparisons were made both for a given line (e.g., Was service better in phase 3 than it was in phase 1 for a given line, direction, or time-of-day combination?) and relative to what occurred on the control line.

The results, in general, indicated that the effects of using AVM real-time control capability were positive, although reliable estimates of the magnitudes of those effects could not be obtained. It had originally been expected that a relatively reliable estimate of the actual magnitude of the impact of using the AVM system would be obtained by comparing operation and performance on the various lines in both before-and-after (the exercise of AVM control) modes and by making comparisons between each line and the control line not subject to intervention by AVM-assisted dispatchers. Unfortunately, the fluctuations and inconsistencies in the performance on the control line over the 62-week period were such that straightforward use of the data obtained from this line for estimating the magnitude of background variation in performance and in rigorous use of the line as a normal baseline condition were rendered impossible. The data were, however, useful for describing general trends in performance.

Similarly, while there were some predictable variations in the statistics of the performance variables for the lines, other variations were inconsistent; e.g., while the schedule deviation at a point varied more toward the end of any line than at the start, statistical measures of schedule deviation were not consistently related to total passenger boardings at the same point. Again, the net result was to make estimation of the actual magnitudes of AVM effects unobtainable.

However, in spite of the problems outlined above, the overall trend of the findings supported an assertion that AVM system control did have a positive impact. Basically, each line except the control was subjected to three increasing levels of AVM control, and data were also collected during an initial configuration when no control capability was available. Statistical measures of performance were then compared for each sequential pair of configurations and between two 10-week periods during the initial and last configurations (the year-to-year comparison).

The results of the between-configuration and year-to-year comparisons showed that, in general, service performance on the AVM lines was more likely to improve than degrade over time with increasing AVM control, as compared with the control line where performance was more likely to degrade over time (and without the exercise of AVM control). The important year-to-year comparison showed that 62 percent of the statistical measures considered showed improvement in the second year for lines where AVM control was exercised, whereas on the control line only 38 percent of the measures improved. In addition, for the control line the best performance occurred during either year 1 or configuration 1 approximately 67 percent of the time, whereas for the lines subject to AVM control only 30 percent of the best performances occurred during those periods (indicating again that the best performance occurred during a period when AVM control was being exercised).

Table 1. Trends in performance.

Line	Configurations 1-2						Configurations 2-3						Configurations 3-3A						Year 1-2						Incident of Configuration 1 or Year 1 Lowest Overall (%)
	Improve		Degrade		Even		Improve		Degrade		Even		Improve		Degrade		Even		Improve		Degrade		Even		
	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	
44[a]	3	7	39	87	3	7	27	60	16	36	2	4	17	38	25	56	3	7	17	38	27	60	1	2	67
41	10	50	10	50	0	0	9	45	9	45	2	10	8	40	7	35	5	25	12	60	8	40	0	0	35
89	6	25	13	54	5	21	7	29	15	63	2	8	16	67	5	21	3	13	7	29	11	46	6	25	58
83[b]	14	58	8	33	2	8	14	58	7	29	3	13	8	33	15	63	1	4	20	83	1	4	3	13	8
83[c]	11	50	8	36	3	14	12	55	8	36	2	9	15	68	6	27	1	5	17	77	3	14	2	9	18
All AVM[d]	41	46	39	43	10	11	42	47	39	43	9	10	47	52	33	37	10	11	56	62	23	26	11	12	30

Note: This is a summary that indicates general trends in statistical measures (e.g., standard deviation) of the performance variables (e.g., schedule deviation). Column entries show the actual number and percentage of measures improving, degrading, or where there was no change.

[a] Control line. [b] Local line. [c] Limited line. [d] Except line 44.

The statistical comparisons (e.g., Did the mean or standard deviation of schedule deviation vary significantly over the several AVM configurations?) were typically not particularly enlightening relative to making an estimate of the magnitude of the AVM effects. For example, in some instances a statistically significant change might be noted although operationally the actual value was quite small (e.g., What is the operational significance of 10-15 s of improvement on a run time of half an hour?). Similarly, while schedule deviation might improve on two different lines, the magnitude of the improvement might vary substantially between the two. For these and similar reasons, the overall AVM effects relative to performance variables were reviewed from a more qualitative point of view: Were the general trends on the AVM lines showing improvement or degradation of service, especially as compared to what was happening on the control line?

Table 1 provides a summary of the qualitative review of conditions on the lines. In general, the statistics (mean, standard deviation, and number of observations occurring in the extreme tail of the distribution) for the performance variables were examined on each line at several locations, and an indication was noted of whether there was improvement from one AVM configuration to the next and in the year-to-year comparison. In addition, a notation was made as to when (over the four configurations and of the two 10-week periods) the best performance occurred.

The entries in the table indicate how many of the statistics were improving, degrading, and where there was no change. The percent figures represent the appropriate percentage of the total number of measures considered in each instance. The far right column shows the percentage of instances when either year 1 or configuration 1 had the lowest value for all of the configurations or years. Thus, Table 1 facilitates direct comparison both among the lines and between each line and the control and provides a good overall picture of what occurred, in general, on the lines as they were subjected to increasing levels of AVM control.

The last line in the table provides an overall indication of how all lines subject to AVM control (lines 41, 89, and 83) compare as a group with line 44, where AVM was never used by the dispatchers. It can be seen that the general improvement was greater than that expected on the basis of what happened on the control line. Perhaps the most telling statistics were the year-to-year comparisons and the indication of when the best performance occurred early in the overall period--fully 62 percent of the measures showed improvement (year to year) on the AVM-assisted lines compared with only 38 percent on the control line, and the best performance was about twice as likely to occur during a period of AVM control.

Based on the examination of each of the transit lines individually as well as the overall trends, it can be concluded that AVM-assisted control capability does, in fact, represent a positive impact on the performance of the lines that were monitored.

It should again be emphasized that an estimate of the actual magnitude of the impact was impossible to make, given the available data and the level of analysis undertaken. Although the overall trend seems clear, the differential impacts from measure to measure and from line to line varied a great deal.

Response to Emergency Situations and Other Vehicle Rendezvous

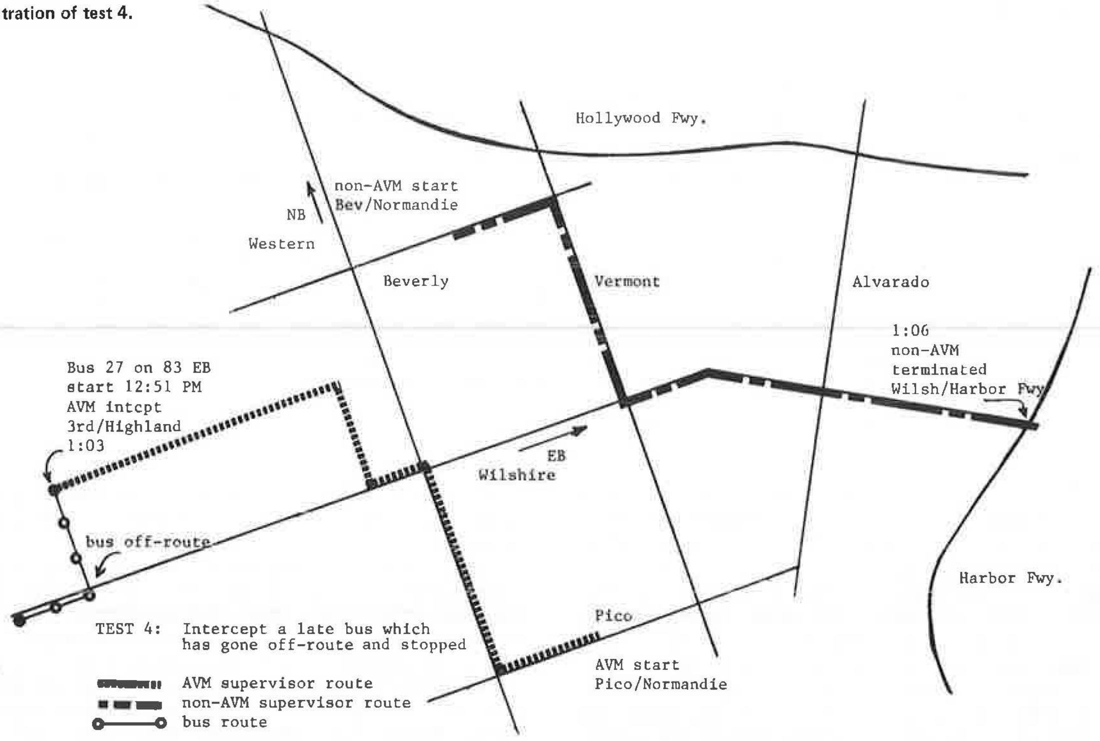
In contrast to the preceding discussion, the impacts of AVM control capability were quite significant and positive when the response to emergency situations and rendezvous of scheduled and nonscheduled vehicles were examined. In this latter examination, the situations were explicitly identified and the effects of background variation minimized.

Two separate types of operations were examined in five common experiment scenarios. The first type was responding to a driver-activated crime alarm [which is used when the driver feels that assistance is required (e.g., a robbery, a rowdy passenger)]. With the existing SCRTD system, the alarm is handled by a separate dispatcher who can automatically identify the bus but has to manually estimate its position based on the printed trip sheets. The dispatcher then identifies the nearest assistance (e.g., mobile supervisor) by using his or her radio and sends it to the bus. With the AVM system, the bus is automatically identified and its position can be determined by using the central control displays. By using another display, the nearest supervisor can also be identified and sent to the effected bus. When the alarm comes in, the system polling rate is increased so that movement of the involved bus can be more closely monitored.

The second type of situation was more general, which involves the rendezvous of a scheduled vehicle and an unscheduled one. Need for such a rendezvous might occur in several instances, such as a mobile supervisor taking material to a scheduled bus, replacing one bus on line with another, or affecting a transfer of passengers between a neighborhood-based demand-responsive van and a regularly scheduled line-haul bus.

In both types of situation, the question is how timely can the rendezvous be made. In the experiment scenarios, two supervisors starting from ap-

Figure 2. Illustration of test 4.



proximately equal distances from the bus to be intercepted were guided to interception, one via existing SCRTD procedures (i.e., manual approximation of the scheduled location of the bus) and one by using AVM capabilities. The supervisors had neither visual or radio contact with each other nor with the bus to be intercepted. The bus was temporarily assigned a real bus run number in each of the five scenarios.

The basic experiment procedure was common to all five scenarios and was as follows:

1. The two supervisors went to assigned locations;
2. The test bus was temporarily assigned a real bus run number and started from a previously assigned location;
3. After starting, the crime alarm on the test bus was activated and then was operated according to a previously defined script;
4. A non-AVM-assisted dispatcher guided one supervisor to the test bus while the AVM-assisted dispatcher guided the other; and
5. Each supervisor was tracked to the eventual point of interception (or test termination).

The five scenarios ranged from intercepting a scheduled bus moving on-route and on-time to intercepting a scheduled bus that was moving off-route, having been on-time when it left the route.

The comparison of the total response times and the search patterns of each supervisor illustrated the basic differences between the existing system and AVM capabilities. In summary, the five scenarios showed the following:

1. The supervisor receiving AVM information was always able to move more or less directly to the test bus without backtracking or making a false start in the wrong direction, which was not the case for the other supervisor [although the latter typically made the correct decision regarding where to attempt to intercept the bus based on available (manually obtained) information];
2. The directness of the routes taken by the AVM-assisted supervisor to intercept the test bus resulted in substantially lower response times (at a minimum, 30 percent less);
3. In two of the scenarios involving off-route buses, it appeared that the non-AVM-assisted supervisor might never make an interception--the supervisor was quite far from the bus with no indication of getting any closer; and
4. AVM-assisted response time did not seriously degrade even when the interception points were off-route.

Figure 2 illustrates one of the worst-case situations where the test bus was moving on-route but late when the crime alarm was activated, and then went off-route a short distance and stopped. The AVM-assisted supervisor received information that the bus was late as well as its location, then that the bus was off-route and where, then that it was stopped at a specific location, and intercepted it in about 12 min overall. The non-AVM-assisted supervisor received only the initial crime alarm report and standard time-point information from the schedule and proceeded to try to intercept the bus where it would have been had it been more or less on time. When the test was terminated, the latter supervisor was moving away from the test bus. This test, although admittedly a worst-case type of situation, is not unrealistic and serves to illustrate that with the current system, in all but the simplest instances, needed assistance can be searching blindly within a large area for a bus (and driver) in trouble.

SUMMARY

Although the evaluation of the AVM system being demonstrated and used in Los Angeles continues, the analysis discussed here has shown that there are considerable benefits to be accrued in several areas of ongoing transit system operations, including route scheduling and information management, improvement of day-to-day system reliability, rendezvous of scheduled and nonscheduled vehicles, and response to emergency situations. Although the day-to-day service improvements were quite difficult to separate from normal background variation, it seems clear that, at an absolute minimum, those operations did not degrade with AVM in place and positive results were obtained in several instances.

No attempt was made to place an absolute dollar value on the benefits noted, as several aspects are quite qualitative and their consideration beyond the scope of this evaluation; e.g., quantifying the value of increased rider confidence in on-time, or at least more predictable, transit performance. Although some of the benefits may be partly quantifiable in the long term (e.g., if increased confidence leads to increased patronage of the system), such data are not currently available or are subject to substantial error in approximation.

It is anticipated that future analysis will provide a more complete picture of the impact of day-to-day improvements that might be expected through exercising AVM control. For example, during the latter stages of the data-collection effort, dispatcher use of system capabilities became more consistent. Thus, subsequent analysis and comparison of results during this time period with those obtained more than a year ago will have the advantage of both the consistent use of AVM capabilities by the dispatchers and the opportunity to ignore some of the temporal (seasonal) variations.

It is also anticipated that as the SCRTD staff becomes more familiar with the types of data available from the system for route scheduling and so forth, greater advantage will be taken of those opportunities in a straightforward fashion (i.e., better scheduling for the instrumented lines) as well as in using the four lines for testing new strategies for controlling day-to-day operations that might be transferable to non-AVM lines.

REFERENCES

1. S.H. Roth. History of Automatic Vehicle Monitoring (AVM). IEEE Trans., Vehicular Technology, Vol. VT-26, Feb. 1977.
2. M. Lukes and R. Shea; Chicago Transit Authority. MONITOR-CTA. UMTA, Final Rept., May 1973.
3. R.E. Fenton. On Future Traffic Control: Advance Systems Hardware. IEEE Trans., Vehicular Technology, Vol. VT-29, No. 2, May 1980.
4. H.G. Miller and W.M. Basham. Evaluation of the MONITOR-CTA Automatic Vehicle Monitoring System. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Rept. UMTA-MA06-0024-74-1, March 1974.
5. O. Bevilacqua, R. Knight, J. Schmidt, R. Waksman, and W. Wade; DeLeuw, Cather and Company. Evaluation of the Cincinnati-Type On-Board Data Collection System (ODCS). Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Final Rept., March 1979.
6. N.A. Irwin, F.D. Catton, and L.E.S. Green. Transit Vehicle Fleet Information and On-Line Management. IEEE Trans., Vehicular Technology, Vol. VT-29, No. 2, May 1980.
7. B.E. Blood and B.W.A. Kleim. Experiments of Four Different Techniques for Automatically Locating Land Vehicles--A Summary of Results. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Rept. UM-721/R-7716, May 1977.
8. D.J. Symes. Automatic Vehicle Monitoring: A Tool for Vehicle Fleet Operations. IEEE Trans., Vehicular Technology, Vol. VT-29, No. 2, May 1980.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Comparing Fixed-Route and Flexible-Route Strategies for Intraurban Bus Transit

OLUSEGUN ADEBISI AND V.F. HURDLE

The usual fixed-route strategy is not the only possible strategy for operating intraurban bus transit. Among the alternatives are flexible-route strategies. This paper focuses on the problem of choosing between fixed-route and flexible-route strategies in order to optimize operations. A mathematical model is used to determine the optimum quantity of service that should be provided under each strategy so as to minimize the costs to operators and users. The quantity of service is characterized by the headways between buses and is given as a function of the average ridership rates, unit costs, and travel times. By comparing the optimum states for the two strategies, the conditions under which one strategy performs better than the other are derived. Findings from the latter are then used to derive a general methodology for comparing the strategies. The highlight of the proposed methodology is that the two strategies must be compared at the extremes of a typical day's ridership levels before one can ascertain whether operating exclusively with either strategy or jointly with both strategies will give the best results. The present study is, however, limited to very small service areas.

The current methodology for comparing fixed-route and flexible-route operating strategies of intraurban bus transit consists essentially of using cost-effectiveness curves (1,2). The cost of providing a preselected level of service is determined as a function of demand for each strategy. That level of demand for which the service cost is equal for the two strategies is referred to as the critical ridership rate. If the design ridership rate is less than the critical ridership rate, then a flexible-route strategy is considered to be more suitable than a fixed-route strategy, while a fixed-route strategy is more suitable where the design ridership rate exceeds the critical ridership rate.

However, the above method ignores the time-variant nature of transit ridership. Because only design demand is considered in the analysis, one cannot be certain that whichever strategy is chosen is actually superior to the other over all ranges of demand encountered on a typical day. Also, a preselected level of service will not always ensure optimality.

In this paper, a new method that seeks to correct the above-mentioned deficiencies is proposed. By using a simple mathematical model of an intraurban bus transit system, the complications of choosing between the strategies are investigated. The proposed method compares the strategies on the basis of their capabilities to minimize costs to operators and users.

The physical setting assumed in developing the mathematical model is one in which the service area is given, but the transit operator wishes to develop optimal operating policies given that only scheduled fixed-route and/or scheduled flexible-route operating strategies will be considered. In order to simplify the analysis, a simple situation that consists of a very small service area that requires only a single fixed-route or a single service zone for flexible-route service is considered. It is hoped that when the mathematical model is extended to cover larger areas, a general comparison method for areawide as well as a corridor-by-corridor analysis would emerge.

The paper consists of two parts. The formulation of the mathematical model, which constitutes the foundation of the comparison method, is presented first. The second part contains relevant deductions and theorems as well as the derivation of the method. Table 1 contains the nomenclature that will be used throughout the paper.

Table 1. List of notations used.

Notation	Definition
γ	Average bus operating cost per unit time
γ_{iv}	Unit value of passenger's travel time spent inside a bus
γ_{ov}	Unit value of passenger's travel time spent outside a bus
λ	Fraction of total passengers who require a communication device to register their demands under flexible-route strategy
σ_2	Fraction of bus trip time for which a typical passenger remains on bus under fixed-route strategy
σ_2'	Fraction of bus trip time for which a typical passenger remains on bus under flexible-route strategy
σ_3	Average increase in the variance of bus trip time per passenger under flexible-route strategy
Ψ	Average communication cost per passenger
C_o	Operators' total cost per unit time
C_u	Users' average cost per unit time
F	Number of buses used for service
T	Time scheduled for a bus trip
T_{rd}	Actual driving time during a bus trip
Q	Average number of passenger trips per unit time
Q_{max}	Maximum value of Q on a typical day
Q_{min}	Minimum value of Q on a typical day
$Q_{c,0}$	Value of Q at and beyond which bus capacity is constrained under fixed-route strategy
$Q_{c,1}$	Value of Q at and beyond which optimum headways are constrained by bus capacity under flexible-route strategy
Z	Total system cost
z_0^*	Minimum value of Z for fixed-route strategy divided by Q
z_0^{**}	Value of z_0^* when optimum headways are constrained by vehicle capacity
z_1^*	Minimum value of Z for flexible-route strategy divided by Q
z_1^{**}	Value of z_1^* when optimum headways are constrained by vehicle capacity
h	Time headway between buses
h_0^*	Theoretically optimum headway under the fixed-route strategy
h_1^*	Theoretically optimum headway under the flexible-route strategy
h_1'	Optimum feasible headway under the flexible-route strategy
q	Mean number of passengers served during a bus trip
q_{cap}	Capacity (passenger spaces) of each vehicle used for fixed-route service
q_{cap}'	Capacity (passenger spaces) of each vehicle used for flexible-route service
t_F	Time scheduled for a bus trip under the fixed-route strategy
t_F'	Fixed component of the time scheduled for a bus trip under the flexible-route strategy
t_1	Average service time per passenger allowed under the flexible-route strategy
t_{iv}	Average travel time spent inside the vehicle per passenger trip
t_{ov}	Average travel time spent outside the vehicle per passenger trip

MATHEMATICAL MODEL

Mathematical Formulation of Problem and Simplifying Assumptions

An objective function of minimizing the sum of operating costs and the expectation of the users' costs are assumed. So, if we define the operating costs to be C_o and the expectation of the users' costs to be C_u, then

$$Z = C_o + C_u \tag{1}$$

where Z is the total cost to be minimized. A similar objective function was used elsewhere (3-5). Because the magnitudes of the defined costs depend on the length of time considered, C_o and C_u are based on unit time.

An average ridership rate of Q passenger trips per unit time is assumed. Although Q is implicitly assumed to be dependent on the time of day, it is assumed to be unaffected by the strategy.

It is assumed that F buses are used for service within the unit time under consideration and that C_o can be approximately modeled as follows:

for fixed-route strategy,

$$C_o \simeq \gamma F \tag{2a}$$

and for flexible-route strategy,

$$C_o \simeq \gamma F + \Psi\lambda Q \tag{2b}$$

where

γ = average total cost of operating a vehicle per unit time,
Ψ = average cost (per passenger) of providing the communication medium for the flexible-route strategy, and
λ = proportion of passengers served who require a communication facility to register their demand.

Actually, the ridership rate is dependent on the time of day and, because of this, transit operators do not always provide the same quantity of service throughout the day. Also, the operators' costs consist partly of a component that directly varies with the quantity of service provided and partly of a fixed component that is independent of the quantity of service provided. Consequently, the values of γ and Ψ may not necessarily be the same throughout a whole day. The question of how to correctly apportion transit costs among different demand periods is not addressed in the present study. Both γ and Ψ are treated as constants. It is also possible that vehicle operating costs would be different for the two strategies. However, because a substantial portion of these costs is labor costs, it is the efficiency of labor use on the system rather than operating strategy that will significantly influence the value of γ. Thus, γ is taken to have approximately the same value for both strategies.

Users' costs consist of the fares and costs attributable to their total time commitment for the trip. The fares are considered to be internal to the bus transit system and are not included in the evaluation.

A user's time commitment to a trip consists of time actually spent on the vehicle and time spent outside of it. If we suppose that each unit of time spent inside the vehicle is worth γ_{iv} units of money while each unit of time spent outside is worth γ_{ov}, then

$$C_u = Q(\gamma_{iv}t_{iv} + \gamma_{ov}t_{ov}) \tag{3}$$

where $\gamma_{ov} \geqslant \gamma_{iv}$, t_{iv} is the average time spent inside the vehicle per passenger trip, and t_{ov} is the average time spent outside the vehicle per passenger trip.

By using a constant time value, it erroneously implies that all users attach the same value of travel time and that the marginal utility of travel time is constant, regardless of the time length of the trip. It should be pointed out, then, that because of the assumption, wealthy users, who are likely to attach higher values to their travel times, and long distance travelers, who are likely to have higher marginal time utility, are disadvantaged. In studies directly aimed at estimating travel demands, both factors would have to be allowed for.

Substituting Equations 2a and b and 3 into Equation 1 gives the following:

for fixed-route strategy,

$$Z = \gamma F + Q(\gamma_{iv}t_{iv} + \gamma_{ov}t_{ov}) \tag{4a}$$

and for flexible-route strategy,

$$Z = \gamma F + Q(\lambda\Psi + \gamma_{iv}t_{iv} + \gamma_{ov}t_{ov}) \tag{4b}$$

In comparing the two strategies, all design variables are first chosen so as to optimize Z within each strategy. After this, the optimal values of Z for the two strategies are compared. It is on the basis of the results that a general methodology for comparing the strategies is proposed.

In general, the two main design variables are the dispatch headways and the dimensions and configuration of the service zones (for flexible-route strategy) or spacing between the routes (for fixed-route strategy). In this paper, the service area is assumed to be given and to be small enough to require only a single route under the fixed-route strategy or a single service zone under the flexible-route strategy. However, some publications (5,6) include the topic of route spacing under the fixed-route strategy while the problem of optimally partitioning an area into service zones is investigated in Ward (1).

If we represent the headway by h and the scheduled vehicle trip time as T, then:

$$F = T/h \tag{5}$$

Normally, T (the scheduled trip time) is comprised of a slack time and the estimated expectation of the actual driving time. The driving time is denoted as T_{rd}. Although T should be a constant for a given ridership rate, T_{rd} is random.

The following mathematical relations are assumed for the travel times contained in Equations 4a and b and 5:

for fixed-route strategy,

$$t_{ov} = 2\sigma_1 \left\{[h^2 + \mathrm{var}(T_{rd})]/h\right\} + t_{acc} \tag{6a}$$

and for flexible-route strategy,

$$t_{ov} = 2\sigma_1 \left\{[h^2 + \mathrm{var}(T_{rd})]/h\right\} \tag{6b}$$

for fixed-route strategy,

$$T = t_F \tag{6c}$$

and for flexible-route strategy,

$$T = t'_F + t_1 q \tag{6d}$$

for fixed-route strategy,

$$t_{iv} = \sigma_2 T \tag{6e}$$

and for flexible-route strategy,

$$t_{iv} = \sigma'_2 T \tag{6f}$$

for fixed-route strategy,

$$\mathrm{var}(T_{rd}) = 0 \tag{6g}$$

and for flexible-route strategy,

$$\mathrm{var}(T_{rd}) = \sigma_3 q \tag{6h}$$

where the notations are as follows:

1. σ_1, σ_2, σ_2', and σ_3 are constants such that $0 < \sigma_1 \leqslant 0.5$, $0 < \sigma_2 \leqslant 1$, $0 \leqslant \sigma_2' < 1$, and $\sigma_3 > 0$;
2. t_F, t'_F, t_1, and t_{acc} are also constants for a given service area; and
3. q is the average number of passengers served during a vehicle trip and, by definition

$$q = Qh \tag{6i}$$

Equations 6a and b indicate that t_{ov} consists of t_{acc}, the access time between the bus route and the passenger's trip-end point and the waiting times. Although t_{acc} is a function of the width of the service zone, the waiting time consists of (a) the waiting time at the point of boarding and (b) the schedule delay, which represents the extra time that a user commits to the trip because his or her preferred arrival time at the destination point differs from the vehicle's schedule. For example, in a situation where all the passengers' trip-inception times coincide with the bus arrival times at the boarding points, and the bus arrival times at the respective destination points are exactly the same as the passengers' preferred arrival times, then $\sigma_1 = 0$. However, in the extreme situation where both the passengers' trip-inception times and preferred arrival times at their destinations, independent of the bus schedule, are distributed uniformly over time, then $\sigma_1 = 0.5$. This latter value of σ_1 is assumed, for illustration purposes, in the subsequent discussion. The actual numerical value of σ_1 (provided $\sigma_1 > 0$) is, however, of little significance to the present work.

Equations 6c and d and 6g and h were found to be approximately true from a computer simulation of bus operation during intraurban service undertaken by Adebisi (7). In the simulations, the scheduled trip time was taken to be the sum of the total driving time expectation and twice its standard deviation. The number of passengers served was taken to be randomly distributed over time and space but with a fixed average value per unit time. Only the variability in the bus trip times due to randomness in the bus load was allowed for. One would expect the results, particularly for Equations 6g and h, to be different if the variability in the trip times due to interaction with other vehicles that use the roadway is considered.

Equations 6c and d imply that it is only with flexible-route service that the passengers' service time constitutes a significant portion of the scheduled trip time and also confirms the assumptions (3,8) that passengers' service time in a scheduled, fixed-route operation has a negligible effect on the scheduled trip time. Results of the simulations indicate that, while t_F is mostly a function of the length of the fixed route and vehicle speed characteristics, it generally increases with the length of the bus route. Similarly, t_F', t_1, and σ_1 generally increase with the area of the service zone.

The parameters σ_2 and σ_2' in Equations 6e and f represent the fraction of a bus trip time that constitutes the average duration of a passenger trip time. Consequently, their numerical values depend on the passengers' origin-destination (O-D) pattern. No specific values are assumed for σ_2 and σ_2' in the subsequent analysis.

Objective Function

Combining Equations 4a and b, 5, and 6a-i gives the following:

for fixed-route strategy,

$$Z = (\gamma t_F/h) + Q\ [\gamma_{iv}\sigma_2 t_F + \gamma_{ov}(h + t_{acc})] \tag{7a}$$

and for flexible-route strategy,

$$Z = (\gamma t_F'/h) + Q\ [\Psi\lambda\gamma + \gamma_{iv}\sigma_2'(t_F' + Qht_1) + \gamma_{ov}(h + \sigma_3 Q)] \tag{7b}$$

In order to minimize Z, we must choose an appropriate value of h, which is the only design variable in the present model. Differentiating Z with respect to h gives h_0^* (the optimum headway for the fixed-route strategy) as follows:

$$h_0^* = (\gamma t_F/\gamma_{ov}Q)^{1/2} \tag{8a}$$

Thus, we have from Equations 6g and h and 8a that the mean vehicle load, when optimal headways are used, is as follows:

$$q = (\gamma t_F Q/\gamma_{ov})^{1/2} \tag{8b}$$

The implication from Equation 8b is that, because vehicles have finite capacities, optimal headways may not always be feasible when Q assumes large values. It is therefore necessary to distinguish between optimal feasible headway and the theoretically optimal headway. If we let q_{cap} represent the vehicle capacity and h_0' represent the optimal feasible headway under fixed-route operation, then:

$$h_0' = \min\ [h_0^*, (q_{cap}/Q)] \tag{8c}$$

Because of the randomness in the vehicle load, one may need to allow a safety factor in selecting the value of q_{cap} and not simply use actual capacity. However, if all passengers board the vehicle at the dispatch point and a sufficient supply of vehicles is always available, one does not need a safety factor.

Similarly, if we let h_1' and h_1^* represent the optimal feasible headway and the theoretically optimal headway, respectively, for the flexible-route strategy, we have the following:

$$h_1^* = \left\{ t_F'/[(\gamma_{ov}Q) + (\gamma_{iv}\sigma_2' t_1 Q^2)] \right\}^{1/2} \tag{9a}$$

while

$$h_1' = \min\ [h_1^*, (q_{cap}'/Q)] \tag{9b}$$

where q_{cap}' is the capacity of each vehicle used for flexible-route service. Based on the present practice whereby transit operators use small buses for flexible-route service and large buses for fixed-route service, it is reasonable to suppose that $q_{cap}' \leq q_{cap}$. It should be observed that any other headways beside h_0' and h_1' would only lead to suboptimal states for the strategies. Comparisons based on such suboptimal headways would obviously give biased results. Let us denote as z_0^* and z_1^* the minimum values of Z divided by Q for the fixed-route strategy and the flexible-route strategy, respectively. Thus, z_0^* and z_1^* represent the average total minimum disutility per passenger. Because Q is assumed to be independent of the strategy, it should not make any difference whether we use Z or z_0^* and z_1^* in our comparison. However, because it considerably simplifies subsequent analysis, the latter option is adopted. Thus,

$$z_0^* = (\gamma t_F/Qh_0') + \gamma_{iv}\sigma_2 t_F + \gamma_{ov}(h_0' + t_{acc}) \tag{10a}$$

and

$$z_1^* = (t_F'/Qh_1') + \gamma\Psi + \gamma_{iv}\sigma_2'(t_F' + Qh_1' t_1) + \gamma_{ov}(h_1' + \sigma_3 Q) \tag{10b}$$

Complications in choosing between the strategies are examined by exploring the sensitiveness of z_0^* and z_1^* to changes in Q. Thus, we differentiate z_0^* and z_1^* with respect to Q and obtain the following:

if theoretically optimal headways are used, i.e., if $h_0' = h_0^*$,

$$\partial z_0^*/\partial Q = -(\gamma_{ov} t_F/Q^3)^{1/2} \tag{11a}$$

and if capacity-constrained headways are used, i.e., if $h_0' = q_{cap}/Q$,

$$\partial z_0^*/\partial Q = -(\gamma_{ov} q_{cap}/Q^2) \tag{11b}$$

and if $h_1' = q_{cap}'/Q$,

$$\partial z_1^*/\partial Q = \gamma_{ov}\sigma_3 - (\gamma_{ov} q_{cap}'/Q^2) \tag{11c}$$

and if $h_1' = h_1^*$,

$$\partial z_1^*/\partial Q = \gamma_{ov}\sigma_3 - \left\{ \gamma_{ov} t_F'/Q^3\ [1 + (\gamma_{iv}\sigma_2/\gamma_{ov}\sigma_2')\cdot t_1 Q] \right\}^{1/2} \tag{11d}$$

Equations 11a and b indicate that z_0^* always decreases with Q and thus confirms the general belief that the fixed-route strategy is characterized by economies of scale. Equations 11c and d, on the other hand, indicate that z_1^* might actually increase with Q such as when

$$\gamma_{ov}\sigma_3 > \left\{ \gamma_{ov} t_F'/Q^3\ [1 + (\gamma_{iv}\sigma_2/\gamma_{ov}\sigma_2')\cdot t_1 Q] \right\}^{1/2}$$

or when

$$\gamma_{ov}\sigma_3 > (\gamma_{ov} q_{cap}'/Q^2).$$

The conditions that lead to z_1^* increasing with Q are more likely to be met when Q assumes large values than when it is small. This finding also affirms the reasonableness of the general aversion to recommend a flexible-route strategy for high levels of ridership demands.

COMPARISON OF STRATEGIES

At this stage, useful inferences on the relative performance of the two strategies under consideration can be drawn. The situation when headways under the two strategies are constrained by vehicle capacity (thus being more straightforward) is discussed first. Numerically, $\sigma_3 > 0$, and $q_{cap}' \leqslant q_{cap}$. It is implied from Equations 11a-d that, when the vehicles used for service under both strategies are always fully loaded, z_1* never decreases with Q at a faster rate than z_0*.

Also, it will be recalled from Equations 6c and d that t_F represents the scheduled trip time under the fixed-route strategy while t_F' represents just the fixed portion of the scheduled trip time under the flexible-route strategy. It is true that when the same demands are served, flexible-route service is likely to take a longer time to complete than fixed-route service, but the service times (i.e., buses' access times to demand points and consequent loading times) that t_{1q} represents make up the bulk of the scheduled times. It will therefore be appreciated that t_F' will almost never exceed t_F. Given that $t_F > t_F'$, it follows from Equations 11a-d that, when theoretically optimal headways are used for both strategies, z_1* never decreases with Q at a faster rate than z_0*.

It is possible that, for some values of Q, theoretically optimal headways are appropriate with one strategy, but for the other strategy the headways must be based on vehicle capacity. Therefore, the above does not constitute sufficient proof that z_1* never decreases with Q at a faster rate than z_0* in all situations, but it is sufficient to prove lemma 1, which is stated as follows: When it is optimal for both strategies that theoretically optimum headways always be used or that the headways always be based on vehicle capacity and if we find that for a specific value of Q that $z_0^* < z_1^*$, then $z_0^* < z_1^*$ for all higher values of Q. Similarly, if we find that for a specific value of Q that $z_1^* < z_0^*$, then $z_1^* < z_0^*$ for all smaller values of Q.

By applying lemma 1, it is shown in the next section that, in most cases, where only one strategy is required the appropriate strategy is uniquely determined by considering the extreme ridership rates only.

Proofs of Relevant Theorems

Let us suppose that the maximum and minimum ridership rates likely to be served on a typical day within the service area are Q_{max} and Q_{min}. Let us also represent the values of Q when the theoretically optimal headways are exactly equal to the headways based on vehicle capacity as $Q_{c,0}$ and $Q_{c,1}$ for the fixed- and flexible-route strategies, respectively. For convenience, we denote the value of z_0* as z_0 when the headways are based on the vehicles' capacities but as z_0** when theoretically optimal headways are used. The corresponding values for z_1* are z_1 and z_1**. Thus,

when $Q \leqslant Q_{c,0}$,

$$z_0^* = z_0^{**} \tag{12a}$$

and when $Q > Q_{c,0}$,

$$z_0^* = z_0 \tag{12b}$$

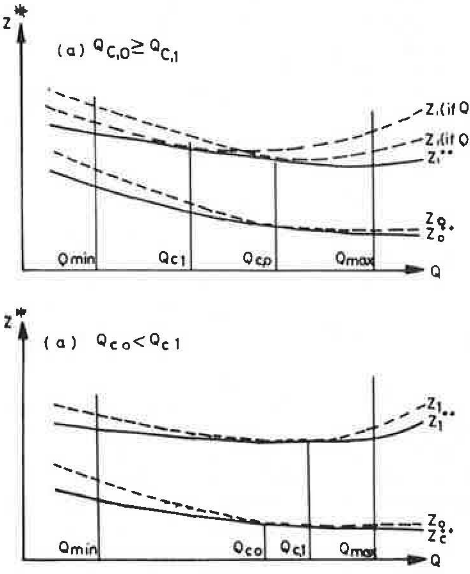
and when $Q \leqslant Q_{c,1}$,

$$z_1^* = z_1^{**} \tag{12c}$$

and when $Q > Q_{c,1}$,

$$z_1^* = z_1 \tag{12d}$$

Figure 1. Conceptualization of some situations when fixed-route strategy is more appropriate at Q_{min}.



Let us first consider the case where the fixed-route strategy performs better than the flexible-route strategy at $Q = Q_{min}$; i.e., $z_0^* < z_1^*$ at $Q = Q_{min}$. The following can then be deduced:

1. If $Q_{c,0} > Q_{max}$ and $Q_{c,1} > Q_{max}$, then theoretically optimal headways should be used under both strategies for all relevant values of Q; i.e., $Q \in [Q_{min}, Q_{max}]$. It follows from Equations 12a-d that $z_0^* = z_0^{**}$ and $z_1^* = z_1^{**}$ for $Q \in [Q_{min}, Q_{max}]$. Since $z_0^{**} < z_1^{**}$ at $Q = Q_{min}$, it follows from lemma 1 that $z_0^{**} < z_1^{**}$ for $Q \in [Q_{min}, Q_{max}]$ and the fixed-route strategy is more appropriate than the flexible-route strategy for all relevant ranges of demand.
2. If $Q_{c,0} < Q_{min}$ and $Q_{c,1} < Q_{min}$, by implication headways should be based on vehicle capacity under both strategies for all $Q \in [Q_{min}, Q_{max}]$. Therefore, it follows from Equations 12a-d that for all $Q \in [Q_{min}, Q_{max}]$, $z_0^* = z_0$ and $z_1^* = z_1$. Because $z_0 < z_1$ at $Q = Q_{min}$, it follows from lemma 1 that $z_0 < z_1$ for all $Q \in [Q_{min}, Q_{max}]$ and the fixed-route strategy is more appropriate within the relevant ranges of demand.
3. If $Q_{min} < Q_{c,0} < Q_{max}$ or $Q_{min} < Q_{c,1} < Q_{max}$, then both optimum headways and headways based on vehicle capacity should be used for at least one of the strategies. Figure 1 gives a conceptual representation of the possible situations where $Q_{c,0} \geqslant Q_{c,1}$ and $Q_{c,0} < Q_{c,1}$.

When $Q_{c,0} \geqslant Q_{c,1}$, then the flexible-route strategy requires that optimal feasible headways be based on vehicle capacity at the same or lower ridership rates than the fixed-route strategy requires. But from lemma 1 and Equations 12a-d we know that if $z_0^* < z_1^*$ at Q_{min}, then $z_0^{**} < z_1^{**}$ and $z_0 < z_1$ for $Q \geqslant Q_{min}$. Furthermore, because $z_1^{**} \leqslant z_1$ for all values of Q, it follows that $z_0^{**} < z_1^{**} \leqslant z_1$ for all $Q \geqslant Q_{min}$. Following from Equations 12a-d, we know that for $Q \in [Q_{min}, Q_{c,1}]$, $z_0^* = z_0^{**}$ and $z_1^* = z_1^{**}$. Thus, $z_0^* < z_1^*$. For $Q \in (Q_{c,1}, Q_{max}]$, $z_1^* = z_1$; but $z_0^* = z_0^{**}$ for $Q \in (Q_{c,1}, Q_{c,0}]$ and $z_0^* = z_0$ for

$Q \in (Q_{c,0}, Q_{max}]$. Thus, $z_0^* < z_1^*$ for $Q \in (Q_{c,1}, Q_{max}]$, too. Hence, $z_0^* < z_1^*$ for all relevant values of Q when $Q_{c,0} \geq Q_{c,1}$. Therefore, the fixed-route strategy should be used throughout. When $Q_{c,0} < Q_{c,1}$, then the fixed-route strategy requires that optimal feasible headways be based on vehicle capacity at lower ridership rates than with the flexible-route strategy. As in the former case, we can deduce that for $Q \geq Q_{min}$, $z_0^{**} < z_1^{**}$ and $z_0 < z_1$.

In addition, $z_0^{**} \leq z_0$ and $z_1^{**} \leq z_1$. However, while it is true that $z_0^{**} \leq z_1^{**} \leq z_1$ for all $Q \geq Q_{min}$, it is not necessarily true that $z_0 < z_1^{**}$ for all $Q \geq Q_{min}$. However, for $Q \in [Q_{min}, Q_{c,0}]$, $z_0^* = z_0^{**}$ and $z_1^* = z_1^{**}$. Thus, $z_0^* < z_1^*$. For $Q \in (Q_{c,0}, Q_{max}]$, $z_0^* = z_0$, but $z_1^* = z_1^{**}$ for $Q \in (Q_{c,0}, Q_{c,1}]$ and $z_1^* = z_1$ for $Q \in (Q_{c,1}, Q_{max}]$. Therefore, the fixed-route strategy is more appropriate for $Q \in [Q_{min}, Q_{c,0}]$ and $Q \in (Q_{c,1}, Q_{max}]$ but, since z_0 is not necessarily less than z_1^{**}, a more detailed investigation is required to determine if, in fact, the fixed-route strategy is appropriate within the interval $(Q_{c,0}, Q_{c,1}]$.

The above deductions are sufficient to prove theorem 1, which is stated below.

Theorem 1

If the fixed-route strategy is found to perform better than the flexible-route strategy at $Q = Q_{min}$, then the fixed-route strategy is more appropriate for all relevant values of Q whenever (a) theoretically optimal headways are always feasible under both strategies or (b) optimal feasible headways are always based on vehicle capacity for both strategies or (c) both theoretically optimal headways and headways based on vehicle capacity are used for at least one of the strategies but $Q_{c,0} \geq Q_{c,1}$.

In addition, if the fixed-route strategy is found to be more appropriate than the flexible-route strategy for $Q = Q_{min}$, and both optimal headways and headways based on the vehicles' capacities are used for the fixed-route strategy with $Q_{c,0} > Q_{c,1}$, then the fixed-route strategy is appropriate for $Q \in [Q_{min}, Q_{c,0}]$ and $Q \in (Q_{c,1}, Q_{max}]$.

By using a similar procedure and logic similar to the above, we can also prove theorem 2.

Theorem 2

If the flexible-route strategy is found to perform better than the fixed-route strategy for $Q = Q_{max}$, then the flexible-route strategy is more appropriate for $Q \in [Q_{min}, Q_{max}]$ whenever (a) optimal headways are always used under both strategies or (b) optimal feasible headways are always based on vehicle capacity for both strategies or (c) both optimal headways and headways based on the buses' capacities are used for at least one of the strategies but $Q_{c,1} \geq Q_{c,0}$.

In addition, if the flexible-route strategy is found to be more appropriate for $Q = Q_{max}$ and both optimal headways and headways based on on the buses' capacities are used for flexible-route strategy with $Q_{c,1} < Q_{c,0}$, then the flexible-route strategy is more appropriate for $Q \in [Q_{min}, Q_{c,1}]$ and $Q \in (Q_{c,0}, Q_{max}]$.

In real-life situations, the values of $Q_{c,1}$ and $Q_{c,0}$ would be very large and, if not larger than Q_{max}, should be close to it. Otherwise, the obvious implication from the present work is that improvements are possible simply by having larger vehicles. Thus, the range of demand $(Q_{c,1}, Q_{c,0}]$ or $(Q_{c,0}, Q_{c,1}]$ where the proven theorems indicate inconclusive results should not be of much practical importance. A general methodology for comparing the strategies is proposed in the next section.

General Methodology for Comparing Strategies

Theorems 1 and 2 indicate that, if a bus transit operator wants to choose one of the fixed-route and flexible-route strategies, he or she needs to compare them at $Q = Q_{max}$ and $Q = Q_{min}$. Theorem 1 indicates that if the fixed-route strategy is found to be superior to the flexible-route strategy at $Q = Q_{min}$, then it will, for most practical situations, be superior for all relevant ranges of demand. It is therefore desirable to operate exclusively with the fixed-route strategy in such a situation. Similarly, if the flexible-route strategy is found to be more appropriate than the fixed-route strategy at $Q = Q_{max}$, then it is desirable to operate exclusively with the flexible-route strategy. However, when the fixed-route strategy is found to be the better strategy at $Q = Q_{max}$ but the flexible-route strategy is more suitable at $Q = Q_{min}$, then neither strategy should be used exclusively. In such a case, the transit operator could determine the range of demand for which one strategy is superior to the other and then draw up a schedule for switching strategies as the demand rate changes from one region to the other. Such an arrangement, however, poses some operational problems. In a situation whereby the fixed-route strategy is exclusively adopted, large vehicles are likely to be used, but where the flexible-route strategy is exclusively adopted, mostly small buses will be used because they are easier to maneuver. Therefore, some compromise may have to be made on fleet composition when operating strategies are routinely switched. This compromise may increase overall operating costs. When allowance is made for the extra costs, it is possible that one finds it better to operate exclusively with one strategy. However, making a rational choice precludes the adoption of an all-or-nothing principle in choosing between the fixed-route and flexible-route strategies.

CONCLUSION

It has been shown that choosing between a fixed-route strategy and a flexible-route strategy is more complicated than earlier studies indicated. Evidently, a comparison method based solely on the determination of the critical ridership rate is inadequate, since it indirectly assumes that demand is invariant with time. Such a comparison method can only identify the dominant strategy that should be used more widely than the other strategy, since it does not conclusively show that the chosen strategy performs better for all relevant ranges of demand. Also, by comparing the strategies on the basis of the costs required to achieve a specified level of service, one runs the risk of comparing them at nonoptimum states. The indication from the present work is that, when the strategies are compared at states other than the optimum, the results are likely to be biased.

The method proposed in this study has tried to remove the limitations described above that are inherent in the current methods of comparing the strategies. However, this new method is also not without its own limitations. For one, the fact that the analysis was focused on a very small service area limits its application. Another limitation is that no allowance was made for the interplay of demand and level of service. Further research on the subject should aim at removing the identified weaknesses in the model.

REFERENCES

1. D.E. Ward. A Theoretical Comparison of Fixed Route Bus and Flexible Route Subscription Bus Feeder Service in Low Density Areas. U.S. Department of Transportation, March 1975.
2. D. Roos and others. The Dial-a-Ride Transportation Systems. Urban Systems Lab, Massachusetts Institute of Technology, Cambridge, Rept. USL-TR-70-10, 1971. NTIS: PB 201-129.
3. G.F. Newell. Dispatching Policies for a Public Transportation Route. Transportation Science, Vol. 5, No. 1, Feb. 1971.
4. V.F. Hurdle. Minimum Cost Schedule for a Public Transportation Route. Transportation Science, Vol. 7, No. 2, March 1973.
5. V.F. Hurdle. Minimum Cost Locations for Parallel Public Transit Lines. Transportation Science, Vol. 7, No. 4, Nov. 1973.
6. B.F. Bryne and R.R. Vuchic. Public Transportation Line Positions and Headways for Minimum Cost. Presented at Fifth International Symposium on Traffic Flow Theory and Transportation, Berkeley, CA, 1973.
7. O. Adebisi. A Comparison of the Fixed-Route and Flexible-Route Strategies of Operating an Intra-urban Bus Transit System. School of Graduate Studies, Univ. of Toronto, Ontario, Canada, Ph.D. dissertation, 1978.
8. E.M. Holroyd. The Optimum Bus Service: A Theoretical Model for a Large Uniform Area. Presented at Third International Symposium on the Theory of Traffic Flow, New York, 1965.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Concurrent-Flow High-Occupancy Vehicle Treatment on Freeways—Success Story in Houston

CHARLES A. FUHS

On March 30, 1981, a 3.3-mile concurrent-flow lane began operation within the median shoulder on North Freeway (Interstate-45). The concurrent-flow lane operates inbound only from 6:00 to 8:30 a.m. and is available to authorized vehicles, which include registered and approved buses and eight-passenger vanpools. The concurrent-flow lane is an extension of contraflow preferential treatment provided further downstream; it provides a travel-time savings of about 4 min. This project is one of seven nationwide that is currently operating, is the only project to be implemented within an existing paved emergency shoulder, and is the first operation to restrict use to authorized vehicles that display an appropriate permit. A general report on the unique characteristics and results of Houston's concurrent-flow operation is presented. Comparative evaluations are presented that measure the success of this project with other concurrent-flow applications on freeways. In the first three months, an average of 257 vehicles (78 percent vanpools and 21 percent buses) traveled the lane inbound during each daily 2.5-h peak period, which facilitated the movement of 3752 commuters. The North Freeway concurrent-flow project was jointly implemented by the Texas State Department of Highways and Public Transportation and the Metropolitan Transit Authority of Harris County. Both agencies funded construction of the project with local monies and jointly managed daily operation. The success of the concurrent-flow project, as illustrated in this paper, has resulted in increased person trips on a severely congested freeway facility and has provided a travel-time incentive to vanpool and bus transit users until such time that a more permanent transitway facility can be constructed.

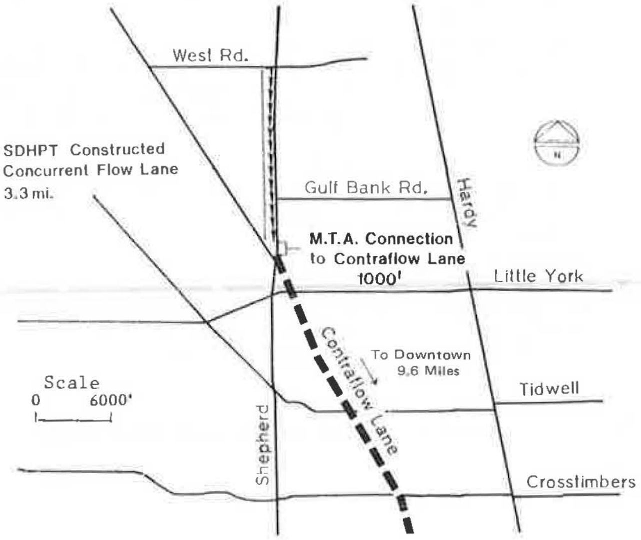
In 1979 the Metropolitan Transit Authority (MTA) of Harris County, Houston, and the Texas State Department of Highways and Public Transportation (TSDHPT) opened a 9.6-mile contraflow lane on Interstate-45N (North Freeway). The $2.1 million project, funded under a Service and Methods Demonstration program grant (Sections 5 and 6) of the Urban Mass Transportation Act of 1964 (as amended), was very successful in attracting riders into vanpools and buses. These were the only authorized vehicles that could benefit from the project, and rather rigid authorization procedures were adopted to help ensure safe operation. The contraflow lane bypassed about 6 miles of severe traffic congestion and saved users about 30-min of travel time daily. Use increased 350 percent from the 1st through 82nd week of operation to 10 900 daily trips (1). However, during the contraflow planning and implementation period from 1975 through 1979, severe traffic congestion was growing and began extending several miles upstream of the northern terminus to contraflow. An extension of the contraflow concept to alleviate this problem was complicated by several factors. Unacceptable traffic conditions upstream did not permit borrowing a lane for contraflow. Also, facility design would not accommodate a safe project termination farther north. Other alternatives were studied for bypassing congestion outside the contraflow limits.

BACKGROUND

The concurrent-flow concept was first proposed as an extension to the contraflow lane in early 1980 to alleviate congestion in the morning period. The concept could be readily implemented within an existing paved median shoulder along a 3.3-mile segment, as shown in Figure 1. The segment was unique in that the termination of the concurrent-flow lane could be transitioned directly to the contraflow lane. This segment encompassed most of the regularly recurring traffic congestion. Median drainage inlets and superelevations prohibited easy conversion of the inbound median any further. In the afternoon peak period, traffic conditions at the time did not warrant implementation of a similar treatment on the outbound shoulder.

TSDHPT subsequently designed the necessary signing and striping modifications to convert the median shoulder for bus and vanpool use. A connection ramp was designed at the downstream terminus to facilitate direct access from the concurrent-flow shoulder to the entry of contraflow. An exception was granted from Interstate standards by the Federal

Figure 1. Concurrent-flow shoulder lane.



Highway Administration in fall 1980. Project implementation was expedited by use of local monies from both agencies to fund construction. TSDHPT installed signs, restriped lanes to accommodate the lane over bridge decks, and reinforced bridge railings. MTA constructed a connection ramp and gate. Total cost to both agencies was about $130 000. Construction began in November 1980 and was completed about four months later.

Operation management of the concurrent-flow lane (CCFL) was already provided under a previous operations plan between MTA and TSDHPT for the contraflow lane. The operating plan finalized and made legal the following:

1. Operating hours and schedule,
2. Requirements for authorized vehicles,
3. Requirements for authorized drivers,
4. Rules and regulations for use of the lane,
5. Enforcement procedures,
6. Maintenance responsibilities, and
7. Emergency and breakdown procedures.

Most of the contraflow policies were transferred to encompass CCFL operation. Authorized users were designated as recipients, which included existing buses and vanpools that operate on the contraflow lane.

The plan has been made the official ruling document by an MTA-TSDHPT operations agreement, which also provided for a project management team to oversee the project and make amendments to the plan by mutual consent of the TSDHPT project engineer and MTA project manager. This arrangement has worked very well for the past two years. Amendments to the plan can be made quickly and effectively at monthly meetings without amending the governing operations agreement.

Of particular interest in securing an operations agreement was the ability of MTA to enforce restrictions that govern the use of the median shoulder. The intent of a previous ordinance enacted by the city empowering their police to enforce the contraflow project was expanded to encompass the concurrent-flow project.

CCFL was opened to authorized users on March 30, 1981, without public fanfare. Notices and driver-training information were distributed in a packet to authorized bus and vanpool drivers a week before the Monday opening. An example of this information is shown in Figure 2. Bus passengers received the notice shown in Figure 3 several days before buses began traveling in the shoulder lane. Newspaper articles represented the only media coverage on the project. Operation is shown in Figure 4.

Characteristics of North Freeway

North Freeway is a standard six-lane Interstate facility within the concurrent-flow project limits. The freeway was built in 1958 and later upgraded with conversion of a 35-ft grass median into continuous asphaltic paved shoulders with standard median rail and glare screen. Average weekday traffic in this portion of the freeway averaged 145 000 vehicles in 1980 (1). Average speeds in the 6:00-8:30 a.m. peak period averaged 26 mph in level-of-service F throughout the three-mile distance (according to the Houston Chronicle, March 24, 1981). Traffic was growing at a rate of more than five percent annually. Much of this growth was being absorbed by contraflow-authorized vanpools and buses that showed a 430 percent increase in ridership to 6200 peak-period trips after 22 months of contraflow operation.

Figure 2. Cover of driver-information package.

Figure 3. Notice issued to bus passengers.

North Freeway Concurrent Flow Lane

Thanks to your patience and support, the North Freeway Concurrent Flow Lane has been added to improve your a.m. commute. A project of the State Department of Highways and Public Transportation, the lane extends preferential treatment an additional 3 miles along the North Freeway. Authorized vehicles will travel "with flow" in the converted median lane and may enter at any point along its length.

The Concurrent Flow Lane is an experimental project, like Contraflow, designed to provide a short-term transit improvement and data for permanent express improvements on I-45.

METRO's Contraflow crew will monitor the lane, and the Houston Police Department will enforce lane regulations.

METRO
Metropolitan Transit Authority

Figure 4. View of CCFL near West Road.

The CCFL project was proposed to encourage more use of the contraflow lane and bypass recurring traffic congestion upstream of the morning contraflow entry.

Several other freeway characteristics favored a concurrent-flow experiment on North Freeway. The 3.3-mile segment contained excellent sight distance and only three bridge structures. Full 12-ft lanes were provided adjacent a median shoulder that averaged 16 ft except at bridge structures. Parallel frontage roads along the outer separation also provided detour opportunity in event of an incident.

Project Design

CCFL extends 3.3 miles along a segment of the median shoulder that is borrowed for 2.5 h each morning. Signing designates shoulder use by authorized vehicles only with standard diamond symbols as described in the Manual on Uniform Traffic Control Devices (MUTCD) (2). Signs are posted about 1000-ft apart on the median rail. Alternate signing is placed between authorized designations that restrict emergency parking in the median during operating periods. There are no high-occupancy vehicle (HOV) diamond pavement markings or other special striping along the lane. The surface texture of the shoulder pavement was retained as white delineation gravel on a hot-mix asphaltic base. Special pavement markings would not have been visible on the white gravel. Also, the effect of driving on the shoulder surface provided better traction and discouraged use of the shoulder as a regular travel lane. The shoulder pavement was sufficient to support a low volume of daily vehicles. A typical at-grade section of the lane is illustrated before and after implementation in Figure 5.

At bridge structures, no median shoulder previously existed, although outer shoulders were provided as shown in Figure 6. Restriping and shifting main lanes to absorb the outer shoulder created sufficient width (10 ft) to accommodate the concurrent-flow median lane over bridge decks. After the first several months of trial operation, temporary paint striping was replaced with thermoplastic markings. The slight weave in main lanes, equivalent to 10 ft over a transition length of 2000 ft, is unnoticeable to general traffic.

At the northern terminus near West Road, as shown in Figure 7, is a larger-than-standard sign in the median shoulder for authorized vehicles. No special edge striping is included that might encourage entry by general traffic. In Figure 8, the plan for treatment of at-grade separations is shown for the Mt. Houston and Gulf Bank locations along the project. The southern terminus and connection to the contraflow lane is highlighted in Figure 9. The CCFL separates from the median as an exclusive connector ramp, tying into an existing contraflow ramp used only in the evening operating period. This ramp feeds inbound vehicles over a short concurrent-flow segment on a bridge structure to the previous contraflow entry. The concurrent-flow segment of this third bridge is separated by yellow pylons from adjacent traffic. This connection is about 10 ft wide. Several gates are employed at the transition to the exclusive ramp and entry to the contraflow lane to prohibit use outside the operating period.

Operating Plan

Procedures and supervision of the CCFL are vested in the Operations Department of MTA. A crew of eight employees used to deploy the contraflow lane monitor the CCFL during operating periods and perform minor setup functions, including pylon installation and gate opening near the contraflow entry. An MTA wrecker previously located along the contraflow lane was moved upstream to the beginning of the CCFL to monitor entering vehicles and respond to incidents in both projects more effectively.

The contraflow crew spend part of their 1.5-h deployment period from 4:30 to 6:00 a.m. removing any debris and towing any stalled vehicles from the median shoulder section. TSDHPT also maintains a regular schedule for sweeping the median.

The minimal operation cost associated with daily deployment of CCFL is absorbed as part of the deployment costs of the contraflow lane. The monthly cost to MTA for contraflow averages $50 000, with approximately two-thirds of this total involving labor. Operation costs are locally funded.

Two groups of eligible CCFL vehicles--vanpools and buses--are included as potential users. In order to be authorized for the contraflow lane, rather rigid requirements must be met. Eligible vehicles include the following:

1. All MTA transit vehicles,
2. Suburban commuter buses operated under contract to MTA,
3. Other full-sized transit vehicles being operated on regularly scheduled services and approved by MTA pursuant to the requirements as listed, and
4. Vans designed to carry eight or more passengers (including driver) and approved by MTA pursuant to the following requirements.

For vehicles defined under 3 and 4 above, authorization also requires that at least eight passengers be registered, that both vehicle and driver must have met minimum insurance requirements, the driver has a good driving record and successfully completed an MTA-sponsored contraflow-training course, and the vehicle has a valid inspection sticker and contraflow-authorization decals affixed appropriately.

Probably the most important aspect of this management procedure is the issuance of authorized MTA decals that appear on the front and back windshields of each vehicle. An example decal is shown in Figure 10. These decals are highly visible to enforcement officers (the printing is black and red on a white background). This unique approach to the authorization or restriction of vehicles to a transit preferential treatment greatly simplified enforcement and provided close controls over Houston's first steps toward a regionwide transitway system.

Figure 5. CCFL at-grade implementation.

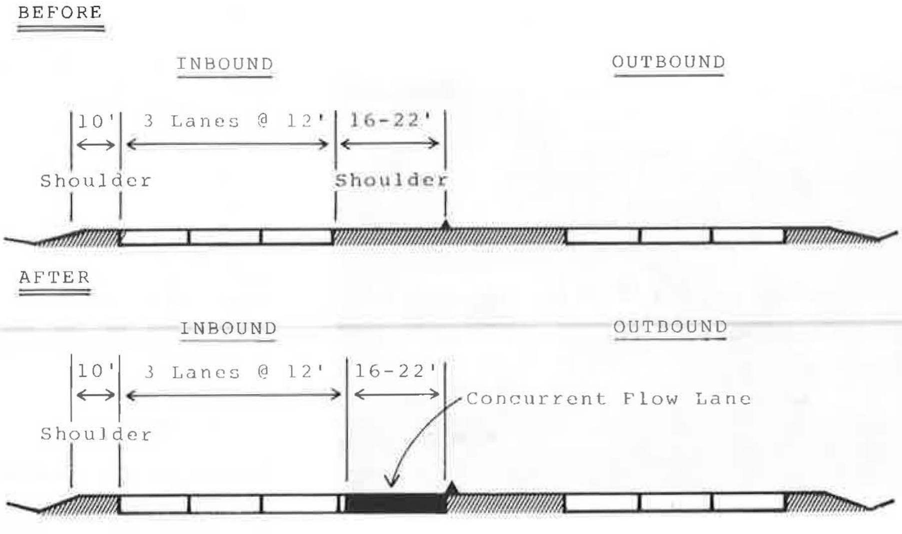
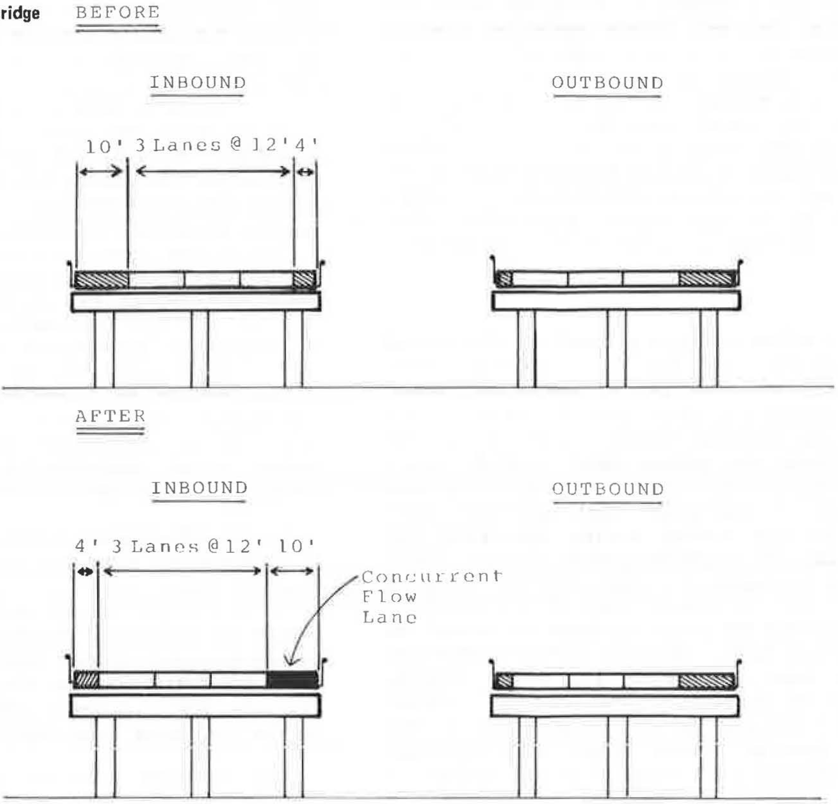


Figure 6. CCFL implementation on bridge decks.



Moreover, this approach was easily adapted to the CCFL after already being established for contraflow operation.

Unauthorized vehicles are easily identifiable if they do not display a proper decal. Police can deter or remove vehicles along the CCFL by setting up a monitoring point anywhere along the shoulder lane. A 16-ft-wide space (wider near bridge structures) is sufficient for patrols to park adjacent to the median rail without disrupting CCFL operation. The number of attempted violations have averaged from three to seven occurrences/day.

During the first month of operation, two patrols of the Houston Police Department were furnished to deter violators. They maintained fixed positions along the lane at locations of high visibility atop approaches to bridges. After a month the police were removed. In the following months, officers assigned to the contraflow-lane project provided infrequent patrols on the concurrent-flow segment. Violation rates have remained low with this support.

Operating rules for drivers of the CCFL are more rigorous than policies and procedures adopted on similar nationwide projects. A summary of these rules, extracted from the MTA driver training manual, is included below (3):

1. Operational rules: (a) Entry at any point along length of shoulder, no exit except in emergency; (b) headlights on; (c) no passing; and (d) 3-s minimum spacing between vehicle ahead and through connection ramps to contraflow entry; at

Figure 7. Concurrent-flow plan--West Houston.

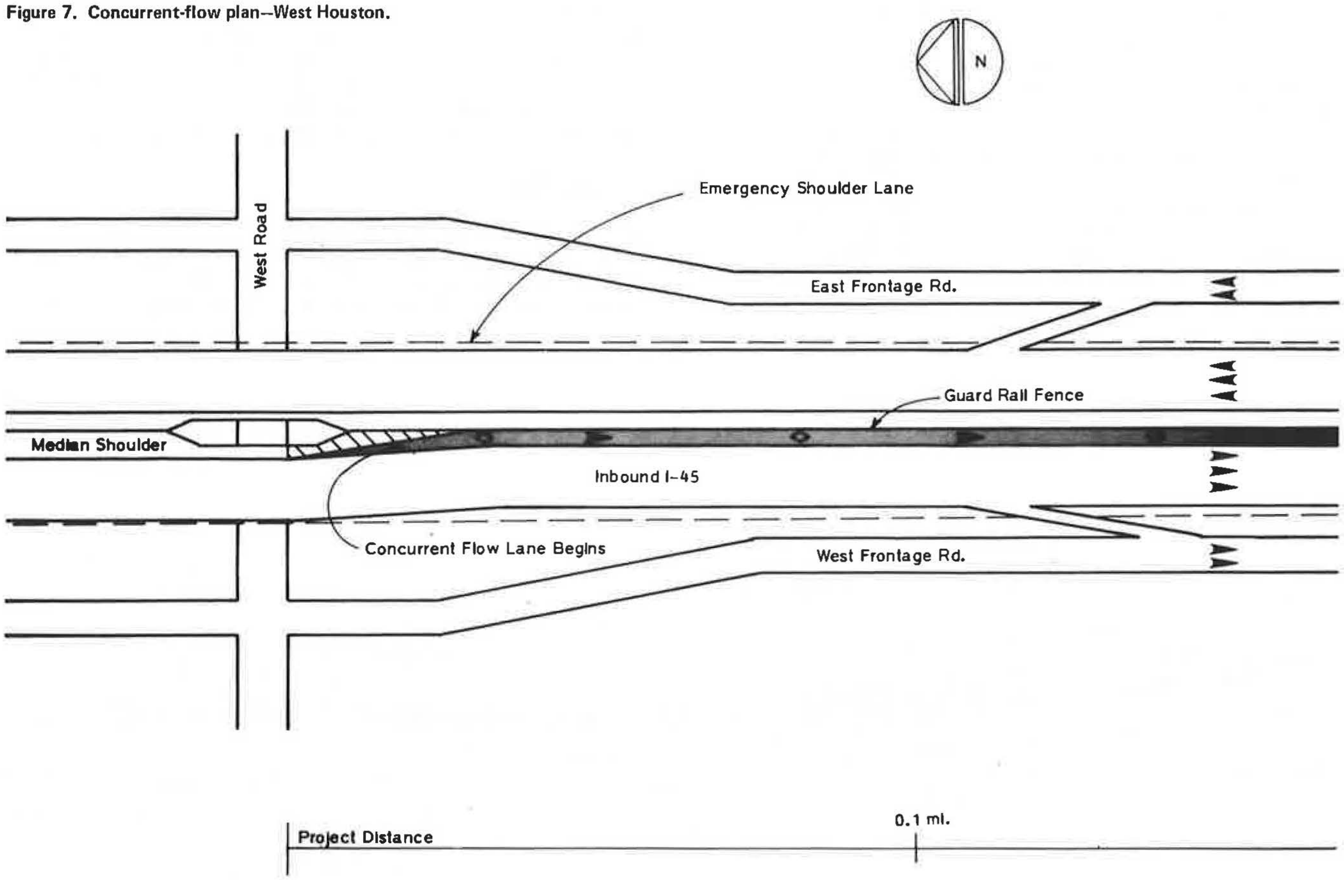
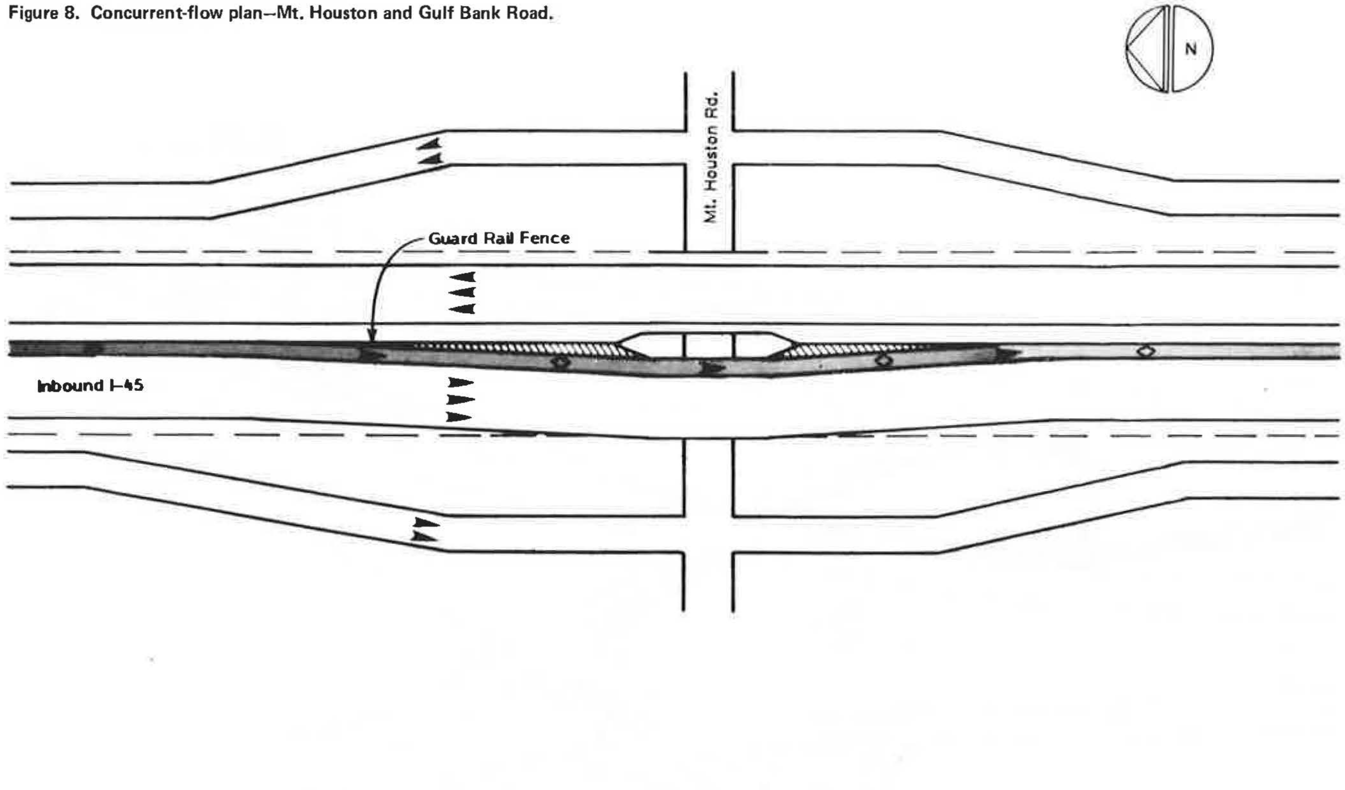


Figure 8. Concurrent-flow plan--Mt. Houston and Gulf Bank Road.



contraflow entry reduce speed to 30 mph when passing police surveillance point.

2. Entering lane: (a) Turn on headlamps; (b) enter from leftmost travel lane next to median shoulder, checking for oncoming vehicles in shoulder; (c) use turn indicator; and (d) yield to oncoming vehicles to your left in shoulder as you merge.

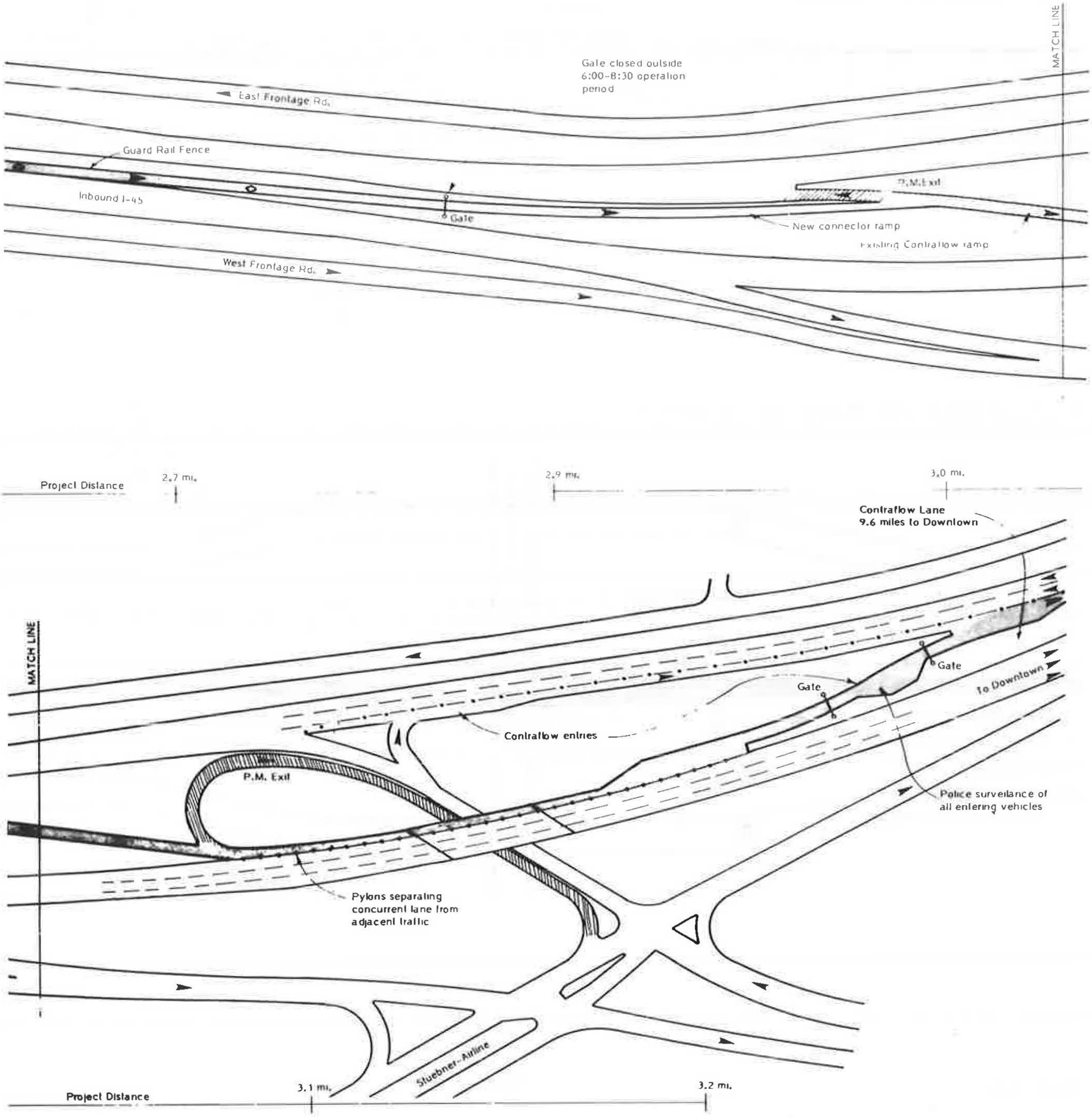
3. Entering contraflow lane: Enter the median shoulder lane or enter via the North Shepherd and Stuebner-Airline ramp. (There is no longer an entry to contraflow from I-45 at the North Shepherd overpass.)

4. Negotiating a disabled vehicle: (a) Slow at least 200 ft behind disabled vehicle, (b) signal right-turn indicator to merge back into leftmost travel lane, (c) maneuver around disabled vehicle, and (d) reenter shoulder lane by using left-turn indicator and carefully merge from leftmost lane.

Project Use

During the first three months of project operation, use of the concurrent-flow shoulder lane grew from 250 to 280 vehicles/operating period (4). Vehicles were composed of 22 percent buses and 78 percent

Figure 9. Concurrent-flow plan at termination.



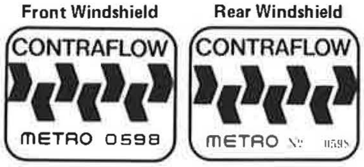
vanpools. A more detailed record of vehicle use along the CCFL is presented in Figure 11 (4) and the table below (4):

1981	Beginning, West Road	Midpoint, Mt. Houston	Termination, Contraflow Entry
February (before CCFL)	151	173	190
April	207	236	249
May	242	254	259
June	254	265	281

Figure 11 illustrates the growth in use at three locations along the lane. Total vehicles increased an average of 14 percent along the lane. This increase was commensurate to increases on the contraflow lane during the previous three months. Since entry to the shoulder is unrestricted, vehicles (particularly vanpools) load onto the lane throughout its length from various freeway feeder ramps. Few buses enter the freeway throughout this segment. There is no user demand for exiting the lane. Before the CCFL project, 21 percent of the users were entering the freeway within the project limits. After three months only 10 percent were entering the freeway downstream of the beginning of the CCFL. Apparently a number of vanpools have shifted travel patterns from parallel arterials to benefit from this project.

The extent of diversion among vanpools is presented in more detail in Figure 12 (4). There are two entry points to the contraflow lane. The first is directly from I-45 inbound via the CCFL; the other is via Stuebner-Airline, a major surface arterial. In a December 1980 origin-destination survey, 78 percent of all vanpools entered the contraflow lane via the I-45 crossover. After CCFL implementation in May, 95 percent were entering from the CCFL. Similar increases in the percentage of total contraflow vanpools loading onto I-45 farther upstream are indicated. No information has been collected after CCFL implementation to determine specific changes on parallel arterials.

Figure 10. Authorized vehicle decals.



This distribution of vehicle volumes in the CCFL is not uniform throughout the operating period. As illustrated in Figure 13 (4), as much as two-thirds of the total volume travels in the lane in the peak hour (6:30-7:30 a.m.). This distribution is similar to earlier distributions made during the first two years on the contraflow project farther downstream. The profile of users by time has not apparently changed due to increased growth or characteristics of enforcement.

RESULTS

Travel Speeds and Time Savings

Travel speeds presented were made between the beginning of the CCFL (West Road) and just upstream of lane termination (Gulf Bank Road), an effective length of 3 miles. Speeds in the main lanes adjacent to the CCFL were determined by floating-car studies. Studies were made during the first month of operation and are presented in Table 1 (4). Vehicles in the CCFL averaged 48 mph while vehicles in main lanes averaged 26 mph. Travel times for this distance are as follows: CCFL, 3.75 min and main lanes, 6.92 min. Thus, a net savings of 3.17 min/user was initially realized, which resulted in a daily savings of about 190 person-h. Travel-time studies are continuing at this time. It is expected that daily savings will continue to increase as a factor of growth in use.

Breakdown Incidents on CCFL

Deployment of the contraflow-project wrecker to respond to incidents on both projects has worked well. The wrecker sweeps the CCFL in advance of each operating period, removing any disabled vehicles. The wrecker is within several minutes response time of a breakdown incident on the CCFL by being stationed at the beginning of the lane. About 10 breakdown incidents have been logged during the first three months, but no incidents have involved authorized vehicles. Because of the width of the

Figure 11. Vehicle use of CCFL.

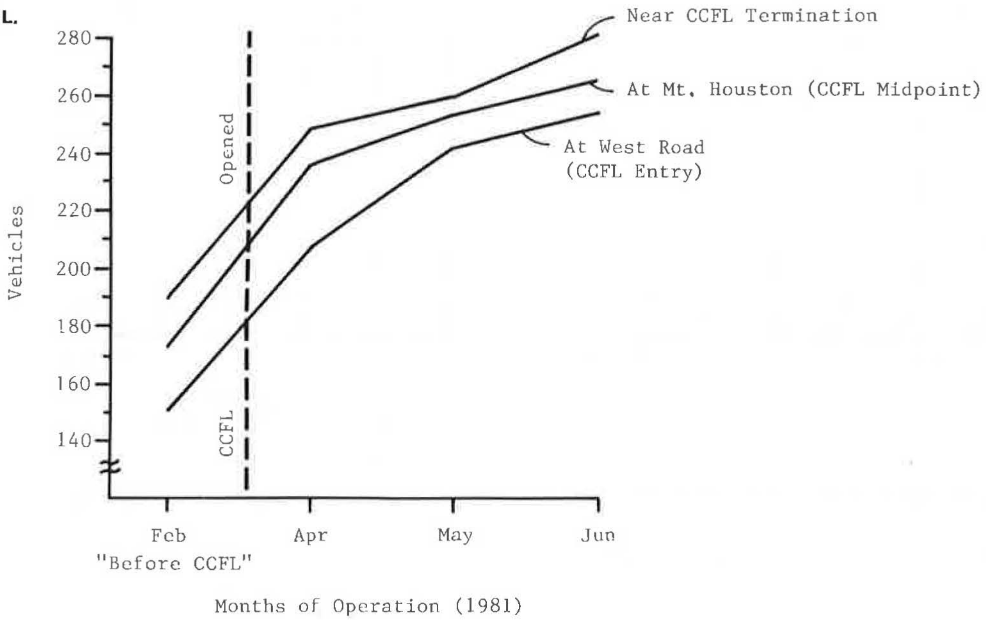
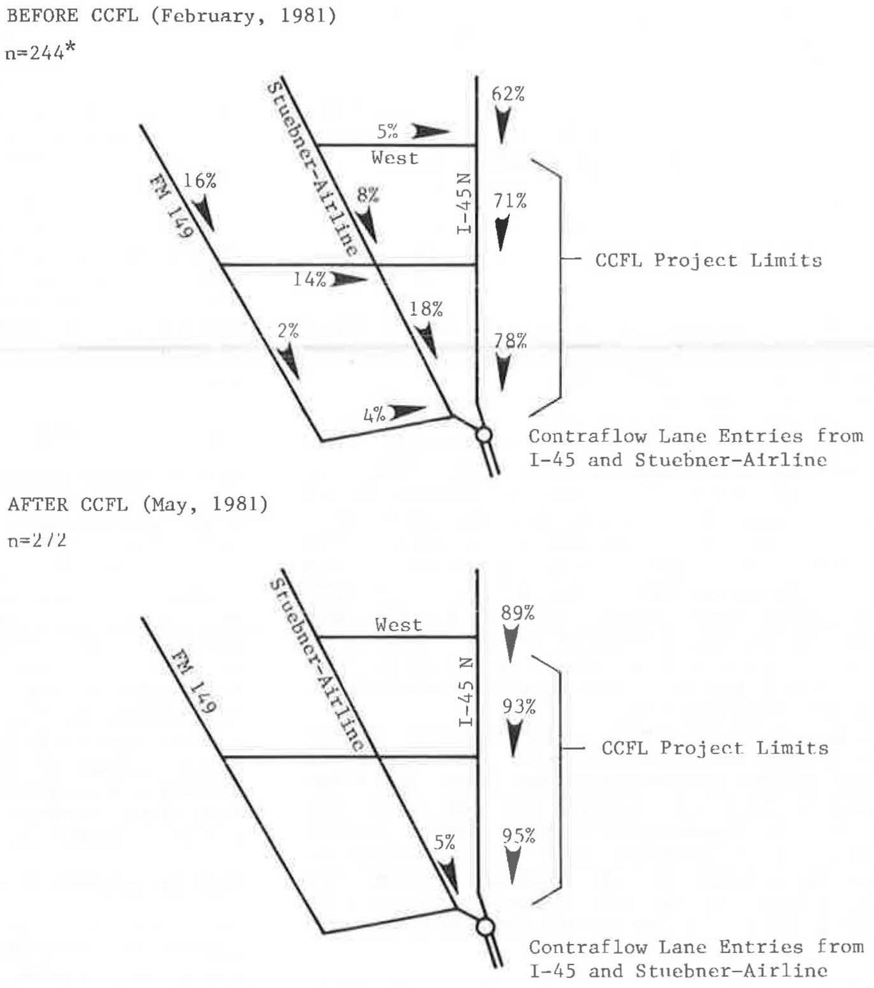


Figure 12. Changes in vanpool routes after implementation.



*Origin/Destination sample size was 175 (December, 1980).

Figure 13. Vehicles traveling in CCFL.

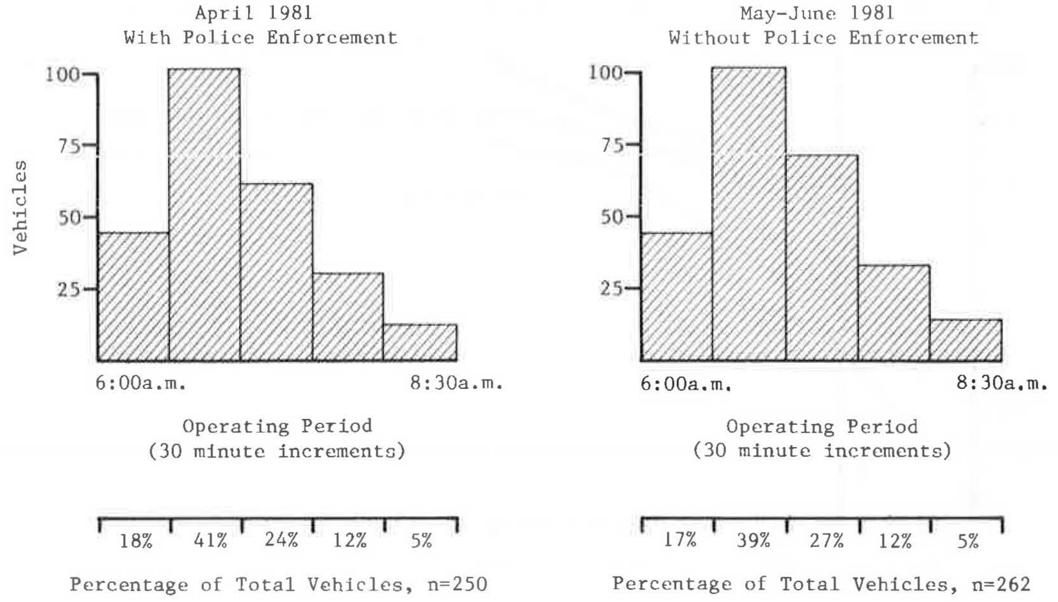


Table 1. Average speeds in CCFL and main lanes.

Date	6:00-6:30 a.m.		6:30-7:00 a.m.		7:00-7:30 a.m.		7:30-8:00 a.m.		8:00-8:30 a.m.		Average	
	CCFL	ML	CCFL	ML	CCFL	ML	CCFL	ML	CCFL	ML	CCFL	ML
4-7-81	35	21	47	27	39	18	44	19	45	27	42	22
4-10-81	50	29	48	28	49	20	41	24	44	35	49	27
4-13-81	48	33	46	20	45	17	47	21	50	26	47	23
4-16-81	51	35	51	35	47	19	48	27	56	45	51	32
4-20-81	48	27	50	20	45	20	49	21	53	32	49	24
4-22-81	50	27	44	29	49	20	52	17	56	31	50	25

Notes: ML = main lanes.
For CCFL, $\bar{X}$ = 48 and S = 3; for ML, $\bar{X}$ = 26 and S = 4.

Table 2. Number of violators by location.

Date	West Road	Mt. Houston	Gulf Bank Road
With Enforcement[a]			
4-7-81	2	3	3
4-10-81	0	1	2
4-13-81	1	2	1
4-16-81	0	4	2
4-20-81	0	0	0
4-22-81	0	3	2
4-24-81	9	7	5
Without Enforcement[b]			
5-5-81	5	10	3
5-6-81	7	7	5
5-14-81	12	13	5
5-21-81	2	6	3
6-19-81	8	8	7
6-22-81	5	2	4

[a] $\bar{X}$ = 2, 3, and 2 and S = 3.0, 2.3, and 1.6 for West Road, Mt. Houston, and Gulf Bank Road, respectively.
[b] $\bar{X}$ = 7, 8, and 4 and S = 3.4, 2.9, and 1.8 for West Road, Mt. Houston, and Gulf Bank Road, respectively.

median, authorized vehicles have either negotiated incidents within the shoulder or merged into adjacent lanes and reentered the CCFL. Slow travel speeds in adjacent lanes usually permit this infrequent maneuver without difficulty.

Violations

During the first month (March 30 to April 27), two patrols of the Houston Police were stationed on the CCFL. Violations during this period are presented in Table 2 (4).

Survey data were collected over seven days during this period, yielding a range of up to nine violations being sited at any given location during an operating period. Violators were sited more often near the beginning of the lane at West Road where police seldom were stationed, but observances were irregular. If an average of 2.3 sitings/day is used throughout the project, the percentage of violators to authorized users represented less than 1 percent of daily vehicles during police enforcement.

During the following months, police enforcement was removed. Average sitings of violators increased slightly, varying from an average of 1.8 to 3.4 observances/location daily and reflecting less than 2 percent of all traffic in the median.

Accidents

The CCFL recorded two minor accidents through June. Both involved minor property damage and involved unauthorized vehicles accidentally or intentionally entering the median shoulder and colliding with an authorized vehicle traveling in the lane. To date the two incidents reflect the only reported accidents during an initial period when 55 000 vehicle miles were logged on the shoulder lane.

There has been no observed change in accidents on adjacent lanes. Information compiled from accident records over the length of the project [taken from State Accident Statistics--Region 4, Educational Service Center (February-May 1981)] is given in the table below for accidents coded for the 8000-10 300 blocks of North Freeway, 6:00-8:30 a.m. (note: inbound accidents are traffic adjacent to the CCFL operation):

	Inbound		Outbound	
Condition	Date	Time (a.m.)	Date	Time (a.m.)
Before	3-6-81	6	3-4-81	7
(February-March)	3-11-81	6		
After				
(April-May)	4-9-81	8	5-6-81	6
	5-20-81	6		
	5-20-81	8		

Only three accidents were reported during the first two months compared with two accidents in a similar 60-day period prior to the opening of the project. This reflects an accident rate of 1.1 accidents/million vehicle miles (MVM) before to 1.7 accidents/MVM after. Rates in the outbound side remained unchanged at about 0.9 accident/MVM in free-flow conditions. Note that this evaluation is rather limited, encompassing only a 60-day period before and after the project. An extended period of operation is needed to more fully assess accident impacts to users and nonusers.

Costs and Benefits

An initial cost/benefit analysis is presented to provide some indication of the relative significance this project has made after a rather short period of operation. Costs of the project include initial construction ($130 000) and daily operation. Manpower requirements needed for the adjoining contraflow-lane project are used to monitor and provide wrecker response on the CCFL. There may be some additional manpower needed, but this cost is absorbed within regular crew shifts required by the contraflow operation. Special police enforcement is no longer used; thus, no enforcement cost is considered necessary at the present time.

The benefits to users involve an initial travel-time savings of 190 person-h. There may also be some savings to main-lane drivers when an average of 250 authorized vehicles are removed, but these impacts are probably negligible when compared with significant latent demand that exists in the corridor for freeway capacity.

A reduction in travel time does reflect a reduction in fuel consumption. This project expedited movement of about 3900 persons in 250 vehicles. If it is assumed that without the CCFL these commuters would be riding in private automobiles with an average vehicle occupancy ratio of 1.4 persons/vehicle, the resulting fuel savings would be about 112 000 gal of gasoline/year, assuming a fuel-consumption rate of 17 miles/gal.

COMPARISON WITH OTHER FREEWAY CONCURRENT-FLOW PROJECTS

Concurrent-flow reserved lanes have been demonstrated on freeways elsewhere, and similar projects are currently operating in Portland, Miami, Honolulu, and the San Francisco area (5-7). There have also been unsuccessful demonstrations in Boston and Los Angeles (8). These projects have been the subject of intense review and evaluation. Applications to date have been criticized because they have experienced increased accident rates and are difficult to enforce. Recommendations made from this experience include the following:

1. A CCFL should only be implemented in conjunction with the addition of a freeway lane; a general-traffic lane should not be designated for HOV use;
2. The CCFL should span a location of normal peak-period freeway congestion; otherwise, HOVs will not receive an adequate travel advantage and will have difficulty merging into and out of the lane;
3. Project implementation should be preceded by a vigorous public-information campaign;
4. Project implementation should contain a thorough, well-planned enforcement program; monitoring by motorcycle officers is encouraged for mobility and selective enforcement techniques are recommended as a minimum approach to violation control, with continuous special enforcement encouraged to achieve a desired level of motorist compliance;
5. The project should contain median shoulders or refuge areas for public safety and enforcement monitoring; and
6. Signing and markings should conform to MUTCD standards to reduce driver confusion.

Most design-related recommendations were incorporated into the I-45N project. These included the following:

1. The CCFL was borrowed from a previously designated emergency shoulder; thus, no general-traffic lane was affected;
2. The length and period of designation were specifically selected to bypass a recurring segment of traffic congestion;
3. Median signing conformed to MUTCD standards; lane markings were not included because of the rough white gravel texture of the shoulder and contrast difficulties; and
4. Refuge areas for vehicle breakdowns and enforcement monitoring were available in wider portions of the project near bridge approaches.

Operational recommendations based on experience from other projects were carefully considered by the CCFL management team. Variances from recommendations were made where appropriate to meet specific needs of this project and develop conformity between concurrent-flow and contraflow project operating rules and regulations. Several examples of these variations with accompanying justifications follow:

1. Project implementation and opening were not widely advertised and covered. Because a specific user group on the contraflow lane had already been identified and authorized, there was no incentive to initially seek more users. Second, the project was implemented as an experiment. As such, the management team could more easily modify or terminate this experiment without affecting a previous success record in HOV applications if the project were introduced without a vigorous information campaign. Finally, the earlier implementation of the contraflow lane on the same freeway already familiarized motorists with the objectives of HOV preferential treatment and the term authorized users.
2. Enforcement was considered a necessary part of CCFL operation but no rigorous program could be implemented. Cooperation from the Houston Police Department was requested and received for the first 30 days of project operation. Following this period a series of steps were taken to ensure compliance. Existing police patrols under contract to MTA were asked to expand their coverage area to include the CCFL on a selective basis. Authorized users were asked to report observed violators to MTA. Reports were followed up by correspondence to vehicle owners. These steps have been effective in keeping observed sitings below 2 percent of lane use.
3. Rules and regulations for operating in the CCFL include use of headlights, procedures for maneuvering around a stalled vehicle, and no exiting except on the termination into the contraflow lane. These procedures are more rigorous than other CCFL projects, as are the authorization procedures. The management team felt that retention of the more stringent policies adopted for the contraflow lane would better ensure safe operation on the CCFL.

A comparison of operating characteristics of the Houston, Miami, San Francisco, and Portland projects is included in Table 3 (5-7,9). The Houston project nets the best comparative travel-time savings shown at 1.1 min/mile. Initial use is second only to US-101 in San Francisco. Occupancies before project implementation were impacted by contraflow-lane operations that were initiated 19 months earlier.

CONCLUSIONS

The concurrent-flow shoulder lane operation on I-45N in Houston has made a perceptible improvement in user travel time without impacting adjacent traffic characteristics. A sizable number of authorized users have rerouted their trips to benefit from this improvement. The $130 000 cost for construction, funded entirely from local sources, has reflected very high cost benefits. Enforcement procedures have been minimized and accident rates have not detrimentally affected the CCFL or adjacent lanes after three months of operation.

These findings are in variance from the consensus of experience collected nationwide on the applicability of a CCFL on freeways. The success experienced in Houston may be a result of the following unique characteristics:

1. Facility design--The CCFL did not remove a regular traffic lane from the general public. The shoulder lane was wide enough in places to facilitate police monitoring and apprehension of violators. The surface texture of rough delineation gravel was retained, thereby alleviating potential preceptions of shoulder use for a regular travel lane. The CCFL also turned into contraflow preferential treatment downstream at a location under constant police surveillance.
2. Management control--MTA maintained stringent authorization procedures for CCFL eligibility. A highly visible windshield sticker was required as

Table 3. Operating characteristics of concurrent-flow HOV projects.

	Miami, I-95			San Francisco, US-101			Portland, Banfield Freeway		Houston, North Freeway	
Variable	Before	Bus and 3-Person Carpool	Bus and 2-Person Carpool	Before	Bus Only	Bus and 3-Person Carpool	Before	Bus and 3-Person Carpool	Before	Bus and Vanpool
Critical peak period	4:00-6:00 p.m.	4:00-6:00 p.m.	4:00-6:00 p.m.	4:00-7:00 p.m.	4:00-7:00 p.m.	4:00-7:00 p.m.	7:00-8:00 a.m.	7:00-8:00 a.m.	6:00-8:30 a.m.	6:00-8:30 a.m.
Length of HOV lane (miles)	-	6.7	6.7	-	3.7	3.7	-	3.3	-	3.3
Total peak directional lanes	3-4	4-5	4-5	3	4	4	2-3	3-4	3	4
No. of HOV lanes	-	1	1	-	1	1	-	1	-	1
Volume										
All lanes	11 355	12 825	15 290	13 600	13 137	13 098	3557	4025	12 382	12 600
HOV lanes	-	618	2057	-	191	647	-	203	-	262
HOV lanes (bus only)	-	23	23	-	148	150	-	23	-	58
HOV lanes per total volume (%)	-	4.8	13.5	-	1.5	4.9	-	5.0	-	2.1
Vehicle occupancy (persons per vehicle)										
All lanes	1.28	1.37	1.42	1.30	1.30	1.36	1.22	1.26	1.62[a]	1.70
HOV lanes	-	2.23	1.79	-	2.21	2.96	-	2.81	-	16.0
Person throughput										
All lanes	14 875	18 221	22 338	24 439	24 567	25 365	4329	5611	12 723	13 461
HOV lanes	-	1981	4347	-	5719	7172	-	1067	-	4194
HOV lanes per total throughput (%)	-	10.9	19.5	-	23.3	28.3	-	19.0	-	31.2
Speed (mph)										
General lanes	29.6	35.6	41.6	34.1	43.3	47.6	38.2	37.9	26	26
HOV lanes	-	50.0	50.4	-	53.4	53.4	-	51.5	-	48
Travel time (min)										
General lanes	13.5	11.3	9.6	6.5	5.1	4.7	5.2	5.2	7.0	6.9
HOV lanes	-	8.0	8.0	-	4.2	4.2	-	3.8	-	3.7
Accident rate per MVM	5.1	4.7	2.4	4.2	9.6	12.8	0.9	0.8	1.1	1.7

[a] Includes contraflow buses and vanpools previously traveling in all lanes.

part of the authorization procedure. Signing restricted CCFL use to authorized vehicles, and sticker visibility, in addition to vehicle appearance, helped police enforcement and public recognition of violators.

3. Public attitudes--Although the CCFL was the first concurrent-flow freeway application in Houston, it was not the first preferential treatment project. North Freeway commuters were exposed to the contraflow concept over a 9.6-mile segment of the freeway in August 1979. The definition for authorized vehicles was not new to commuters. Many people, including the news media, called the CCFL an extension of the contraflow project. This sequence of staging the concurrent-flow experiment after contraflow may have improved the chances for public acceptance of the concept.

After three months of project operation, the general conclusion of the TSDHPT-MTA project management team, the users, and the public is that the North Freeway CCFL has proved successful. The level of use and its continued increase have met expectations. The fact that an additional 190 person-h are saved daily and 260 buses and vanpools have been afforded exclusive access around a congestion bottleneck has enhanced transit and vanpooling as a desired alternative to the automobile in the corridor. Both the CCFL and contraflow projects to date have accomplished a daily savings of 3300 person-h and removal of 4500 automobiles from peak-direction traffic, significantly impacting expectations for regional transitways on many of Houston's corridors in the future.

I hope that the information presented substantiates the initial conclusions drawn regarding the concurrent-flow application in Houston. The project will continue to be monitored and modified by the management team, as appropriate, until such time that a more permanent transitway facility can be incorporated into the North Freeway.

REFERENCES

1. I-45 Contraflow Interim Report. Metropolitan Transit Authority, Houston, May 1980.
2. Manual on Uniform Traffic Control Devices. FHWA, 1978.
3. Addendum to Driver Training Manual. Metropolitan Transit Authority, Houston, March 1981.
4. Summary of Operations of the Concurrent Flow Lane. Texas Transportation Institute, Texas A&M Univ., College Station, Draft Rept., July 1981.
5. C.A. Fuhs and others. Operational Experience with Concurrent Flow Reserved Lanes. Texas Transportation Institute, Texas A&M Univ., College Station, Rept. 205-4, July 1977.
6. Beiswenger, Hoch, and Associates. Enforcement Requirements for High-Occupancy Vehicles. FHWA, Dec. 1978.
7. Beiswenger, Hoch, and Associates. Safety Evaluation of Priority Techniques for High-Occupancy Vehicles. FHWA, Feb. 1979.
8. H.J. Simkowitz. A Comparative Analysis of Results from Three Recent Non-Separated Concurrent Flow High-Occupancy Freeway Lane Projects: Boston, Santa Monica, and Miami. Cambridge Systematics, Inc., Transportation Systems Center, Cambridge, MA, Draft Rept., Jan. 1978.
9. Beiswenger, Hoch, and Associates; Texas Transportation Institute. Occupancy, Volume, and Park-and-Ride Surveys. Metropolitan Transit Authority, Houston, Jan. 1981.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Review of Bus Costing Procedures

WALTER CHERWONY, SUBHASH R. MUNDLE, BENJAMIN D. PORTER, AND GREGORY R. GLEICHMAN

With changing policies regarding transit funding at all levels of government, transit planners will be required to monitor more carefully existing bus systems as well as examine intensively proposed service changes. A key aspect of this responsibility will be an assessment of transit finances. During the past two decades, the focus of research has been placed on the estimation of demand and revenue. In the next few years, increasing efforts will be directed to the estimation of bus operating costs and the underlying relation that impacts expenditures. A discussion of various procedures and techniques that have been developed and applied in the past to estimate operating costs is presented. The methods have been grouped to form broad generic types, which in turn have been subdivided further by unique approaches. To illustrate the present state of the art, each approach has been illustrated by a single model. This cost-estimation review clearly indicates the evolutionary nature of cost-estimation procedures. The latest research efforts are typically more accurate and sensitive to drivers' wages and work rules that reflect the labor-intensive nature of bus transportation. It is anticipated that an understanding of the prevailing cost-estimation procedures will aid transit planners in their activities and enable them to contribute to the literature on costing procedures.

Almost every transit system today has established a mechanism to monitor existing bus service performance and conduct service planning in a systematic fashion. The techniques and approaches vary widely; some systems perform cursory reviews of their needs and others use sophisticated techniques to perform detailed operations and planning activities. A key element of this analysis involves estimating the costs to provide present service as well as computing the cost impacts of proposed service changes. This need has become acute due to the limited financial resources of all public services, including public transportation. More than ever, transit managers are focusing their attention on improving the productivity, effectiveness, and efficiency of their transit systems. A key component of this new cost consciousness is a strong interest in developing a technique that accurately estimates the cost of present routes and the cost of proposed service changes.

Recognizing this need, the Urban Mass Transportation Administration (UMTA) has commissioned Booz, Allen and Hamilton to develop a uniform technique or set of techniques that will accurately reflect the cost of providing bus service. An initial step in this study is a review of cost-estimation techniques that have been used previously in the transit industry. The objective of this paper is to present the results of this review and provide an overview of techniques and procedures that appear in the technical literature.

GENERIC TYPES OF COSTING MODELS

To provide an analytical framework for review, the various estimation techniques were catalogued into several generic types. Some techniques are combinations or hybrids of more than a single generic type. For purposes of this paper, each procedure has been designated as representative of a particular genre based on the model's concentration of effort. No simple classification system can account for the various permutations of cost models presented in the literature. However, three generic types of models are prevalent, as described below:

1. Causal factors: The causal-factors approach is similar in nature to the preparation of a bid estimate for a construction project. Various quantities required to provide bus service, such as drivers, buses, fuel, tires, etc., are estimated and multiplied by an appropriate unit-cost factor. The products of each quantity estimate and unit cost are summed to arrive at the transit cost.
2. Cost-allocation model: The cost-allocation technique appears widely in the literature as a means to disaggregate system costs into individual route expenditures. Unlike the causal-factors approach, transit costs are estimated on a top-down basis. The key assumption of this approach is that each operating expense item can be assigned or allocated to a specific operating statistic such as vehicle miles. Unit costs are developed that comprise the coefficients of the cost-allocation model.
3. Temporal variation: Many researchers have concentrated their analyses on the differences in costs for providing service by time of day or day of week. By analyzing the underlying relations that influence bus costs, an attempt is made to quantify the temporal variation in costs. Because the emphasis of this research is usually on drivers' wages, these techniques often embrace other generic types to estimate nondriver expenditures.

Regression models were reviewed in this study but are not fully described here. Typically, the regression approach has been applied to identify the underlying relations that influence transit costs rather than to compute the cost of existing routes or to determine the incremental cost associated with service changes.

A three-level hierarchy for classifying the models was developed to aid the following discussion. First, model groups are classified by their generic type. A generic type, as described above, is a grouping of approaches that all share one distinctive characteristic. Each of the next three sections of the paper covers a single generic type. Second, model groups are classified by their approach. An approach is a grouping of models that generally use a similar technique but vary at the detailed level. Finally, each model is discussed in the context of its generic type and approach. Models are distinct techniques developed by a single researcher or research team.

Another point to note is that transit operating expenditures can be described in four ways--fixed, variable, average, and marginal cost. For the most part, these cost categories and nomenclature are drawn from economics and accounting and are not unique to the transit industry. It should be recognized that some authors differ in their use of these terms. To facilitate a uniform nomenclature, the following definitions are used:

1. Fixed costs: Fixed costs are those expenses that do not vary with the level of production. In bus systems, this means that these costs are unchanged with respect to the number of hours, miles, or buses operated. Fixed costs typically include costs such as general manager salary and maintenance expenses for buildings.
2. Variable costs: Variable costs are those costs that do vary with the amount of service provided. These expenses would include costs for fuel, drivers' wages, and a host of transit operating costs. The differences between fixed and variable costs are portrayed in Figure 1.
3. Average cost: As the name implies, average cost is merely the cost divided by the level of output. As shown in Figure 1, the average cost at output level O_1 is merely the slope of the line from the origin (C_1/O_1). Similarly, at output level O_2, the average cost is C_2/O_2.
4. Marginal cost: Sometimes referred to as incremental cost, marginal cost refers to the additional costs associated with an increase in the level of output. As shown in Figure 1, it is merely the change in costs $(C_2 - C_1)$ associated with a change in output level $(O_2 - O_1)$.

CAUSAL FACTORS

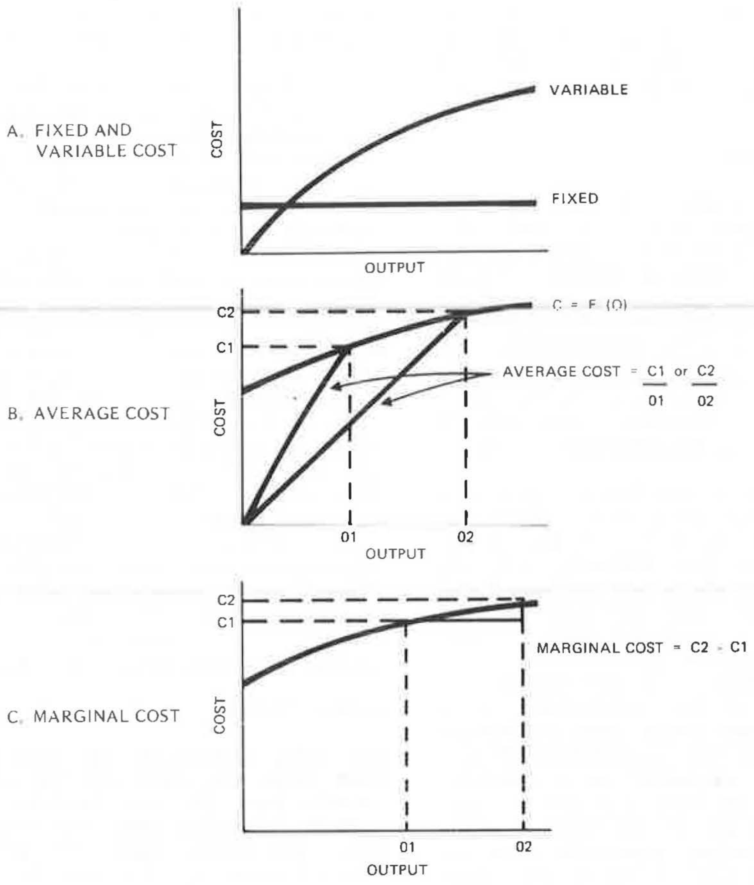
The idea underlying the causal-factors method is that total bus costs are the sum of the individual amounts paid for each resource item consumed. For example, resource items may include drivers' wages, tires and tubes, fuel, oil, and repair parts. The cost of each resource item is found by multiplying the quantity consumed by the unit price or unit cost of the item.

The causal-factors method, not being unique to bus costing, is well known and understood. The process is analogous to the cost-takeoff procedure used in the construction industry and is similar to the budgeting process used in almost all industries. The method is distinguished by the large number of resources included in the cost equation. Note that by selecting which cost items are included in the analysis, the issue of fixed and variable costs can be addressed as well as the incremental out-of-pocket expenses for a specific service change.

An example of this method for a single resource item can be illustrated with expenditures for drivers' wages. For example, a service change that requires an additional 80 vehicle-h daily would first be converted to hours paid based on productivity statistics. At a productivity rate of 1.5 h paid/vehicle-h, 120 pay-h would be required. Based on an average hourly wage of $7.50/h, the driver cost of the service change would be $900/day. In a similar manner, other expense items could be addressed with the causal-factor method.

Approaches that emphasize more accurate estimation of the driver labor resource requirement through detailed scheduling represent a subset of the causal-factors method. Schedule making may be facilitated through the use of computer programs such as RUCUS or other programs that offer simplifications of the driver assignment task. Once the labor requirement has been found by using one of these detailed approaches, the results are used as inputs to a cost model. Thus, detailed scheduling cannot stand alone as a cost-estimation method but can be regarded as an optional step within the causal-factors method.

Figure 1. Cost descriptions.



COST ALLOCATION

The basic concept underlying the cost-allocation method is that the cost of a route or service is a function of a few resource quantities. In the cost-allocation sense, resources are aggregate measures of transit service, such as vehicle miles, vehicle hours, and peak vehicles (1). For example, a commonly used cost-allocation model takes the following form:

$$C = U_H(VH) + U_M(VM) + U_V(PV) \quad (1)$$

where

C = cost of route,
U_H = unit cost per vehicle hour,
VH = vehicle hours of route,
U_M = unit cost per vehicle mile,
VM = vehicle miles of route,
U_V = unit cost per peak vehicle, and
PV = peak vehicles used on route.

The unit costs are found by completing three tasks. First, each expense object (e.g., drivers' wages, fuel) is assigned to one or more resource variables (e.g., vehicle hours). Second, the expense objects assigned to each resource are summed to obtain the overall cost assigned to that resource. Third, unit costs (e.g., cost per vehicle hour) are derived by dividing the overall resource cost by the quantity of that resource. The method received its name because it is commonly used to allocate total system costs to individual routes on a proportional basis (2).

The cost-allocation method differs from the causal-factors method in that it is a top-down approach. In the causal-factors approach, for instance, unit costs are based on actual market prices for specific items. In contrast, the cost-allocation model derives unit costs from system expense-account data and operating statistics. Unit costs for the cost-allocation model, then, are not defined in terms of goods normally purchased. For example, transit systems do not buy vehicle hours in the same sense that they buy diesel fuel. Rather, unit costs represent the cost for providing some aggregate measure of transit service. Although some could be considered input measures, such as peak vehicles, others are more accurately termed output measures, such as vehicle miles.

Two approaches have been followed in the development and application of cost-allocation models. The first is denoted fully allocated in that all operating costs are included. Another approach, favored by British bus systems, is the fixed-variable procedure. In this latter approach, costs are stratified by whether they are fixed or variable.

Fully Allocated

The first step in applying the fully allocated approach is selecting the resource variables for inclusion in the model. This step effectively defines the number of terms in the model's equation. For illustrative purposes, the following discussion is based on the application of a three-variable cost-allocation model to the Birmingham-Jefferson County Transit Authority. The Birmingham application used the model form presented earlier.

The second step in this approach is to derive unit costs. As described previously, three tasks are involved. First, one must assign the expense accounts to the resources. The following discussion illustrates the rationale used to make some of the assignments:

1. Vehicle hours: Employees engaged in operating the vehicles are, of course, paid on an hourly basis. Thus, the assignment of this wage expense is properly made on the basis of hours of service. Likewise, other expenses that are related to service hours, such as supervision of transportation operations, are assigned to this category.
2. Vehicle miles: Many costs are related directly to the miles of operation of each route. Expenses such as fuel, tires, parts, and maintenance of revenue equipment are a direct function of the number of miles operated.
3. Peak vehicle needs: Many individual expense items do not vary as functions of either of the foregoing parameters--vehicle miles or vehicle hours. Rather, many overhead expenses are related to the scale of the system. Peak vehicles provide a reasonable measure to assess certain cost consequences of orienting the transit system to peak requirements of service.

By summing the expenses assigned to each resource and then dividing by the appropriate operating statistic, the unit-cost coefficients of the model are determined. The calibrated model for the Birmingham example is as follows:

$$C = 9.34\,(VH) + 0.32\,(VM) + 3459\,(PV) \quad (2)$$

Although the preceding discussion has centered on a three-variable model, other fully allocated models have used more or less variables. The resources used to define the variables also differ from model to model. No matter what number or type of resources are used, the basic algorithm for all fully allocated models is essentially the same as that described for the three-variable model. Only minor modifications are necessary to accommodate the additional (or deleted) variables. It should be noted that average costs, such as $2.25/mile or $23.50/h, represent the simplest cost-allocation model--one with a single variable.

Fixed-Variable Procedure

The fixed-variable cost-allocation models differentiate between fixed and variable costs. Such models modify the fully allocated approach by classifying each expense account as either a fixed or variable cost (3). Once classified, unit costs can be derived from the expense accounts in two dimensions: (a) according to resource, as is done with the fully allocated approach, and (b) according to cost classification.

As noted previously, this approach is typically employed by British bus systems. Frequently, the fixed costs have been stratified into two cost types--variable overhead and fixed overhead. An example (4) of a cost model for the Merseyside Bus Company is presented below:

	Cost Type		
Cost Variable	Direct	Variable Overhead	Fixed Overhead
Vehicle hours	1.08	0.39	0.82
Vehicle miles	0.03	0.04	--
Peak vehicles	--	53.53	22.35

One attractive feature of the fixed-variable cost-allocation-model approach is that it can be used to allocate existing bus route costs as well as estimate the incremental costs associated with proposed service changes.

TEMPORAL VARIATION

It is generally accepted in the transit industry that the cost of peak-period service is higher than the cost of base-period service. Costing models that specifically address this variation of peak and base costs have been termed temporal-variation models. Temporal cost variation arises from two sources: (a) the labor cost differential associated with labor agreement provisions that specify wages and work rules and (b) the vehicle cost differential associated with supporting peak-period vehicle requirements. All temporal-variation models focus on the first source, since labor costs are by far the single most significant component of operating cost. However, several models also treat the vehicle cost differential, although in a less complex manner.

The focus on labor costs takes the form of a detailed examination of productivity and wage costs for each period of the day and, in some cases, day of the week. Productivity is typically viewed in terms of the number of driver pay hours required to provide a platform hour of bus service. Generally, the ratio of pay hour to platform hour is higher for peak periods due to inefficiencies introduced by split shifts, spread penalties, guarantee time, and other labor agreement provisions. Wage cost variations result from bonuses, overtime rates, penalty pay rates, and other bonus or penalty provisions. Temporal-variation models use a variety of techniques to incorporate these types of cost differences into the cost-estimation procedure.

Temporal-variation models are all enhanced cost-allocation models that focus on time period cost variations. Typically, nondriver costs are handled within the traditional cost-allocation framework while special methods are reserved for driver and vehicle cost calculations. As a result, the subsequent discussion focuses on the unique features of the temporal models, i.e., their examination of labor and vehicle costs, and only briefly describes those aspects similar to the cost-allocation method described previously.

The models identified as belonging to the temporal-variation generic type have been classified as representing one of three approaches:

1. Cost-adjustment approach, in which vehicle hour unit cost is adjusted relative to peak and base labor productivity;
2. Statistical approach, in which sample data are used to determine the relative productivity of peak-period service and cost; and
3. Resource approach, in which labor assignment practices are used to estimate labor requirements that reflect time-of-day variations.

Models of the temporal-variation type are certainly the most complex and perhaps the most important in understanding relations that affect bus operating costs. Because of the evaluation and nature of research of transit cost, temporal-variation models represent the latest efforts in this field. Numerous models can be categorized into the three approaches described above. Because of space limitations, only a single representative model is described below for each approach.

Peak-Base Model: Cost Adjustment

The peak-base model modifies the standard three-variable cost-allocation model by defining two different vehicle hour unit-cost coefficients, one for vehicle hours operated during the peak period and another for vehicle hours operated during the base period (5). The peak-period vehicle unit cost generally is higher than the base-period vehicle unit cost.

The two unit-cost coefficients are found by adjusting the standard allocation model's single vehicle hour coefficient. Two indices are used for the adjustment, one representing the relative productivity of labor and one representing the ratio of peak to base service. The indices are based on an audit of a sample month's data regarding vehicle hours and pay hours consumed during the peak and base periods. Vehicle mile unit cost is applied to both peak and base service. Peak vehicle unit cost is used for only the peak period.

The first step in the model is to assign the audit month's vehicle hours and pay hours to either the peak or base period. The labor productivity (i.e., ratio of pay hours to vehicle hours) is greater for the peak than base time period. Through various algebraic manipulations, it is shown that the new vehicle hour unit costs (peak and base) are calculated by the following formulas:

$$UC_p = [n(1 + s)/(1 + ns)] \cdot UC_s \qquad (3)$$

$$UC_b = [(1 + s)/(1 + ns)] \cdot UC_s \qquad (4)$$

where

UC_s = vehicle hour unit cost (traditional allocation model),
UC_p = peak-period vehicle hour unit cost,
UC_b = base-period vehicle hour unit cost,
n = relative labor productivity, and
s = service index (ratio of peak to base vehicle hours).

Equations 3 and 4 represent the adjustment factors for the vehicle hour unit costs. The resulting cost-allocation model for the Minneapolis-St. Paul example is presented below:

Traditional: $C = 9.90H + 0.31M + 1353V$

Peak: $C = 10.57H + 0.31M + 1353V$

Base: $C = 9.20H + 0.31M$

where C is cost, H is vehicle hours, M is vehicle miles, and V is peak vehicles.

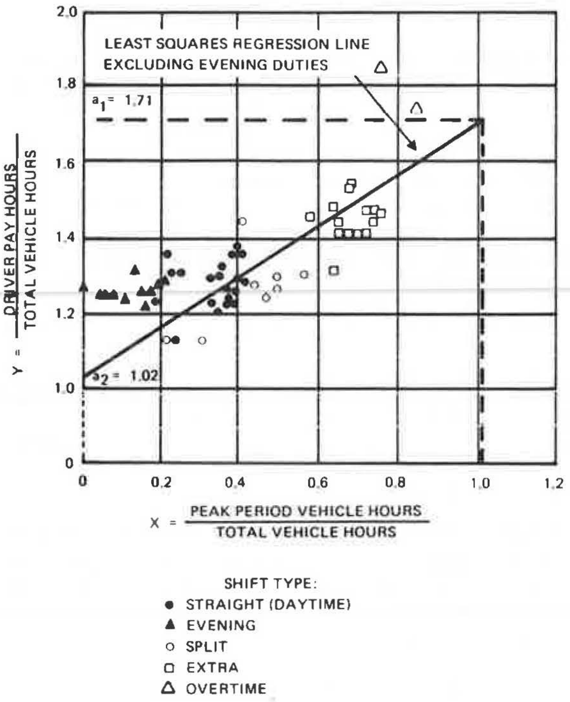
Arthur Andersen Model: Statistical

The Arthur Andersen model (6) is basically an enhanced fixed-variable cost-allocation model. Thus, the first step toward using the model is the development of the cost-allocation portion. Expense accounts are assigned to one of three cost types [i.e., direct costs, variable overheads (semifixed), and fixed costs] as well as three resources (i.e., vehicle hours, vehicle miles, and peak vehicles). Nine combinations are possible. Direct driver cost is included in the combination of vehicle hours and direct costs. Direct driver cost is analyzed in detail separately from the other combinations. Indirect driver cost and all other costs are estimated with the fixed-variable cost-allocation technique previously described.

To analyze driver costs, the initial step is to define the peak and base periods. Next, the sample shift data are used to estimate the coefficients of the following equation:

$$D_a = a_1(P) + a_2(B) \qquad (5)$$

Figure 2. Arthur Andersen model—regression analysis example.



where

D_a = total driver pay hours under the Andersen model,
a_1 = pay hours per peak-period vehicle hour,
a_2 = pay hours per base-period vehicle hour,
P = peak-period vehicle hours, and
B = base-period vehicle hours.

The coefficients a_1 and a_2 are found by plotting the sample data and fitting a curve. Each sample point is a shift that includes a combination of peak- and base-period vehicle hours (P and B). The proportion of peak and base hours depends on the shift's type, as shown in Figure 2 (6, Figure 5; 7). Generally, split shifts will have a higher proportion of peak-period vehicle hours than straight shifts. Extra shifts have a higher ratio of pay hour to vehicle hour than split shifts. Overtime shifts have the highest ratio. Regression analysis is performed to find the curve that relates the ratio of peak-period vehicle hours to total vehicle hours to the ratio of driver pay hours to total vehicle hours.

Estimates of the coefficients a_1 and a_2 can be found from the graph of the regression analysis results (Figure 2). Coefficient a_1 is the value on the vertical axis when the horizontal axis value is unity. The y-intercept of the graph gives the value of a_2. Once estimated, the parameters are converted to costs by multiplying them by the wage rate. As shown below, this calculation produces estimates of driver unit costs for peak and off-peak periods:

Item	Peak	Base
Pay hours per vehicle hour (a_1 and a_2)	1.71	1.02
Wage rate per pay hour (£)	2.00	2.00
Driver cost per vehicle hour (£)	3.42	2.04

Driver cost is combined with the results obtained from the fixed-variable cost-allocation model to produce total cost.

Bradford Model: Resource Allocation

The Bradford cost model was developed by R. Travers Morgan and Partners for their cost analysis of the Bradford (England) bus system (8). In addition to the development of cost procedures, this research presents a lengthy discussion of factors that influence operating costs and the quantification of these costs impacts. Because of space limitations, only the salient dimensions of this research effort are presented here. Of interest are the cost variations by day of the week, time period within a typical weekday, and a scheduling algorithm to cost new services.

The model is basically a fixed-variable cost-allocation model with pay hours, bus hours, and peak vehicles as the resources and driver labor costs and direct operating and overhead expenses as the cost categories. Expense accounts are assigned to resource and cost categories. The peak vehicle cost calculation follows the traditional cost-allocation approach. The calculation of the unit costs per pay hour and per vehicle hour involve slightly different procedures.

Unit costs per pay hour are obtained exclusively from expense accounts classified as driver labor. The initial step is to calculate the wage cost per 40-h week. Next, the driver schedule audit month data are used to find the ratio of pay hours to worked hours. Results for the audit month, which reflect various premium and penalty provisions of the labor contract, are presented below:

Day of Week	Ratio
Monday-Friday	1.08
Saturday	1.55
Sunday	1.75
Total	1.20

In addition, the research focuses on labor costs by time period. Much of this work examines the ratio of pay hours to work hours by time period.

Another feature of the Bradford study is the estimation of daily vehicle costs for weekdays, Saturdays, and Sundays. This analysis is based on the number of days each bus operates. As shown below, the vehicle cost by day varies considerably, since the use of buses differs substantially during each day:

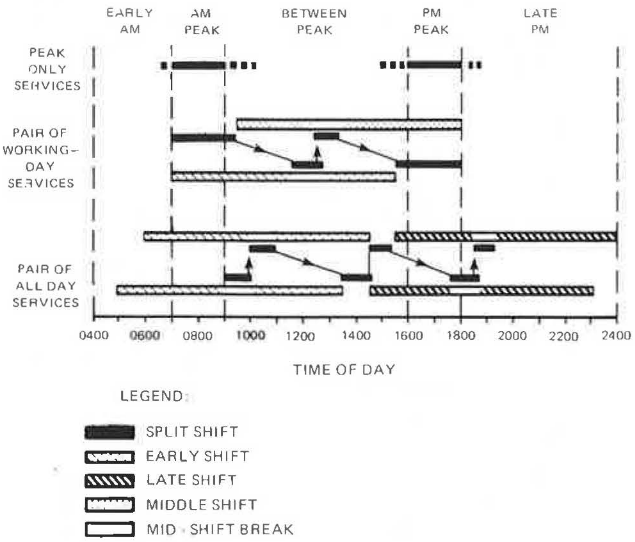
Day of Week	Vehicle Cost (£)
Monday-Friday	35.9
Saturday	32.5
Sunday	30.3

Another issue treated in the research is the vehicle cost variation by time period for weekday service. The examination is similar to the apportionment exercise carried out for the day-of-week variations. The analysis was performed for three layers of weekday service:

1. Peak only--Average duration is about 4 h, typically from 7:00 to 9:00 a.m. and 4:00 to 6:00 p.m.;
2. Working day--Average duration is about 11 h, typically from 7:00 a.m. to 6:00 p.m.; and
3. All day--Average duration is about 18 h, typically having staggered starting times from 4:00 to 7:00 a.m. and finishing times from 11:00 p.m. to midnight.

By using these definitions, the values from various intermediate steps (not presented here) were summed to obtain the appropriate vehicle costs for each service layer. A summary of this vehicle cost analysis is presented below:

Figure 3. Bradford model–driver scheduling model.



Layer	Hours	Avg Cost per Bus (£)
Peak only	4	22.2
Between peaks	7	16.0
Early morning, late evening	7	11.8

Total costs for each layer were found by adding the appropriate pay hour and vehicle hour cost components to the vehicle cost.

Another unique feature of the Bradford work was the development of a simple scheduling model to estimate the number of straight and split shifts required to implement service changes. The model is based on the labor scheduling practices prevailing at the transit property that typically reflect wage and work rule provisions of the labor agreement. As shown in Figure 3 (8, Figure 7.01), the model assumes that a single split shift staffs a peak-only service, that two straight shifts and one split shift staff a pair of working-day services, and four straight shifts and one split shift staff an all-day service.

CONCLUSIONS

The previous discussion provides a brief overview of the various cost-estimation procedures that have been developed and applied in the past. They vary considerably in terms of their level of sophistication, ease of use, and sensitivity to various dimensions of the bus system. The most commonly used cost procedure is the allocation model, which can be used in cost analysis of existing systems as well as in estimation of cost impacts of service changes. The more recent research modifies and enhances this basic analytical framework. Some researchers have segregated costs into various categories of fixed and variable components. Not surprisingly, the latest research places a common focus on examining the major cost element of transit service: drivers' wages. Although these methods differ considerably, they all recognize the labor-intensive nature of transit operations.

With greater emphasis on cost containment and resource allocation in the future, planners will need to understand the factors that influence bus operating costs. This will mandate a knowledge of the prevailing state of the art in bus cost-estimation procedures. It is anticipated that with greater emphasis on this topic, further enhancements and innovative approaches to estimating transit costs will evolve.

ACKNOWLEDGMENT

This research was performed under the sponsorship of UMTA, U.S. Department of Transportation. The contents of this paper were prepared by the Simpson and Curtin Division of Booz, Allen and Hamilton, Inc., and do not necessarily reflect the official views or policies of the U.S. Department of Transportation or UMTA.

A more complete discussion of the concepts presented in this paper is available elsewhere (9).

REFERENCES

1. W. Cherwony and B.E. McCollom. Development of Multi-Modal Cost Allocation Models. Proc., Fourth Annual Intersociety Conference on Transportation, ASME, New York, 1976.
2. M.G. Ferreri. Development of Transit Cost Allocation Formula. HRB, Highway Research Record 285, 1969, pp. 1-9.
3. H.W. Taylor. A Method of Bus Operations Costing Developed by NBC. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1975, pp. 6-13.
4. J.S. McClenahan and D.R. Kaye. A Method of Bus Costing Developed by Arthur Andersen and Company. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1975, pp. 31-51.
5. W. Cherwony and S.R. Mundle. Peak-Base Cost Allocation Models. TRB, Transportation Research Record 663, 1978, pp. 52-56.
6. J.W. McClenahan, D. Nicholls, M. Elms, and P.H. Bly. Two Methods of Estimating the Crew Costs of Bus Service. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplementary Rept. 364, 1978.
7. Symposium on the Costing of Bus Operations. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplementary Rept. 180UC, 1975, p. 51.
8. R. Travers Morgan and Partners. Bradford Bus Study. West Yorkshire Metropolitan County Council and West Yorkshire Passenger Transport Executive, England, Final Rept., 1976.
9. Simpson and Curtin; Booz, Allen and Hamilton. Bus Route Costing Proceedures: A Review. UMTA, Rept. UMTA-IT-09-9014-81-1, May 1981.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Potential Impacts of Transit Service Changes Based on Analytical Service Standards

GEORGE KOCUR

The results of a hypothetical case study of the Hartford, Connecticut, bus transit system in which service and fares are redesigned based on service standards derived from an analytical optimization model are presented. The key variables in the analysis are route spacing, headway, fare, and route length for both local and express routes. Three different sets of possible local objectives are treated, which place varying emphasis on profit (or deficit) and user benefits. Comparisons of the results with the existing system are made, and several policy issues are addressed. The analysis concludes that major increases in productivity are technically possible, based in large part on route restructuring and the introduction of substantial express service. Because relatively large changes from current operations are entailed, equity and political feasibility may be large issues in making the proposed changes.

The next decade promises to be a period of transition in urban transportation services. Urban public transportation was provided by private firms in most U.S. cities until the mid-1960s when most systems came under public management and subsidization. Few major changes in bus operating policies or system design have been made in this period of public ownership except for maintaining fares at a lower level than a private firm would have required. This strategy may be reassessed in many cities in the next decade for two major reasons. First, transit deficits have grown sharply over the levels originally anticipated when the systems became public operations. In 1965 the total U.S. transit deficit was $11 million and revenues covered 99 percent of expenses. However, in 1977 the U.S. transit deficit had risen to $2 billion and the percentage of operating expenses covered by farebox revenues had dropped to 53 percent (1). Part of this rapid increase in deficits had been absorbed by the federal government, but already tight state and municipal budgets will be forced to absorb most of the additional operating losses that may occur. This is likely to lead to consideration of service reductions, fare increases, and means of increasing productivity at the local level.

A second major impetus to the analysis of bus systems is energy policy. Expansions in bus service may reduce urban transportation energy requirements, but the deficits of such service require that any expansion in service must be designed very carefully to maintain economic feasibility.

SUMMARY OF SERVICE AND FARE STANDARDS METHODOLOGY

In this case study, the Hartford, Connecticut, system was redesigned according to service standards based on three sets of goals (or objective functions), and the results were compared with current operations. The case study treats peak-hour service only for simplicity. The service standards are based on optimization model results described in two previous papers (2,3). The model incorporates components that correspond to demand, supply, cost, revenue, and benefit models--the same set of components used by the traditional transportation planning process.

A linear approximation to a disaggregate logit model is used as the demand function. In the peak-period service standards used in this paper, only modal choice is considered; however, the results are valid for a general linear demand function. The measure of net user benefits (consumers' surplus) is based on the linear demand function. A simple cost model based on bus hours of service is used. Revenue, service level, and load-factor equations complete the basic model. These submodels are expressed as a series of equations for which optimal results are found through calculus and the technique of Lagrange multipliers [see Kocur and Hendrickson (2) for a description of the technical details].

Service and fare standards are developed for three objective functions:

1. Maximize profit or, in some circumstances, minimize deficit, subject to the constraint that a positive number of passengers must be carried;
2. Maximize a weighted sum of operator profit and net user benefit; and
3. Maximize net user benefit subject to a deficit constraint.

The second and third functions are possibly representative of current transit objective functions; the first is included primarily as a limiting case.

The service standards that emerge from these analyses are average values of the variables over the service area, expressed as equations into which local values of parameters such as demand coefficients, unit costs, operating speed, and population density are substituted. For example, the average headway for a profit-maximizing (deficit-minimizing) operator on local routes in an area of constant density is derived to be the following:

$$h = (4ca_4/3jk^2vpa_2A)^{1/3} \qquad (1)$$

where

h = headway (minutes);
c = bus operating cost (cents/minute);
a_4 = demand coefficient of fare;
j = average walk speed (miles per minute);
k = ratio of expected user wait time to headway;
v = local bus speed, including stops (miles per minute);
p = average trip density by all modes to the central business district (CBD) (trips per square mile per minute);
a_2 = demand coefficient of out-of-vehicle (wait-and-walk) time; and
A = constant in demand model.

These results vary by objective and are also affected by the operation of express service in the area and by trip or population density distribution.

The service standards for route spacing are functions of the same variables as headway, with very similar relations. In general, the express and local headway and spacing standards are proportional to the cube root of most of the parameters. Because the magnitude of some of the parameters is not known with precision, this robustness is reassuring. The fare standard for the same case given above is as follows:

$$f = -A/2a_4 - (4ca_2^2k/3jvpa_4^2A)^{1/3} \qquad (2)$$

The express-fare standard is similar. A warrant for operating express service is also included in the standards and depends strongly on the ratio of express speed to local speed. Generally, if express speed is 25-50 percent faster than local speed, express service is warranted. Formal vehicle-size standards have not yet been developed for the cases treated in this paper, although average passenger loads are displayed in the tables below to show the vehicle type required. Vehicle-size standards have been derived for simpler cases (2), and similar results are expected to hold in more complex ones.

These service standards are intended only to set the broad outlines of transit operations. Detailed route locations, schedules, and stop locations must still be decided by the operator based on experience, the constraints of available streets, variations in population density and markets, and so on. These standards also differ from the usual ones in that they specify average values rather than minimum standards. In this view, average standards are set to obtain maximum service effectiveness while minimum standards are set to address equity issues. Because equity issues are generally best dealt with outside the realm of technical analysis, they are not considered in this paper. The actual systems described in this paper follow the average standards closely.

DESCRIPTION OF CURRENT TRANSIT SYSTEM

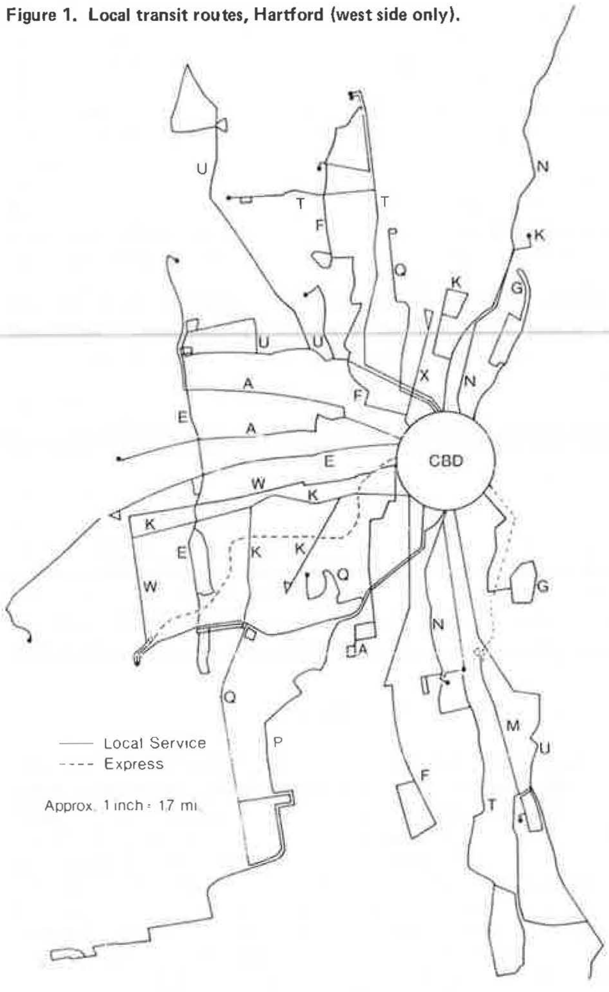
The Hartford urbanized area has a population of almost 500 000; it is the 53rd largest urbanized area in the United States. Public transportation is provided by a bus system operated by Connecticut Transit, which is funded by the Connecticut Department of Transportation (ConnDOT). In 1980 the system operated a fleet of 257 buses over 40 local and 14 express routes. The system carried approximately 10 500 passengers in the peak hour and about 1 500 000 passengers/month. Monthly operating costs were approximately $1 500 000 and monthly operating revenues were about $700 000, resulting in an operating ratio of just under 50 percent. The base fare was $0.50, with zone fare increments of $0.25 for a few long local routes and most express routes. (Fares were raised on March 1, 1980, from a previous level of $0.35.) Average operating cost per vehicle hour is about $30, according to monthly issues of One Month Comparison of Existing Systems published by ConnDOT.

Figure 1 shows the current local bus service in Hartford. Almost all express routes are operated as park-and-ride services from suburban parking lots. Because this paper does not address the design of park-and-ride services and because these services operate from isolated points beyond the general service area, they are not considered further.

The local routes serve roughly a 6-mile radius around the CBD. The Connecticut River divides the area into two sectors. On the more densely populated west side, which includes Hartford, more transit service is provided than on the east side. Thirty-one local routes operate on the west side with an average headway of about 15 min in the peak period. If these routes are assumed to serve an area of π radians (half a circle), the average spacing between them is $\pi/31$ or 0.101 radians. Nine routes operate on the east side at an average peak headway of about 21 min. Assuming that these routes serve an area of 0.8π radians (40 percent of a circle), the average spacing is 0.279 radians.

Peak-hour service employs approximately 140 buses for west-side local routes and about 40 vehicles for east-side routes. At an operating cost of $30/bus hour, this yields a cost of approximately $5400 for

Figure 1. Local transit routes, Hartford (west side only).



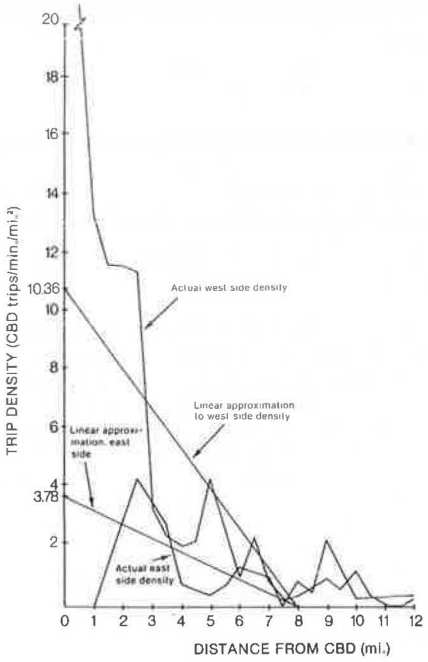
the peak hour. In the absence of detailed data, the operating cost per bus hour is assumed equal for peak and off-peak periods.

Ridership on west-side local routes in January 1980 was 5480 in the peak hour; on east-side local routes it was 1394. At a $0.50 fare these riders yield about $3437 in revenue. Thus, in the peak period revenues cover about 64 percent of operating cost; this is higher than the overall operating ratio of about 50 percent. However, this comparison between peak and off-peak operating ratios does not reflect the higher costs of peak operations. The average number of passengers per bus (peak direction and peak load point) in the peak hour is about 44, which yields a ratio of total riders to seats of 1.03. The majority of buses used in local service seat 45, with a small number holding 37. Table 1 summarizes the current local service for the west- and east-side sectors.

Table 1. Summary of current local bus service.

Avg Value	West Side	East Side	Total
Route length (miles)	6.0	7.0	--
Route spacing (radians)	0.101	0.279	--
Route headway (minutes)	15.6	21.6	--
Fare ($)	0.50	0.50	--
Peak revenue ($)	2740	697	3437
Peak ridership	5480	1394	6874
Peak costs ($)	4200	1200	5400
Profit ($)	-1460	-503	-1963
Passenger load per bus	44	45	44
No. of routes	31	9	40
Peak vehicles required	140	40	180

Figure 2. Density of trip to Hartford CBD.



SUMMARY OF TRAVEL-DEMAND DATA AND MODELS

As shown in the example service-standard equations, several key parameters are based on demand coefficients and trip density. This section briefly summarizes the Hartford data used to derive these parameters.

The density of travel to the CBD was derived from a 1 percent statewide home-interview survey conducted by ConnDOT (4). As all transit trips in the analysis underlying the service standards are modeled as being to or from the CBD, a half-circle 0.75 mile in radius is defined as the CBD. This is slightly larger than the actual CBD as defined by local planners and the census. Also, a peaking factor of 10 percent of all one-way trips is applied to convert the total person trip table from a 24-h period to a peak hour. These steps yield the graph of trip density against density from the CBD shown in Figure 2. Also shown are the linear approximations to the density functions used in the standards. These approximations are fitted by choosing a distance for the edge of the urban area and requiring that the total trips under the linear approximation equal the actual number of trips.

The demand coefficients for the linear modal-choice model used in the analysis are derived from a binary logit work modal-choice model calibrated for the Hartford area (5). There are separate models for three market segments (carless, one-car, and two-or-more-car households), but only the one-car model is used in the analysis. (The one-car and two-or-more-car models are virtually identical, while the carless model does not appear to have reasonable coefficients.) The one-car model is as follows:

$$t = \exp(U_t)/[\exp(U_t) + \exp(U_a)] \tag{3}$$

$$U_a = -0.061x_1 - 0.0244x_2 - 0.0137x_3 \tag{4}$$

$$U_t = -1.7636 - 0.0637x_4 - 0.061x_5 - 0.0244x_6 - 0.0137x_7 \tag{5}$$

Figure 3. Linear approximation to logit modal-choice model.

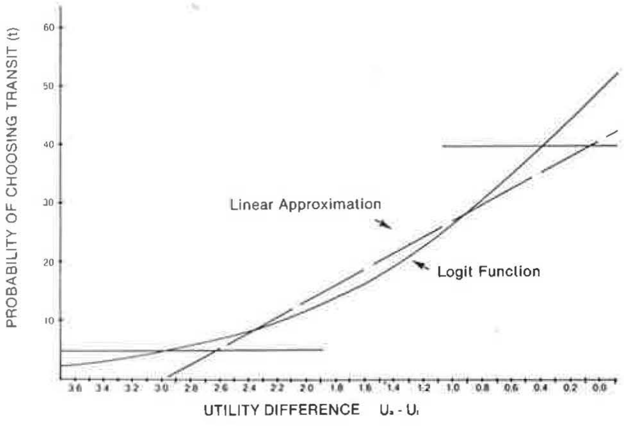


Table 2. Hartford validation results, peak hour.

Item	West Side Model	West Side Actual	East Side Model	East Side Actual
Transit ridership	5039	5480	1242	1394
Transit revenue ($)	2520	2740	621	697
Transit cost ($)	4298	4200	1125	1200
Transit net user benefit ($)	5171	--	1074	--
Avg transit passenger load per bus	42	44	50	45

Table 3. Profit maximization with express and local service.

Item	West Side	East Side	Total
Route length (miles)			
Local	4.0	3.6	--
Express	7.0	6.2	--
Route spacing (radians)			
Local	0.445	0.682	--
Express	0.328	0.517	--
Route headway (minutes)			
Local	14.0	19.3	--
Express	17.3	24.3	--
Fare ($)			
Local	1.20	1.08	--
Express	1.27	1.09	--
Peak revenue ($)			
Local	1898	359	2257
Express	2143	422	2565
Peak ridership			
Local	1576	332	1908
Express	1690	387	2077
Peak costs ($)			
Local	731	246	977
Express	992	319	1311
Net user benefit ($)			
Local	842	154	996
Express	1089	196	1285
Profit ($)			
Local	1167	113	1280
Express	1511	103	1254
Passenger load per bus			
Local	52	29	--
Express	51	32	--

where

t = transit modal share;
U_a = automobile utility;
U_t = transit utility;
x_1 = automobile out-of-vehicle time (8 min);
x_2 = automobile in-vehicle time (computed at 24 mph plus 8 min access time);
x_3 = automobile operating cost (computed at 16¢/mile plus 67.5¢ parking);
x_4 = transit wait time (minutes), computed at 0.4 times headway;
x_5 = transit walk time (minutes), computed at 3 mph from route spacing;
x_6 = transit in-vehicle time (minutes); and
x_7 = transit fare (cents).

The logit model is approximated by a linear function as shown in Figure 3. The logit function is S shaped; only its right-hand half with transit modal shares less than 50 percent is shown. The linear approximation was chosen by inspection to fit a range of modal splits from 5 to 40 percent fairly closely, as this is felt to be the range of interest.

These demand coefficients correspond to the following elasticities of transit use when the transit modal share is 20 percent and all variables are at their mean values:

1. Wait and walk time: -0.60,
2. In-vehicle travel time: -0.25, and
3. Fare: -0.35.

This linear modal-choice model together with the linear trip-density function were validated on the current Hartford data to ensure that the approximation errors were tolerable. The results are given in Table 2. Predicted ridership is 6-8 percent too low in both the east and west sectors, but predicted costs are quite close to actual values. Transit net user benefit (or consumers' surplus) cannot be measured, so no comparison with actual benefits is possible; this figure is reported as a base of comparison for later policy options.

The linear modal-choice model and the linear trip-density function form the basis of the simple sketch-planning model used to estimate the impacts of implementing the service and fare standards described below; it is also described in Kocur (6).

SYSTEM DESIGNS FOR ALTERNATIVE OBJECTIVE FUNCTIONS

A series of analyses was conducted to design transit services for the Hartford area within 8 miles of the CBD. The analyses consider service standards based on all three objective functions, each with local and express service.

Profit Maximization with Express and Local Service

In this analysis, a set of service standards is computed and applied based on the assumption that profit maximization (excluding capital costs) is the system objective. Service standards are computed for eight variables: local headway, route spacing, fare, and route length, and express headway, route spacing, fare, and route length. These values are given in Table 3 with separate values for the west and east sides, as the service patterns are quite

Figure 4. General structure of transit service based on analytical service standards.

CBD - Central Business District

different. The express speed of 20 mph possible in Hartford due to its extensive expressway network is double the average local speed of 10 mph in the peak, so express service is warranted according to one of the equations in the service standards.

The transit service area, under these standards, is divided into two concentric rings that encompass roughly equal numbers of trips to the CBD, as shown in Figure 4. The inner ring is served by local routes. The outer ring is served by routes that, on inbound trips, begin at the outer edge of the area, make local stops until reaching the boundary of the local service area, and then run express (nonstop) to the CBD. Transfer points are established at the local and express service-area boundary for intracorridor trips.

On the west side, the area within 4 miles of the CBD is served by local routes, and the area 4-7 miles from the CBD is served by express routes. The express routes make local stops in the outer area but then run express (without stops) on an expressway facility for the last 4 miles to the CBD. The 7 local routes are spaced 0.445 radians apart, which results in an average walk time of 5.6 min or an average walk distance of 0.28 mile. The local routes operate at 14-min headways and charge a one-way fare of $1.20.

The west-side express routes are spaced 0.328 radians apart, which results in an average walk time of 6.9 min or an average walk distance of 0.35 mile. The express routes operate at 17.3-min headways and charge a one-way fare of $1.27, only $0.07 more than the local routes. Express routes have higher speeds and are thus more attractive relative to local service in this regard, but they also have wider route spacings and longer headways. Because the standard fare is strongly dependent on the service quality, these offsetting effects decrease the fare differential from local service.

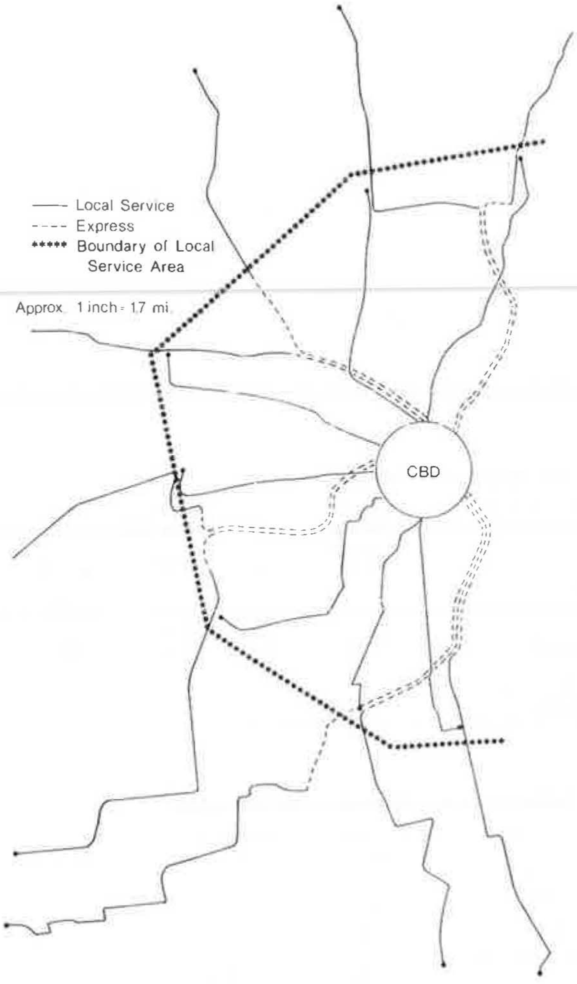
East-side results follow a similar pattern, although with a smaller service area and lower service levels due to the lower trip density. The express fare is about the same as the local fare because the express headways and route spacing are so much poorer than the local ones.

Figure 5 shows a route pattern based on these results. It is quite different than the current service. Routes are much more widely spaced in the inner area, which provides a lower level of service than the current system. In the express zone, however, there is more service than currently provided. These results raise equity issues in the distribution of transit service and its benefits and costs.

These services are then assessed by using a sketch-planning model, which predicts the impacts given in Table 3. The peak-hour ridership is 3985 as opposed to the current total of 6874, almost a 50 percent reduction. Revenues are $4822 as opposed to the current $3437, a 40 percent increase. Operating costs are $2288, a reduction of over 50 percent. Net user benefits are estimated to be $2281 as opposed to the current $6245, a reduction of over 60 percent. Average passenger loads are near 50 on the west side and 30 on the east side.

These results are, of course, quite extreme. This is due principally to the use of an objective function that is unlikely to represent current transit goals in most cities, and possibly also to the extension of the demand coefficients beyond the range of data on which they were originally calibrated. The traveler reaction to the long walk distances that emerge from the model may be quite different than their reactions to the shorter distances they now walk. However, recalling that these service standards are functions of the cube root of the demand parameter for walk time, the walk coefficient would have to be in error by a factor of 8 to halve the route spacing, which still leaves an average walk distance of 0.14 mile or a walk time of 2.8 min. In general, the issue of sensitivity to walk distance is one on which demand models and operators have different beliefs.

Figure 5. Route structure for profit maximization (west side only).



Profit Maximization with All Local Service

Table 4 gives the results of profit maximization with all local service to contrast it with the express- and local-service patterns. The major differences are that the service area is slightly smaller, headways and route spacing are a compromise between the separate express and local optima, fares are slightly lower, and passenger loads per bus are much higher. The load of 85 found for the west side exceeds the capacity of standard buses and suggests the use of articulated buses. The all-local-system case is inferior to the express system in all impacts, even if the heroic assumption is maintained that unit operating costs and average speeds are equal for standard and articulated buses. Operating profit is $2379 instead of $2570 and user benefits are $2149 instead of $2298.

In fact, the magnitude of the passenger loads in this all-local case (and the all-local case for other objectives) may suggest the consideration of articulated buses in areas of similar or higher density than Hartford. If express service can be

Table 4. Profit maximization with all local service.

Item	West Side	East Side	Total
Local route length (miles)	6.8	5.8	--
Local route spacing (radians)	0.328	0.491	--
Local route headway (minutes)	15.5	20.8	--
Local fare ($)	1.17	1.05	--
Local peak revenue ($)	3678	713	4391
Local peak ridership	3146	682	3828
Local peak costs ($)	1504	508	2012
Local net user benefit ($)	1810	339	2149
Local profit ($)	2174	205	2379
Local passenger load per bus	85	46	--

Table 5. Maximization of profit plus half of net user benefit with express and local service.

Item	West Side	East Side	Total
Route length (miles)			
Local	4.1	3.8	--
Express	7.2	6.5	--
Route spacing (radians)			
Local	0.396	0.596	--
Express	0.290	0.447	--
Route headway (minutes)			
Local	12.8	17.6	--
Express	15.7	22.1	--
Fare ($)			
Local	0.82	0.75	--
Express	0.87	0.77	--
Peak revenue ($)			
Local	1905	390	2295
Express	2019	441	2460
Peak ridership			
Local	2316	522	2838
Express	2314	576	2890
Peak costs ($)			
Local	922	324	1246
Express	1265	427	1692
Net user benefit ($)			
Local	1751	351	2102
Express	2068	429	2497
Profit ($)			
Local	983	66	1049
Express	754	14	768
Objective value ($)	3646	470	4116
Passenger load per bus			
Local	62	36	--
Express	56	38	--

operated, it appears that this option dominates the all-local articulated-bus option on the technical grounds being considered.

Maximization of Weighted Profit plus Net User Benefit with Express and Local Service

Table 5 gives the results of adopting service standards to achieve the objective of maximizing the sum of operator profit and net user benefit, with user benefit being weighed only half as much as profit. Compared with the profit-maximization objective, the service area is larger, the route spacings and headways are improved, and fares are lower. Compared with the current system, there is less local service but more express service.

Ridership is 5728 in the peak hour, slightly lower than the current level of 6874. Revenues are $4755 as opposed to the current $3437, and costs are $2938 instead of approximately $5400. Thus, an operating profit of $1817 is still obtained. Net user benefits are $4599 as opposed to the current $6245.

Many of the same patterns appear in these results as in the previous case. Route spacings are quite wide, which results in an average walk of 5.1 min (0.26 mile) for west-side local users and 6.3 min (0.32 mile) for west-side express users. Headways are fairly long and fares are somewhat higher than the current $0.50 fare. Passenger loads per bus are relatively high but not beyond the capacity of standard buses with standees.

Table 6. Maximization of net user benefit, subject to break-even financial constraint, with express and local service.

Item	West Side	East Side	Total
Route length (miles)			
Local	4.2	3.8	--
Express	7.3	6.6	--
Route spacing (radians)			
Local	0.362	0.580	--
Express	0.264	0.434	--
Route headway (minutes)			
Local	11.8	17.3	--
Express	14.5	21.6	--
Fare ($)			
Local	0.39	0.66	--
Express	0.42	0.68	--
Peak revenue ($)			
Local	1227	376	1603
Express	1243	422	1665
Peak ridership			
Local	3121	570	3691
Express	2964	621	3585
Peak cost ($)			
Local	1108	342	1450
Express	1532	453	1985
Net user benefit ($)			
Local	3103	413	3516
Express	3432	499	3931
Profit ($)			
Local	119	34	153
Express	-289	-31	-320
Passenger load per bus			
Local	71	38	--
Express	60	39	--
Shadow price (Y_2)	1.23	1.70	--

Thus, a substantially different operating policy results from this analysis than the status quo: 60 percent of all service operated is express, route spacings are about triple the current ones, fares are about 50 percent higher, and an operating profit is made.

Maximization of Net User Benefit, Subject to Break-Even Financial Constraint, with Express and Local Service

Table 6 gives the results of applying service standards to achieve the objective of maximizing net user benefits subject to a break-even financial constraint in the peak period. Again, express and local services are warranted. The west-side service is very similar to the previous case except that a lower, $0.40 fare is charged. The service area is the largest of any of the cases examined, and the headways, fares, and route spacings are the lowest. Even so, the average walk time is 4.8 min (0.24 mile) for west local routes and 5.9 min (0.30 mile) for west express routes. One other impact measure can be computed in this case--the shadow price related to the break-even financial constraint. The approximate shadow price is $1.23, which is interpreted that an extra dollar of subsidy would produce $1.23 in extra net user benefit. This measure can help decisionmakers in assessing the level of deficit they are willing to support.

The total ridership in this case is 7276, slightly higher than the current 6874. Revenues are $3268 as opposed to the current $3437. Costs are approximately $3435 as opposed to the current $5400. (The deficit is $167, which is an approximation error.) Net user benefits increase to $7447 from an estimated $6245 in the current system.

This solution demonstrates a system in which fares are approximately equal to current fares and which operates at a break-even level instead of only covering two-thirds of its peak-hour operating costs as does the current system. Net user benefits and ridership increase over current levels, and passenger loads are quite manageable, although large vehicles are required for west-side local service.

This result emerges because routes are widely spaced in the inner area where little ridership is lost through these changes, while service is improved over current levels in the express zone where considerable ridership can be gained. The express service lowers running times significantly that, in turn, reduces costs, increases ridership, and allows a higher fare to be charged, which generates more revenue. Even in the express zone, however, routes are widely spaced.

CONCLUSIONS

Several general conclusions can be drawn from this assessment of service and fare standards, and other conclusions can be drawn with reference to specific objectives. These conclusions can be viewed as general directions of change for which there is significant analytical support, but detailed recommendations must be based on further analysis and consideration of institutional factors.

It appears that service standards can be used to improve the productivity and performance of transit systems substantially. Especially in the area of route structure, many current systems do not appear to have been assessed or designed systematically and general restructuring could provide substantial benefits. A possible substantial expansion of express service and a coordinated fare policy to reinforce the service objectives are also key elements of the standards.

Bus transit systems may also be able to substantially improve their financial performance, at least in peak periods. This result comes from possibly raising fares above current levels, increasing average route spacings (or distance between routes) substantially over current levels, leaving headways near typical current levels, and operating over half of total transit service as express service. Under some sets of goals, little or no user benefit is sacrificed to attain these financial results.

Average passenger loads per bus at the peak load point in the peak direction under these operating strategies are typically near the capacity of standard bus vehicles. In areas in which express service is not feasible due to lack of expressways or other suitable roadways, all routes must be operated in local service; these generally produce passenger loads higher than the capacity of standard buses. In these cases, either headways and route spacing must be decreased to meet capacity requirements or larger vehicles must be used. It is possible that low-cost light-rail transit systems might be effective alternatives in high-density areas in which express service cannot be offered, but that analysis is not carried out in this paper.

Major issues of equity and the distribution of benefits are raised by the analysis. For most cities, the analysis implies service cuts for the inner area in which many carless and low-income people live, and service expansions or at least smaller service cuts in the relatively more affluent and mobile outer sections of the city or suburbs. However, it must be noted that these results are based on demand-model parameters that do not treat market segments such as low-income groups separately but assume that all travelers in the city are identical. These equity issues must be addressed seriously if any service charges are considered by a local area. This is a function for minimum service standards.

Turning to objective-specific findings, the profit-maximizing standard fare is over $1.00 (in 1980 dollars) in most cases. Typical headways of 15-20 min and average distances of about 1 mile between routes result, which are larger than current practice in Hartford and most other cities. Revenue-to-operating cost ratios near 1.5 in the peak period appear possible.

The objective of maximizing the weighted sum of operator profit and net user benefit yields a range of service standards, depending on the weight placed on net user benefit. In most areas, standard bus vehicles will be filled to capacity under this objective. About 60 percent more service is operated in this case than in the profit-maximizing case. As the weight placed on benefit decreases, the solutions approach the profit-maximizing case. Even with net user benefit weighted at half the value of operator profit, the transit system may still make a small operating profit in the peak period at fares slightly higher than current levels.

The third objective of maximizing net transit user benefits subject to a deficit constraint was examined by using a break-even criterion. Results are similar to the previous case with benefit weighted less than profit.

To summarize, the analysis suggests that major restructuring of transit routes, fares, and headways is possible, and that service standards can be a vehicle for making these changes. A heavier reliance is placed on users' walking to bus stops, substantial express service is initiated, and some level of operating profitability (at least in peak periods) appears possible. These major changes may result in disbenefits to inner-area residents and increased benefits to outer-area residents. Systematic analysis of transit service nevertheless appears to have significant potential for increasing transit productivity, benefits, and financial performance.

ACKNOWLEDGMENT

The views expressed in this paper are solely mine and do not necessarily reflect the position of Connecticut Transit, ConnDOT, or any other official body. The support of the University Research Program, U.S. Department of Transportation, is gratefully acknowledged.

REFERENCES

1. Transit Fact Book, 1978 ed. American Public Transit Association, Washington, DC, 1979.
2. G. Kocur and C. Hendrickson. Design of Local Bus Service with Demand Equilibration. Transportation Science, Vol. 16, No. 2, May 1982.
3. G. Kocur. Optimal Design of Express and Local Bus Services. Presented at International Symposium on Transportation Equilibrium and Supply Models, Montreal, Quebec, Canada, Nov. 1981.
4. Analysis of Survey Responses to 1976-77 Connecticut Statewide Household Travel Survey. Bureau of Planning and Research, Connecticut Department of Transportation, Wethersfield, draft, 1978.
5. Wilbur Smith and Associates. Hartford Capitol Region TRANS PLAN Project: Transportation Demand and Systems Usage Models. Capitol Region Council of Governments, Hartford, CT, Sept. 1979.

6. G. Kocur. Optimal Design of Urban Bus Systems with Demand Sensitive to Service Levels. School of Urban and Public Affairs, Carnegie-Mellon Univ., Pittsburgh, Ph.D dissertation, 1981.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Abridgment

Assessing User Needs in Design of a Management Information System for Rural Public Transportation Services

JOHN COLLURA AND DALE FERGUSON COPE

The feasibility of designing a comprehensive management information system (MIS), which will assist in the performance of management and operations tasks involved in the provision of rural public transportation, depends on the definition and understanding of the relations among information gathering, processing, and reporting activities and those areas of administration and operations that might avail themselves of the advantages of a comprehensive MIS. The initial analysis of user needs and constraints, which is the focus of this paper, should indicate in specific terms the ways in which an MIS will contribute to the performance of information-related activities within the transportation operation. A step-by-step process is outlined to provide guidance to the initial MIS development activities and to assist in the structuring of a low-cost, comprehensive, and efficient MIS that will meet the information gathering, processing, and reporting requirements of rural public transportation authorities and operators.

The development of information gathering and processing techniques and the construction and maintenance of data sets for the purposes of administering and operating public transportation services are becoming ever more important objectives for government agencies, transportation authorities, operators, and the riding public. Current fiscal constraints, energy shortages, and the resulting emphasis on service efficiency and productivity are leading to increased dependency on the availability of appropriate and accessible information for transportation policymaking and management.

The continued viability of public transportation, particularly in rural areas where the service population is widely dispersed, where costs per trip tend to be high, and where systems are currently facing reductions in operating subsidies (Section 18 funds, Urban Mass Transportation Act of 1964, as amended) as well as other cutbacks in local, state, and human service agency budgets, rests especially on the ability of those who are responsible for the delivery of services to monitor carefully both the technical and social efficiency of these systems. The design and implementation of a comprehensive management information system (MIS), the advantages of which have long been recognized in the private sector, which will meet the data requirements of rural transportation agencies with full recognition of the limited financial and personnel resources available to such agencies, is a key factor in the maintenance of efficient and coordinated programs.

For the purposes of this paper, we have grouped the major management information needs of regional and local public transportation agencies and operators in rural areas into four general areas of activity:

1. Billing and accounting needs (1-3),
2. Monitoring and evaluating needs (4, and Section 15 of the Urban Mass Transportation Act of 1964, as amended),
3. Reporting requirements (5), and
4. Routing, scheduling, and dispatching needs (6).

Particularly in smaller transit facilities, these four broad categories encompass the minimum information-related activities that must occur to meet funding and report-filing regulations and ensure effective delivery of service.

MIS OBJECTIVES

Once the decision has been made to investigate the possibilities of designing a comprehensive MIS, and the proposal has been justified in the light of other uses for the start-up funds and time elsewhere in the operation, the process of identifying system objectives, constraints, and desirable features may begin (7). MIS objectives should be defined as specific targets that indicate how the MIS will support various aspects of the transit service and should be expressed in terms of what managers and operators will be able to do once their information requirements have been satisfied.

Thus, the basic sequential flow of the MIS development process is initiated with an analysis of existing data needs and current system capabilities and deficiencies. Participants in this first definitive step might categorize information system deficiencies as either those gaps that result from what information is lacking in the current system or as deficiencies in the structure, organization, storage, or use of information.

A review of the work tasks and schedules of each employee, including managers, bookkeepers, dispatchers, drivers, and others, and the information requirements that correlate to their tasks will reveal both the data needs and the deficiencies in the data and/or data structure that may be present in the existing system. As a result of this effort, the preliminary outlines of the MIS that might be designed to maintain and manipulate the necessary information and the specific technology and personnel required to process the information will become evident. The delineation of appropriate questions for the transit manager to ask with regard to specific goals in each category will aid in the clarification of the point along the simple-complex spectrum at which the service's MIS should be directed.

Another crucial aspect of assessing MIS needs involves the selection of an approach to the in-place information activity. An empirical approach involves the study of the in-place system slated to be replaced or improved; the idea is that the existing information system may have some good features and that the inherent deficiencies have been calibrated over time such that personnel have learned to compensate for the inadequacies of the system. In addition, the empirical approach also stresses that there may be a chance that inexpensive modifications to current methods might solve any information problems that exist (6).

Even if the decision has been made to scrap the existing information activity, an empirical examination of the information gathering and processing methods central to the present system will lead to an understanding of the volume and kinds of data needed in the new MIS, the level of accuracy and precision of current inputs, and the ways in which information currently generated is used by various personnel within the organization.

The alternative approach might be termed the logical method and may appear to be the mode of choice in a situation where the present information activities are hopelessly bound up in methods that result in wastes of time, money, and energy. The logical approach, essentially starting from scratch, may be preferable, for example, when a new manager is hired or a new service is being implemented. Under the logical approach, the information system designed will be free to experiment with various concepts of information management and, in this case, ignorance of how information-related activities have been carried out in the past may be a blessing. The end result of either approach, however, should be a general idea of the major elements of the new or redefined MIS and an understanding of how these elements are interrelated.

After the initial tasks of defining, refining, and reconciling information requirements have been completed, the user's needs assessment should result in the following outputs:

1. Statements of MIS objectives,
2. List of cost and personnel constraints and other restrictions,
3. Statements (or lists) of information requirements, and
4. Statements of what personnel should be able to do when the MIS is implemented.

The process of defining needs will have revealed data categories of essentially two dimensions: (a) a set of information-related activities that the MIS will encompass, and (b) a listing of the sources of data needed to accomplish those activities. At this point, it will be helpful for the information manager to construct a matrix of these information activities and the sources of data to aid in illustrating both the multiple uses of data gathered for the MIS and the processing steps necessary to match the activity with the source of information. Such a matrix will also illustrate the empty cells in which new activities and/or new sources of data are needed.

Each individual in the organization will contribute to the construction of the activity-source matrix a list of information-dependent tasks for which he or she presently takes responsibility for and a series of current sources of the requisite data for completion of those tasks. The conduct of this exercise, and staff discussion of the resulting matrix, will not only provide information regarding data storage locations and the full range of information-dependent activities carried out within the operation, but it will also tend to reveal certain duplicative efforts (if they exist) and any access problems encountered by particular members of the staff (7). The exercise might have several other beneficial results in pointing out areas of excess or missing information and thus lead to the reorganization of particular data on file. This streamlining of files will be especially important if the decision has been made to computerize the MIS, and it will demonstrate the true storage needs of such a system.

Figure 1 represents a slice of a matrix constructed of the information activities and data sources for a hypothetical transit service. It is assumed that the person within the organization who has taken on the role of MIS coordinator has worked up a list of information requirements and has ascertained the location of various files, card boxes, etc., where the necessary data are stored for manual or computerized processing. For the purposes of this hypothetical example, let us say that Figure 1 represents a slice of the matrix wherein an activity is specified and the data storage locations listed. The activity includes the billing of area human service agencies for transportation services provided to eligible clients under a specific funding source.

USING ACTIVITY-SOURCE MATRIX TO MODEL MIS

By examining a specific activity within the matrix and its corresponding data sources, a process for conducting that activity begins to take shape. The matrix assists in the task of converting verbal descriptions of information needs (collected from the involved personnel) into pictorial representations of work tasks and work flow. A flowchart of this activity sequence, as displayed in Figure 2, will serve as an abstract representation of the MIS components and their interrelations (6).

The flowchart illustrates inputs and their sources, a process (manual or automated), and outputs. The information and understanding collected within the framework of the activity matrix will provide assistance in the construction of a flowchart specific to the MIS activity. As mentioned above, a flowchart will prove useful for either a manual or computerized system; the differences will appear only in the processing block. At a minimum, the blocks shown in Figure 2 should be included in the graphic representation of the activity.

Figure 3 presents a flowchart that might be constructed for the carrying out of the first hypothetical activity in the matrix: Billing participating human service agencies for transportation services delivered to clients.

The input step specifies those files to be used in the activity, the location of these files, and the specific data to be retrieved from the files; coding information; and timing of the activity (monthly, annually, etc.). The processing step specifies how the data are to be sorted, what calculations are necessary, what files are to be updated with the new data, and what and where processed information is to be stored. The output step specifies what should be done with the data. In this case, an invoice for each participating agency will be prepared, mailed, and stored.

CONCLUSIONS

As demonstrated in this paper, the definition of sources and processes required to carry out the information activities of small transit services is a critical first step toward increasing the efficiency and productivity of both the transit staff

Figure 1. Matrix of information activities and sources of data.

SOURCE / ACTIVITY	CLIENT FILE(S)	VEHICLE FILE(S)	TRIP FILE(S)	AGENCY FILE(S)	REVENUE FILE(S)	PERSONNEL FILE(S)
	• ID Number • Name, Address • Handicap • Age • Eligibility • Agency Affiliation • Etc.	• Capacity • Insurance • History • Maintenance • Etc.	• OD Data • Vehicle Hours • Vehicle Miles • Purpose • Eligibility	• Rate Data • Subsidy Data • Funding Sources • Billing Proc.	• Fares • Subsidies • History • Etc.	• ID Number • Wage Data • Benefit Data • Time Data • Employment History • Etc.
BILLING HUMAN SERVICE AGENCY FOR SERVICES TO CLIENT IN MONTH	X		X	X	X	
FORM 406 UMTA SECTION 15 REPORT	X	X	X		X	X
OTHER ACTIVITIES • • • •						

Figure 2. Flowchart of theoretical input, processing, and output.

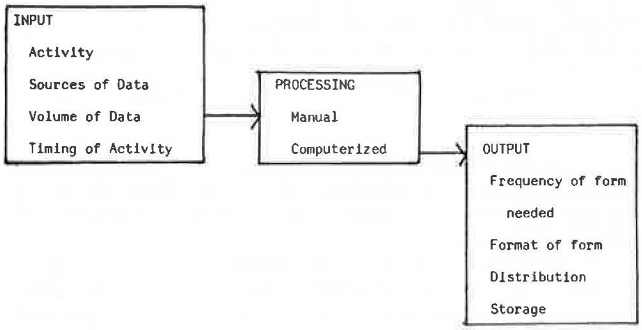
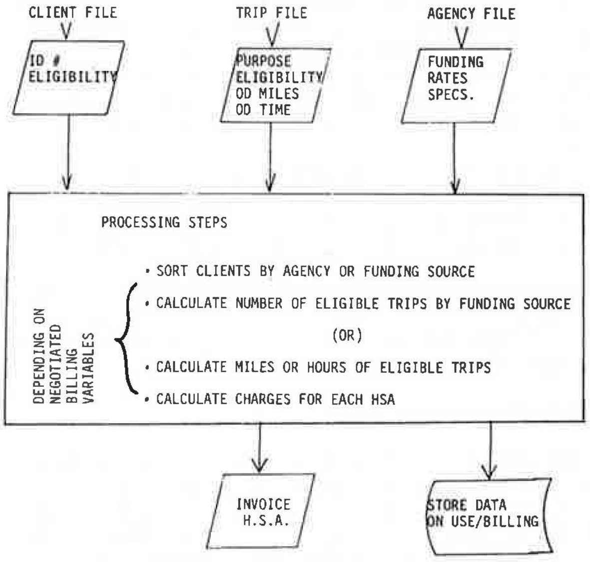


Figure 3. Sample flowchart for billing human service agencies.



and the service itself. Preparing a flowchart similar to that shown in Figure 3 for each information activity may seem an onerous task at first; however, the understanding gained through the process of defining needs, specifying activities and data sources in matrix form, and constructing a flowchart for each activity will be valuable to all participants in the service. The clarification of all the elements involved in the conduct of each information-related task will lead to more efficient procedures in the four major areas of concern.

ACKNOWLEDGMENT

This paper is part of a research project sponsored by the Office of University Research, U.S. Department of Transportation. The assistance of Theodore R. Higgs, Jr., contracting officer, and B. Paul Bushueff, Jr., contracting officer's technical representative, is gratefully acknowledged.

The overall goal of this ongoing research is to develop and test a comprehensive MIS that will be flexible enough to meet the varying data needs of rural public transportation operations. The MIS will be developed with full recognition of the limited resources available to agencies in rural areas.

The opinions and conclusions expressed or implied in this paper are ours and not necessarily those of the U.S. Department of Transportation.

REFERENCES

1. J. Collura, J.N. Nkonge, D.F. Cope, and A. Mobolurin. Charging Human Service Agencies for Public Transportation Services in Rural Areas. TRB, Transportation Research Record 830, 1981, pp. 15-21.
2. D.M. Alshuler and M. Flusberg. Establishing Contractual Relationships for Demand-Responsive Transportation Services. TRB, Transportation Research Record 608, 1976, pp. 107-112.
3. J. Collura and D. Cope. A Manual of Procedures to Allocate Costs of Rural Public Transportation Among Participating Towns and Human Service Agencies. Office of University Research, U.S. Department of Transportation, Draft Rept., July 1981.

4. J. Collura and R.P. Warren. Regional Paratransit Services: An Evaluation. Transportation Engineering Journal, ASCE, Vol. 105, No. TE6, Nov. 1979, pp. 683-697.
5. L.J. Harman and R.V. Giangrande. Microcomputer Applications in the Management of Paratransit Operations. Paper presented at Second International Conference on Transport for the Elderly and the Handicapped, Cambridge Univ., Cambridge, England, July 1981.
6. B.P. Bushueff, Jr. An Analysis of the Automation Requirements for Small Coordinated Paratransit Systems. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Sept. 1980.
7. G.B. Davis. Management Information Systems: Conceptual Foundations, Structure, and Development. McGraw-Hill, New York, 1974.

Publication of this paper sponsored by Committee on Rural Public Transportation.

Madison Avenue Dual-Width Bus Lane Project

SAMUEL I. SCHWARTZ, ANDREW HOLLANDER, CHARLES LOUIE, AND RAYMOND AMORUSO

On May 26, 1981, New York City implemented an exclusive dual-width bus lane on Madison Avenue in midtown Manhattan, which was funded by a one-year federal demonstration grant. The facility operates from 2:00 to 7:00 p.m. on weekdays and carries 25 000 passengers daily. It shares a roadway with three lanes of mixed traffic and is defined by pavement markings and overhead signs, accompanied by intense enforcement. Initial results indicated that (a) peak-hour bus speed was increased by 83 percent, (b) peak-hour bus reliability was increased by 57 percent, (c) peak-hour bus density was reduced by 45 percent, (d) traffic speed on Madison Avenue was increased by 10 percent, (e) average speed on parallel avenues was unchanged, and (f) average speed on eastbound cross streets was unchanged and on westbound cross streets was reduced by 6 percent. This project represents one of the most ambitious transit-priority projects for an urban arterial short of a complete ban of other traffic. The evolution and results of the project are described, and the implementation process is emphasized.

The concept of exclusive bus lanes is well established. It has been tested on expressways and urban streets throughout the United States and is now an accepted method of moving more people faster. But the institution of a dual-width bus lane on the congested streets of midtown Manhattan must be one of the severest tests of this approach.

This paper presents the rationale for selecting Madison Avenue as the locale for such a project and describes the implementation of the project and its impacts.

DUAL-WIDTH BUS LANE PROJECT

Project Background

Planning for a major surface transit improvement in midtown Manhattan began in 1979. All major avenues in midtown Manhattan were examined as possible candidates. Madison Avenue was selected because it was characterized by the following:

1. The highest bus volumes on any midtown arterial--approximately 200 buses during the peak hour (approximately 24 000 people travel by bus between 2:00 and 7:00 p.m. on Madison Avenue),
2. The lowest bus travel speeds on any midtown avenue during midday and evening periods--approximately 4 mph, and
3. The lowest automobile travel speeds on any midtown avenue during the evening period--approximately 5 mph.

These characteristics of Madison Avenue stem from its location as the central corridor for office development in midtown. Five local bus routes (with a combined headway of 53 s during the peak hour) and 32 express bus routes traverse its length. [Express buses run nonstop between the Manhattan central business district (CBD) and residential areas in each of the city's boroughs.] Subway lines flank it two blocks away on both sides. A major commuter railroad terminal (Grand Central Station) is one block away on Park Avenue at 42nd Street (see Figure 1).

The site conditions of Madison Avenue are as follows:

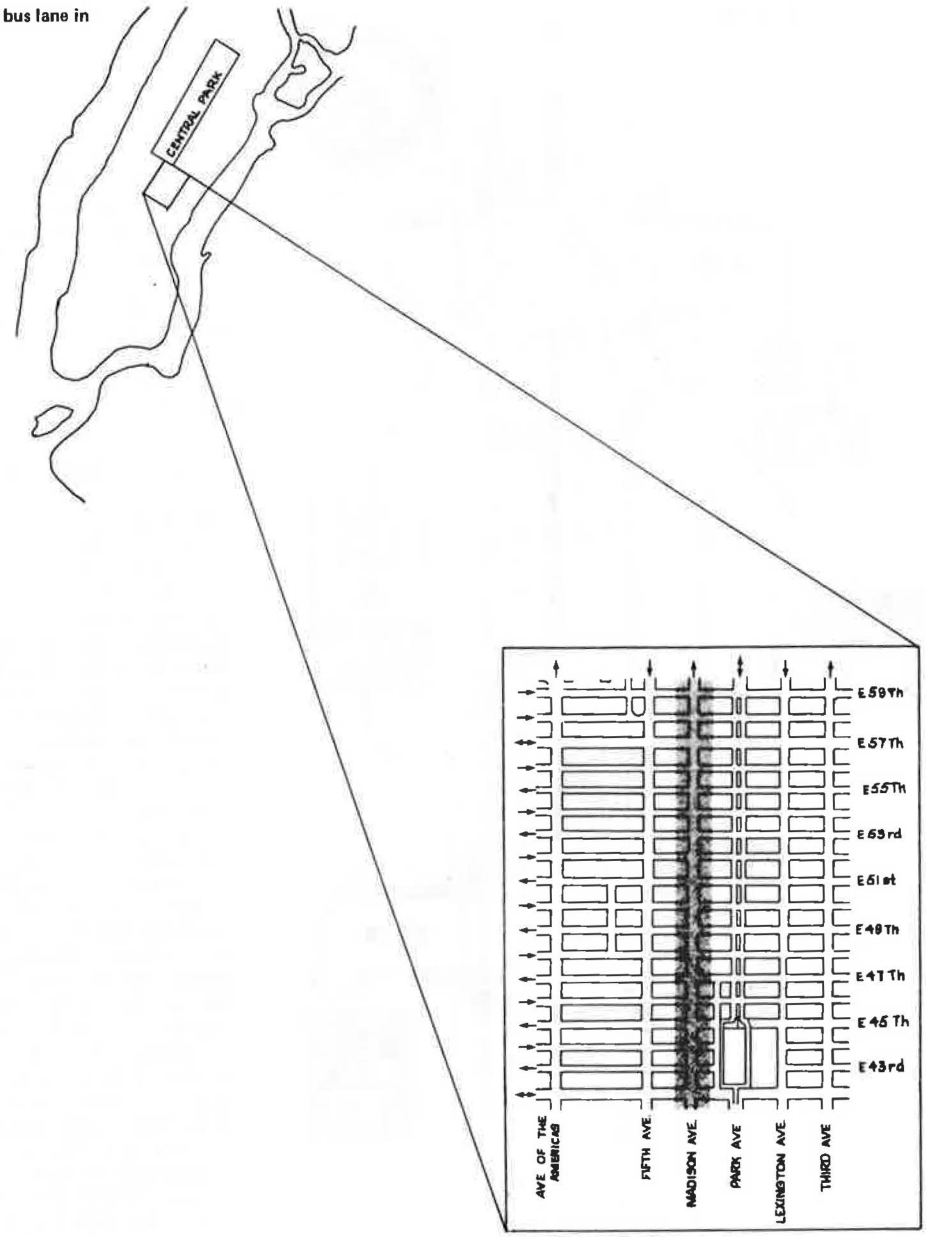
1. Roadway widths--Madison Avenue occupies an 80-ft right-of-way between 42nd and 60th Streets. The right-of-way consists of a 54-ft roadway and 13-ft sidewalks.
2. Traffic control devices--Madison Avenue is a one-way northbound arterial. Left turns are prohibited at the two-way cross streets in the project corridor. The remainder of the cross streets are one way. All intersections are signalized. There is a 27-mph northbound signal progression.
3. On-street parking regulations--Before implementation, the entire curb lane along the east (right) side of Madison Avenue was signed "No Standing, Bus Zone". Between 38th and 60th Streets, parking was prohibited along the west (left) curb, except for 54 spaces allocated for diplomats, 7 for the press, 11 for cars of handicapped drivers, and 5 for taxis.
4. Surface transit system--Between 42nd and 59th Streets Madison Avenue is directly served by 5 New York City Transit Authority (TA) local bus routes, 15 TA express bus routes, and 17 private express bus routes.
5. Land use--Both sides of Madison Avenue are characterized by office towers. At the time of project implementation, four major buildings were under construction.

Project Design

After consideration of several approaches, including single- and double-width contraflow lanes, a transit mall, and rerouting of buses, the dual-width concurrent-flow approach was selected as optimal. The final design consisted of the following elements:

1. Reorganization of bus stops along the right curb. The frequency of bus stops for local buses was changed from every other block, on average about every 500 ft, to every third block, about every 750

Figure 1. Location of Madison Avenue duel-width bus lane in midtown Manhattan.



ft. The frequency of stops for express buses was changed from an average of every five blocks (1250 ft) to every seven blocks (1750 ft). In addition, bus stops were removed from critical block faces at points of anticipated high congestion.

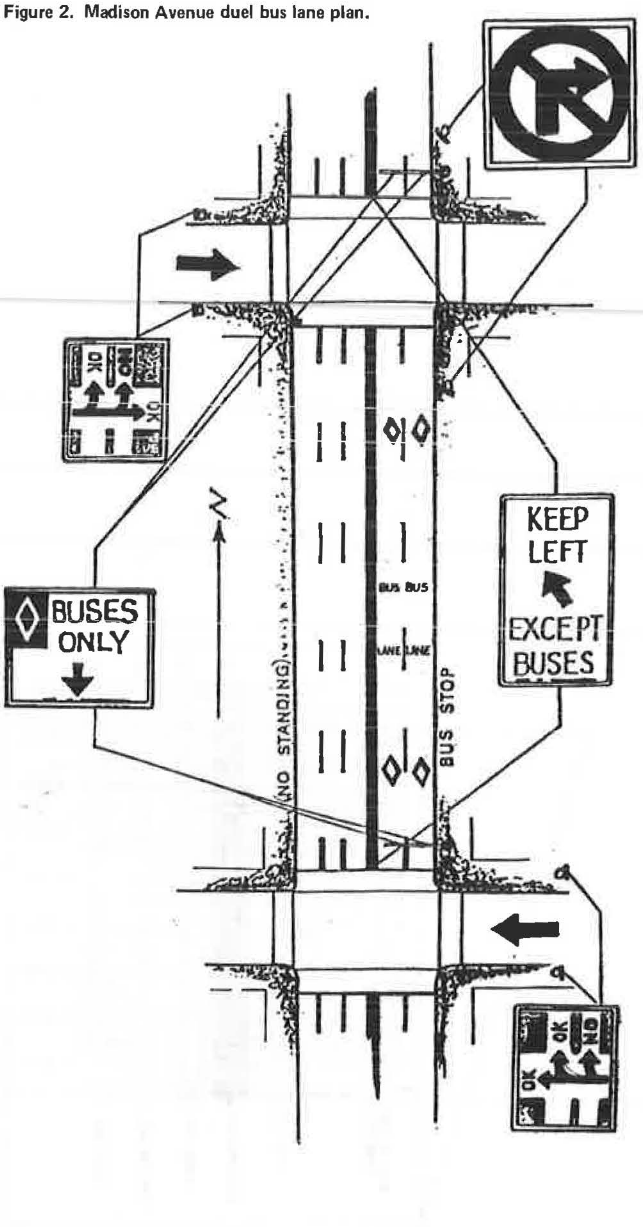
2. Removal of all parking from the left curb during hours of bus lane operations. Authorized parking was relocated to various cross streets, and taxi stands were eliminated. Replacing these were two regulations: "No Standing Except Trucks Loading and Unloading, 7 a.m.-1 p.m., Except Sunday", and "No Standing, 1 p.m.-7 p.m., Except Sunday." This was to allow vehicles to turn left from the left curb lane during hours of bus lane operations and to change the second-from-the-left lane from a turning lane to a through lane.

3. Dedication of the right two lanes exclusively for buses between 42nd and 59th Streets (0.85 mile), 2:00-7:00 p.m., weekdays. The selected cross section of the bus lane (from right to left) consists of two 11-ft lanes for buses, a 3-ft solid white thermoplastic mall to separate the bus lanes from the mixed-traffic lanes, two 10-ft mixed-traffic lanes, and a 9-ft mixed-traffic curb lane. The bus lanes are identified by overhead signs, pavement markings of thermoplastic diamonds with the word message "Bus Lane", and roll-out signs at the head of each block (see Figure 2). Vehicles from the cross streets are allowed to turn into Madison Avenue but not into the bus lane. Taxis and trucks as well as cars are prohibited from the bus lane, except as described in 5 below.

4. Prohibition of right turns. For capacity and safety reasons, and to prevent confusion in enforcement, right turns were banned from north of 42nd Street to south of 62nd Street, a distance of a little under 1 mile. Within these limits, traffic destined for areas east of Madison Avenue either had to avoid Madison Avenue or execute three left turns instead of a right turn.

5. Allowance of taxis to 46th Street. As part of the public transportation system, taxis are allowed certain privileges not accorded other vehicles. In the case of the Madison Avenue bus lane,

Figure 2. Madison Avenue duel bus lane plan.



taxis with passengers are allowed to enter the bus lane at 42nd Street and proceed without stopping to 44th or 46th Street, where they must turn right. This eliminates considerable circuity for taxis from Grand Central Station.

6. Design of termination points. The project starts at 42nd Street, where the avenue widens from 45 to 54 ft, curb to curb. Advance warning signs were placed upstream at several locations. The project ends at 59th Street, the northern boundary of the CBD where traffic generally becomes lighter, but the first eastbound street where a right turn is possible is 62nd Street. Heavy turning volumes were anticipated at this location. To accommodate these vehicles, the block face on Madison Avenue between 61st and 62nd Streets was cleared of bus stops, the parking regulations on 62nd Street were changed to eliminate standing along one curb during the hours of bus lane operation, and additional traffic signal green time was provided.

7. Hours of operation. The Madison Avenue bus lane was originally envisioned as a 24-h installation. However, as planning progressed, it became apparent that there were constituencies with strong feelings against disrupting traffic during hours when bus volumes were light. Because Madison Avenue serves primarily as an outbound artery, the strategy that was finally adopted was to establish the bus lane for the period of highest congestion (2:00-7:00 p.m.) with intense enforcement.

8. Enforcement. The decision to adopt a part-time operation implied a reliance on enforcement rather than on a physical barrier to limit violations. A staff of 24 civilian enforcement agents plus supervision was funded for one year by a federal grant.

9. Restriction on construction and street openings. Rules were established forbidding the use of Madison Avenue, including the curb lane, for storage of construction equipment or supplies between 1:00 and 7:00 p.m. All street openings except for verified emergencies were prohibited except at night.

PROJECT IMPLEMENTATION

Preliminary planning for the project was performed by the consulting firm of Edwards and Kelcey, which produced a report on the project in April 1980. As the project developed, interested community and professional groups were brought into the planning process. By the time it was implemented, 64 separate organizations had been exposed to, or participated in, its design. The planning culminated in a one-year Section 6 (Urban Mass Transportation Act of 1964, as amended) demonstration grant request for $788 000, which was awarded on September 30, 1980.

The Madison Avenue dual-width bus lane project was organized into three phases for implementation. Phase 1 was the reorganization of bus stops along Madison Avenue between 38th and 63rd Streets. This was completed by March 30, 1981. Phase 2 was the relocation of authorized parking and the change in parking regulations on the left (west) curb of Madison Avenue between 38th and 62nd Streets. This was completed by May 15, 1981. Phase 3 was the implementation of the Madison Avenue dual-width bus lane itself on May 26, 1981.

Prior to the implementation of each phase, a public-information campaign was conducted, primarily through the distribution of fliers and direct contacts with affected individuals and groups (see Figure 3). In addition, during the first week immediately following bus stop reorganization, professional staff was assigned to aid travelers looking for their new bus stop locations.

Construction of four major buildings on Madison Avenue and street openings by utilities had to be controlled for smooth implementation of the project on schedule. Special meetings were held with groups involved in these functions, and rules were laid down forbidding activity that would impact the avenue during hours of bus lane operation, except for verified emergencies. Experience has shown that this prohibition has to be constantly monitored to ensure compliance.

The implementation of the bus lane on May 26 went smoothly, owing principally to the following special procedures:

1. An early morning tour of inspection by all responsible agencies to eliminate any last minute problems.

2. The establishment of an on-site communications command post. This was a trailer supplied by the Police Department equipped with radios and telephone lines. A list of emergency telephone numbers was developed. The command post was constantly manned by the commanding officer for the traffic control agents, by the planning staff, and by the police.

Figure 3. Public information fliers.

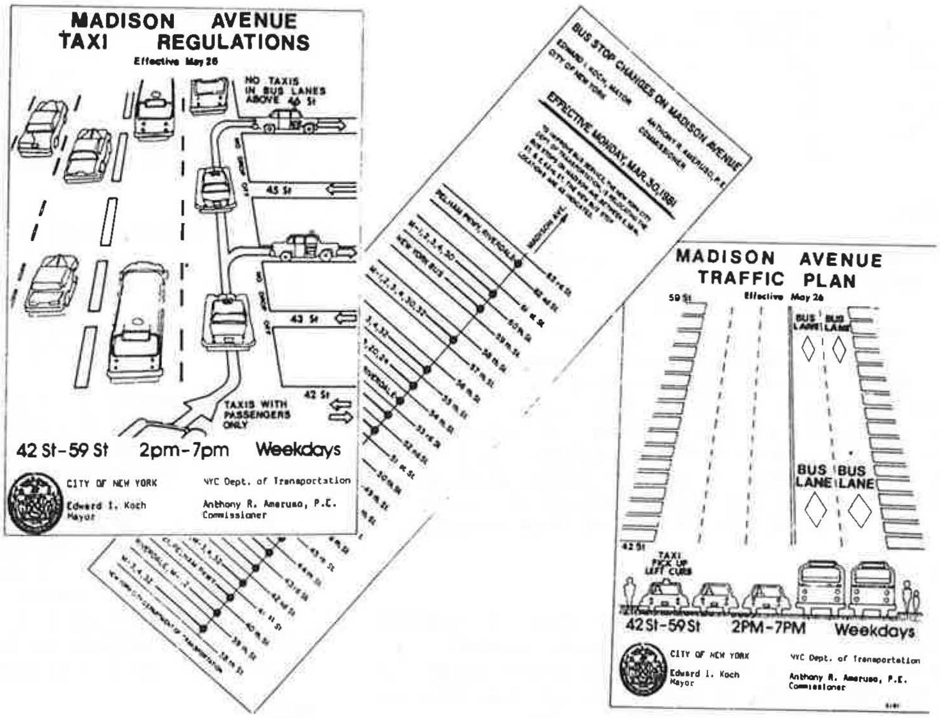
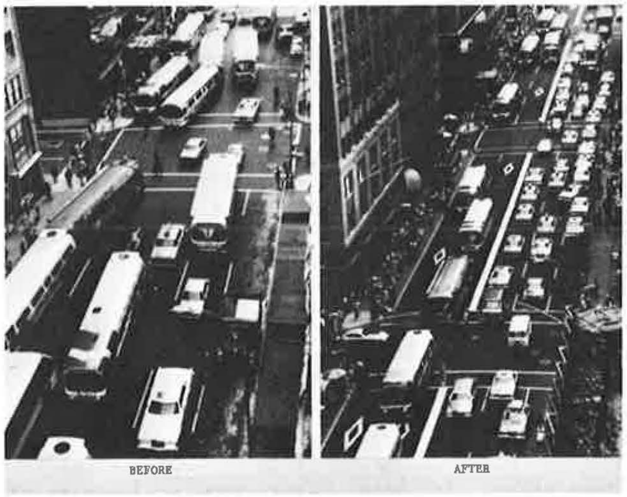


Figure 4. Before and after photographs of bus lane.



3. Constant surveillance during the first weeks at critical locations by radio-equipped members of the planning staff.
4. Establishment of a radio-equipped observation post on a high building.
5. Intensified enforcement during the first two weeks.

The normal complement of traffic and parking agents was nearly tripled and a police car was assigned to the bus lane. Six tow trucks were posted throughout the corridor to quickly respond to disabled or illegally parked vehicles, and the bus companies were required to provide trucks capable of towing buses.

IMPACT ANALYSIS

The Madison Avenue dual-width bus lane project was implemented by the New York City Department of Transportation on May 26, 1981 (Figure 4). The results show great benefits and limited adverse impacts on displaced traffic.

Bus Speeds and Reliability

Average peak-hour (5:00-6:00 p.m.) bus travel time through the project corridor decreased from approximately 18 min to less than 10 min (-45 percent). Speeds increased from 2.9 to 5.3 mph (+83 percent). During the entire 2:00-7:00 p.m. period, average bus travel time declined from 14.5 min to less than 9 min (-40 percent). Speeds increased from 3.5 to 6.0 mph (+71 percent).

An even more important effect than improvement in average bus speeds was an improvement in bus reliability. For peak hours the standard deviation of travel was cut by 59 percent (2.7 min) for local buses and by 56 percent (3.5 min) for express buses. The standard deviation as a fraction of the average travel time dropped from 26 to 18 percent for local buses and from 35 to 31 percent for express buses. In terms of the 85th percentile, travel times went from 22 to 13 min for local buses and from 25 to 11 min for express buses.

Figure 5 shows this information graphically. In comparing before and after trip times, note that the graphs are shifted to the left and are more compact. This illustrates how both trip times and dispersion in trip times were dramatically reduced.

Bus Volumes

Bus volumes for the bus lane operating period remained essentially unchanged. The total number of buses that use the bus lane for the entire 5 h is approximately 680 buses; there is a peak-hour (5:00-6:00 p.m.) average of 218 buses/h, as shown in the table below:

	Bus Volume	
Time (p.m.)	Before	After
2:00-3:00	80	78

	Bus Volume	
Time (p.m.)	Before	After
3:00-4:00	82	83
4:00-5:00	160	174
5:00-6:00	221	218
6:00-7:00	140	125
Total	683	678

Bus Ridership

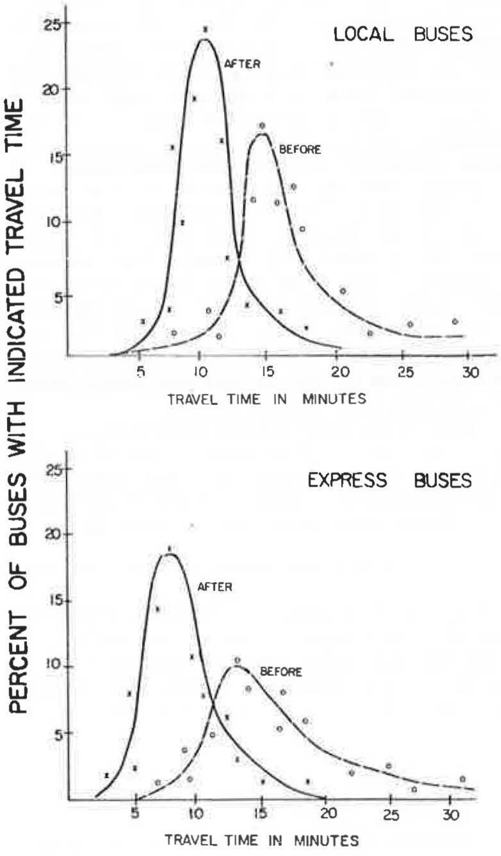
Before implementation, Madison Avenue buses carried approximately 24 000 passengers between 2:00 and 7:00 p.m. Surveys conducted after the implementation showed that the number of buses remained unchanged, yet passengers carried increased 7 percent on the local service and 4 percent on the express service. Comparable figures for Sixth Avenue, a nearby parallel avenue, showed essentially no change. It is assumed that this volume increase was due to the improvement in transit speeds produced by the bus lanes. A passenger attitudinal survey is scheduled for spring 1982 to confirm this result.

Bus Density

Density is a measurement of the number of vehicles that occupy a unit length at a given instant. In this instance, it is related to the visual impact of buses on pedestrians, a key source of dissatisfaction among local residents and merchants.

The relation used in calculating density is as follows: density = flow ÷ speed. Density calculation of buses for the peak hour (5:00-6:00 p.m.) indicated a reduction from 76 buses/mile to 42 buses/mile (-45 percent), with an associated reduction in visual impact and air pollution.

Figure 5. Madison Avenue improvement of travel time and reliability (5:00-6:00 p.m.).



Madison Avenue Automobile Traffic

The effect of the bus lanes on the remaining traffic was manifested primarily in the redistribution of traffic across the remaining lanes. A discussion of this subject is followed by an assessment of this effect on overall speeds, volumes, and other traffic measures.

Distribution of Volume by Lane

In spite of the dedication of two lanes exclusively for buses, the project did not reduce the capacity of Madison Avenue to handle the remaining traffic. This was accomplished by four actions:

1. Removal of buses from mixed traffic,
2. Removal of all parking from the west curb,
3. Elimination of right turns, and
4. Increased enforcement.

To assess these effects, lane-distribution data were collected on Madison Avenue at 47th Street, approximately at the midpoint of the project area. The data show a dramatic increase in the proportion of volume carried in lane 2 and a slight increase in lane 3. The proportion of volume carried in lane 4 dropped because it carried only buses (see Figure 6).

Speeds

Speeds on Madison Avenue improved from 5.7 to 6.0 mph during the 2:00-7:00 p.m. period. During the peak hour (5:00-6:00 p.m.) the automobile speed changes were even greater. Speeds during this period went from 4.8 to 5.3 mph, a 10 percent improvement. There were also corresponding improvements in automobile travel times, as shown in the tables below:

	Speed (min)			
Time (p.m.)	Before	After	Difference	Difference (%)
5:00-6:00	18.2	16.3	-1.9	-10.4
2:00-7:00	15.3	14.5	-0.8	-5.2

	Speed (mph)			
Time (p.m.)	Before	After	Difference	Difference (%)
5:00-6:00	4.8	5.3	+0.5	+10.4
2:00-7:00	5.7	6.0	+0.3	+5.3

Volumes

Volume counts, including buses, show an increase of about 10 percent for the 5:00-6:00 p.m. rush hour, the period with the heaviest congestion, and also the 2:00-7:00 p.m. period (see Table 1).

Classification

A comparison of after data taken in week 2 with before data shows essentially no change in the distribution of vehicle types, as shown in the table below (note that data for 2:00-4:00 p.m. were unavailable):

		Classification (%)			
Time (p.m.)	Period	Car	Taxi	Truck	Bus
5:00-6:00	Before	44	32	8	16
	After	48	32	6	14
4:00-7:00	Before	41	39	9	11
	After	47	35	7	11

Taxi Use of Bus Lane

As mentioned previously, taxis with passengers are

Figure 6. Madison Avenue volume distribution by lane at 47th Street.

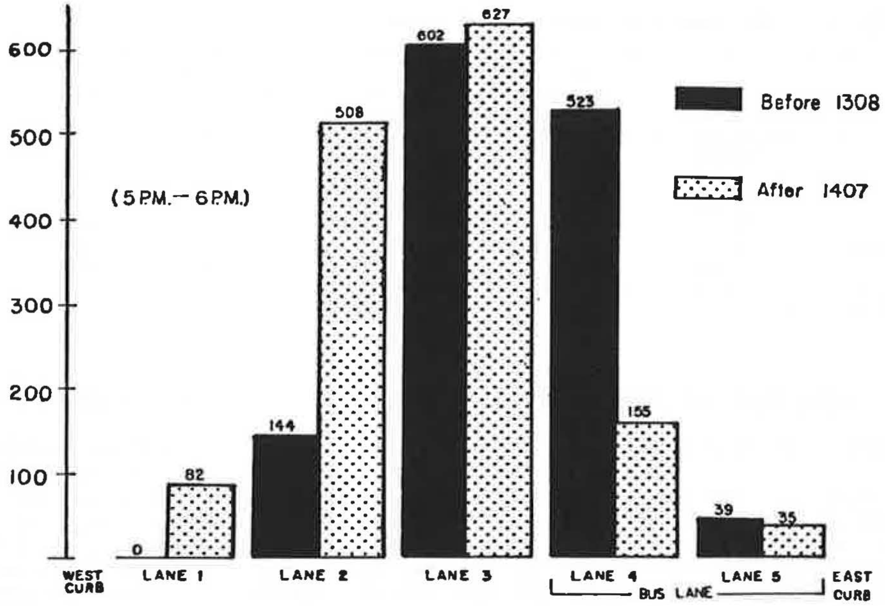


Table 1. Madison Avenue volumes.

		Volume (no. of vehicles)		
Time (p.m.)	Location	Before	After	Change (%)
5:00-6:00	46th-47th Streets	1308	1420	+8
	52nd-53rd Streets	1233	1386	+12
2:00-7:00	46th-47th Streets	6423	6797	+6
	52nd-53rd Streets	6269	6998	+12

permitted to use the bus lane between 42nd and 46th Streets, where they must turn. An analysis was conducted to determine what proportion of taxis on Madison Avenue below 46th Street took advantage of this arrangement. The figures show that about 10 percent of the total taxi volume used the bus lane during both the 5:00-6:00 p.m. rush-hour period and the whole 2:00-7:00 p.m. bus lane operating period.

Impacts on Avenues Parallel to Madison Avenue

To determine the effect of the Madison Avenue bus lane on nearby parallel avenues, the avenues were separated into two groups: northbound (which is the same direction as Madison Avenue) and southbound.

Northbound Avenues

Any effects on other avenues would be expected to manifest themselves primarily on those avenues going in the same direction as Madison Avenue because these would be the routes likely to be selected by diverted traffic. But, as shown in the previous sections, traffic engineering changes on Madison Avenue resulted in no loss of capacity. This minimized the effect on other northbound avenues.

Field data showed that average northbound speed was essentially unchanged. Between 2:00 and 7:00 p.m. the average change is only +1 percent, and between 5:00 and 6:00 p.m. the average change is -2.7 percent (see Table 2). The respective changes in volumes were also small: Average increases were 4 percent between 2:00 and 7:00 p.m. and 2 percent between 5:00 and 6:00 p.m.

Southbound Avenues

Because Madison Avenue is northbound, one would not expect southbound avenues to be very much affected. The one exception if Fifth Avenue. Because vehicles on Madison Avenue that have destinations farther east can no longer turn right, they must either divert to other avenues or make three left turns. A portion of the path of the three left turns involves Fifth Avenue, which might be adversely affected. But, as shown in Table 3, this effect is small.

The average changes in speeds on southbound avenues were slight: a -3 percent change in speed between 2:00 and 7:00 p.m., and a +5 percent increase in speed between 5:00 and 6:00 p.m. The average volume changes were -2 percent between 2:00 and 7:00 p.m. and -2 percent between 5:00 and 6:00 p.m.

Table 2. Changes in speeds and volumes on northbound avenues.

		Speed			Volume		
Avenue	Time (p.m.)	Before (mph)	After (mph)	Change (%)	Before (no.)	After (no.)	Change (%)
Third	2:00-7:00	8.1	9.5	+17	8 829	10 076	+14
	5:00-6:00	6.5	6.9	+6	1 746	1 967	+13
Park[a]	2:00-7:00	7.6	7.6	0	7 530	7 441	-1
	5:00-6:00	6.3	6.2	-2	1 704	1 591	-7
Sixth	2:00-7:00	8.5	7.5	-12	11 252	11 227	0
	5:00-6:00	7.3	6.6	-10	2 301	2 268	-1

[a] Northbound.

Table 3. Changes in speeds and volumes on southbound avenues.

		Speed			Volume		
Avenue	Time (p.m.)	Before (mph)	After (mph)	Change (%)	Before (no.)	After (no.)	Change (%)
Lexington	2:00-7:00	10.6	11.3	+7	6436	5838	9
	5:00-6:00	10.0	11.6	+16	1151	1081	-6
Park[a]	2:00-7:00	9.6	9.1	-5	6216	6221	0
	5:00-6:00	10.1	10.9	+8	1174	1178	0
Fifth	2:00-7:00	7.9	7.3	-8	8259	8421	+2
	5:00-6:00	8.1	7.8	-4	1583	1590	0

[a] Southbound.

Impacts on Streets that Cross Madison Avenue

Westbound Streets

Because of the right-turn ban, vehicles on Madison Avenue with destinations farther east are required to make three left turns. This affects primarily the block segments between Madison and Fifth Avenues. For the surveyed streets, average speeds declined from 5.0 to 4.4 mph (-12 percent) for the 2:00-7:00 p.m. period and from 5.2 to 4.9 mph (-6 percent) during the 5:00-6:00 p.m. rush hour.

Less volume data were collected than speed data, but these indicate a change in the expected direction. For the streets surveyed, the average volume between Madison and Fifth Avenues increased by 6 percent from 2:00 to 7:00 p.m. and also during the 5:00-6:00 p.m. rush-hour period.

Eastbound Streets

The effects of the right-turn ban should influence only the block segments between Fifth and Madison Avenues, with two exceptions. The first is 62nd Street. This is the first eastbound street accessible from Madison Avenue north of 42nd Street. Consequently, increased volume on this street was expected, and techniques were developed to increase its capacity, as previously described.

The second exception is the group of streets that includes 40th and 41st Streets and 42nd Street eastbound. These are the last eastbound corridors south of the bus lane and its associated right-turn ban. It was expected that these streets might absorb some of the eastbound traffic that previously turned right between 44th and 59th Streets.

The average speed on the surveyed eastbound crosstown streets declined slightly from 5.3 to 5.1 mph (-2 percent) for the 2:00-7:00 p.m. period and was unchanged for the 5:00-6:00 p.m. period (before and after speeds were 4.8 mph). The speed on 62nd Street increased 39 percent (from 4.4 to 6.1 mph) for the 2:00-7:00 p.m. period and 19 percent (from 4.2 to 5.0 mph) from 5:00 to 6:00 p.m. Excluding 62nd Street, the average volume change between 2:00 and 7:00 p.m. was -6 percent and between 5:00 and 6:00 p.m. it was -3 percent. For 62nd Street the corresponding volume figures were +22 percent for 2:00-7:00 p.m. and +15 percent for 5:00-6:00 p.m.

The speeds on 40th and 41st Streets did not decline. In fact, they increased. This implies that they were not used as shunts to the east for traffic previously turning right between 42nd and 59th Streets. This is confirmed by examination of the turning volume from Madison Avenue onto these two streets, which did not increase. It is assumed that some of this traffic made three left turns to go right farther north. The remainder presumably avoided the corridor entirely, as designed for in the original plans.

CONCLUSIONS AND RECOMMENDATIONS

The Madison Avenue dual-width bus lane imposed major changes on traffic and access patterns in one of the most intensely used corridors in the nation. In spite of this, the implementation went remarkably smoothly. Some of the important considerations that surfaced in developing and implementing the project are the following:

1. Involvement of relevant groups throughout project definition, design, and installation. About half of this contact was made at public meetings that included invited participants in favor of the project, as well as some who might be opposed. The remainder of the contacts was made at meetings to address specific issues within the context of a project that had already gained considerable momentum.
2. Support by an activist administration willing to take risks. The project involved little in the way of permanent installation and was always billed as an experiment that would be withdrawn if it failed. This stance had credibility, since the same administration had shortly before removed a bicycle lane that had proved unpopular.
3. Modest beginning. Originally conceived as a 24-h, 7-day/week facility with physical barriers to prevent violations, the project was reduced in scope to 5 h/day on weekdays without a physical barrier. This minimized the disruption and ensured that there would always be a high frequency of buses visibly benefiting from the lane. The hope is that the success of the bus lane will build support for making the project permanent and for expansion in terms of hours or to other areas.
4. Consistent enforcement. Without physical barriers, the project is completely dependent on consistent enforcement for success. For the first year this is ensured by the federal grant. Thereafter, New York City will have to fund the project. This has its drawbacks, because the city will have to resist the temptation to shift its limited number of enforcement agents from area to area in response to changing needs.
5. Initial enforcement saturation. To ensure a smooth operation during the critical initial period, normal enforcement levels were tripled and professional staff equipped with radios closely monitored every block. A radio control center and elevated observation post were set up, and arrangements were made to respond instantly to disruptions of any sort.
6. Anticipation of problems. One of the subjects we knew would be most difficult was the reduction in access to garages directly east of Madison Avenue. Because of automobile arrival patterns, we felt that the impact of the bus lane on business would be minimal, but we also knew that the garage owners might dispute this. Consequently, special before surveys were conducted to have a measure

against which the effect of the bus lane could be judged.

7. Continuing involvement of planning staff. Although the project became routine after the first two weeks, unusual conditions continued to arise, e.g., plates over street openings shifted to create hazards, construction equipment that obstructed a lane was used without authorization, enforcement personnel were shifted to other locations, etc. Continuous monitoring and interest in the project by the planning staff enabled these problems to be addressed before they seriously degraded bus lane operation.

Data-gathering efforts and analysis are continuing. In the coming period, the following topics will receive particular attention:

1. Experimentation with differing enforcement strategies, including various mixes of signing, personnel, and traffic cone placement, to determine the most cost-effective method of keeping violation rates at an acceptable level;
2. Development of benefit/cost ratios, including the real operating cost savings to the bus companies; and
3. Assessment of impact on access to cross streets where right turns are banned.

ACKNOWLEDGMENT

The project and analysis were funded by a Section 6 demonstration grant under the Urban Mass Transportation Act of 1964, as amended.

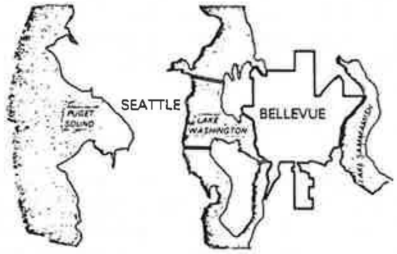
Publication of this paper sponsored by Committee on Bus Transit Systems.

Public Transit Planning by Using Interactive Computer Graphics in Bellevue, Washington

ROBERT A. WHITE, JAMES W. CLARK, AND TOMOKI NOGUCHI

The interactive graphic transit design system (IGTDS) demonstration study was conducted in the City of Bellevue, Washington. IGTDS is a set of computer programs that enable the planner to design and evaluate alternative transit systems through the use of computer graphic techniques. The IGTDS model estimates travelers' choices between automobile and transit modes for systems that serve trips from many origins to a single destination. IGTDS is easy to use; it was especially designed for transportation planners who do not have computer programming backgrounds. Facility requirements are accessed to a time-shared computer system and a computer graphics display terminal. The IGTDS demonstration study successfully accomplished its three primary objectives. IGTDS was used to evaluate different transit service concepts that ranged from the do-nothing alternative (reference case for other alternatives) to the 1990 regional transit plan with park-and-ride service to the Bellevue central business district and transit service to the Crossroads shopping center area. Comparison with the Bellevue manual sketch-planning subarea study revealed that approximately one-half as much effort was required for the IGTDS method as for manual sketch planning. The IGTDS demonstration study evaluated approximately 300 transit service designs, an increase in design productivity over the manual method by a factor of 60 to 1. The different transit system design results produced by IGTDS were presented in graphical form at a high level of detail. The graphic presentation allowed rapid comprehension of the results, and rapid feedback of information also increased understanding of the sensitivity of transportation performance to policy changes. The demonstration study showed that IGTDS is a very useful transportation sketch-planning tool.

Figure 1. Location of Bellevue.



Bellevue, Washington, is one of the principal suburbs of Seattle and has a population of approximately 80 000. Bellevue was selected for the interactive graphic transit design system (IGTDS) demonstration study because IGTDS is well-suited to planning new transportation services for small or medium-sized urban areas. Bellevue has a well-defined central business district (CBD), and the current public transportation services that serve Bellevue are provided specifically for the Seattle CBD (Figure 1).

The objectives of the IGTDS demonstration study (1) were as follows:

1. Apply IGTDS to the solution of actual transit planning problems in a real-world planning effort;
2. Develop comparisons between IGTDS and more conventional transit planning techniques in terms of design results, resource requirements, and other factors; and
3. Test the usefulness of this technology as a communication medium for facilitating decisionmaker understanding of transit patronage and cost variables in an actual transit plan development environment.

An important constraint on the first objective was to perform the study without collecting new data. That is, the input data needed for IGTDS were obtained from previous transportation studies and from readily available local sources.

BACKGROUND

Currently, passenger transportation to and within the Bellevue CBD is provided primarily by private automobiles. In 1979 only about 2 percent of all trips to the CBD were made by public transportation. Island-like building developments surrounded by large parking lots, lack of pedestrian amenities, and wide arterial streets with many curb cuts for driveways cause numerous automobile-pedestrian conflicts. These factors result in an environment with low pedestrian attractiveness that is difficult to serve efficiently by public transportation. As a result of the heavy reliance on personal automobile transportation, the street system is congested throughout the day, and the single freeway interchange that serves the CBD has already reached its capacity during peak demand periods. The anticipated growth of the CBD during the next decade is expected to exacerbate these problems if reliance on personal automobile transportation continues as at present.

In recognition of this situation, city officials, business leaders, and citizens have been searching for ways to increase the use of public transportation in Bellevue. The Mayor's CBD Action Committee has established the objective that, by 1990, 20 percent of all trips to the CBD during peak hours arrive by transit as described in the recent Bellevue CBD Action Plan (2). Transportation planning for the 1980s is being conducted by the regional transit operator, Metro Transit, in a process called Metro TRANSITion Phase IV.

In addition to developing plans for Metro's entire system, this study has produced a Bellevue subarea study that focused on the particular needs of the Bellevue CBD (3). Parallel to Metro's studies, Bellevue applied to the Urban Mass Transportation Administration (UMTA) for a grant to demonstrate the usefulness and effectiveness of IGTDS in a planning process such as the Metro TRANSITion Phase IV studies. In 1978, UMTA awarded Bellevue technical grant and research and development grant funds to conduct the IGTDS demonstration study under Sections 9 and 6, respectively, of the Urban Mass Transportation Act of 1964, as amended. The General Motors (GM) Transportation Systems Center (TSC) of Warren, Michigan, was selected as the project consultant and began the demonstration study in June 1979.

GRAPHICS LABORATORY

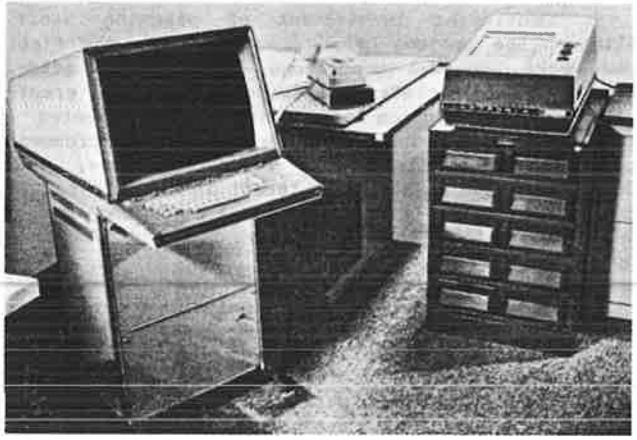
A graphics laboratory was established at the Bellevue City Hall. The laboratory consisted of a work space and the computer graphics terminal equipment needed to operate IGTDS. Three pieces of equipment were obtained: a Tektronix model 4014-1 graphics display terminal, which has a 15-in-wide by 11-in-high display screen; a Bell Systems Model 212A data communications unit, which supports communications at 120 characters/s; and a Tektronix model 4631 hard copy unit (Figure 2).

The City of Bellevue was provided with access to the IGTDS software system that GM TSC designates as IGTDS GM TSC Release No. 1 (4). This version of IGTDS was operational on an IBM Model 3033 computer system (IBM's replacement for the System 370 series computer), which uses the time sharing system (TSS) operating system. The computer facilities were located at the GM Technical Center in Warren, Michigan.

LOCAL PLANNING GROUP

A local planning group was formed to demonstrate the capability of IGTDS to local transportation planners. The group consisted of two planners from Bellevue, two from the Puget Sound Council of Governments (PSCOG), two from Metro Transit, and two from local consulting firms. Members were trained to use IGTDS to design fixed-route transit service to an activity center, and they participated in performing sensitivity analyses on the high-performance designs.

Figure 2. Computer graphics terminal equipment set.



A questionnaire was distributed to obtain reactions to and assessments of IGTDS. The members of the planning group were unanimous in their ratings of IGTDS as easy to learn, easy to use, and effective in providing easy-to-interpret results. Members were also able to envision the use of IGTDS in situations in which they were involved professionally, including (a) macrolevel sketch planning and (b) microlevel policy analysis, such as parking policies, fare policies, routing options, and route productivity.

However, members also expressed generally a low level of confidence in the IGTDS predictions of ridership and costs, since the modal-choice model had not been calibrated with survey data. Other improvements that they felt would improve the usefulness of IGTDS to problems they faced were the following:

1. Ability to consider at least two destinations simultaneously;
2. Better modal-choice model; for example, a model that includes income of travelers as a variable;
3. Better cost-estimation models; and
4. Ability to handle a larger problem.

DATA DEVELOPMENT

In order to apply IGTDS, three kinds of data are needed: transportation supply-side data (including transportation network, vehicle characteristics, and cost parameters needed to estimate transit system operating costs), travel-demand data, and traveler-behavior data.

The transportation supply-side data included several elements. A network structure that represents the street system within Bellevue was obtained from the PSCOG network developed for the regional transportation plan update. Link travel times for driving were obtained from this network, and travel times for the transit mode were obtained from PSCOG's transit network. For the portion of the region outside of Bellevue, the network structure of the Bellevue microzone forecasting model was used. The network data base was digitized by using a network editor (5) and digitizing tablet. A plot of the digitized network is shown in Figure 3. A close-up, 10x10-mile view of Bellevue is shown in Figure 4. (Both figures were produced by IGTDS with annotations added manually.)

Operating cost data were taken from the Bellevue subarea study. Three categories of vehicles were

Figure 3. IGTDS network—Bellevue with external network.

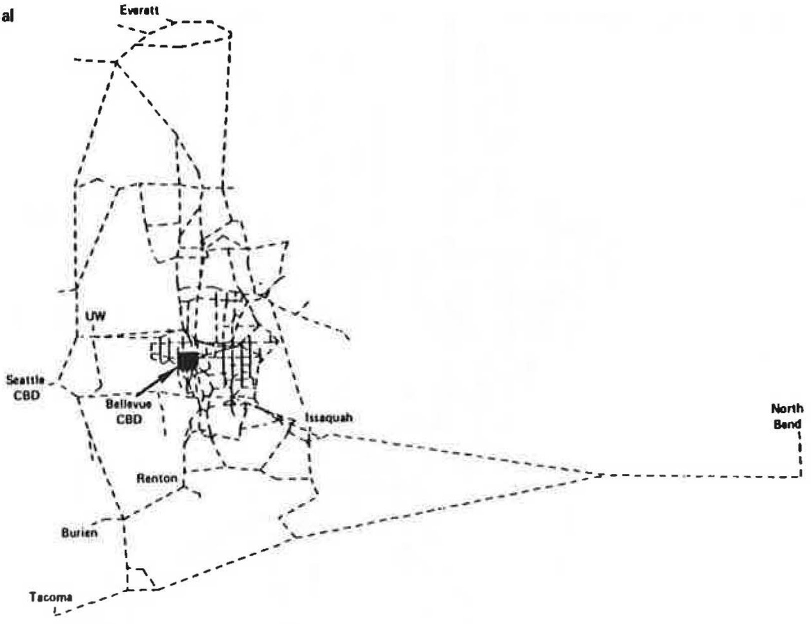
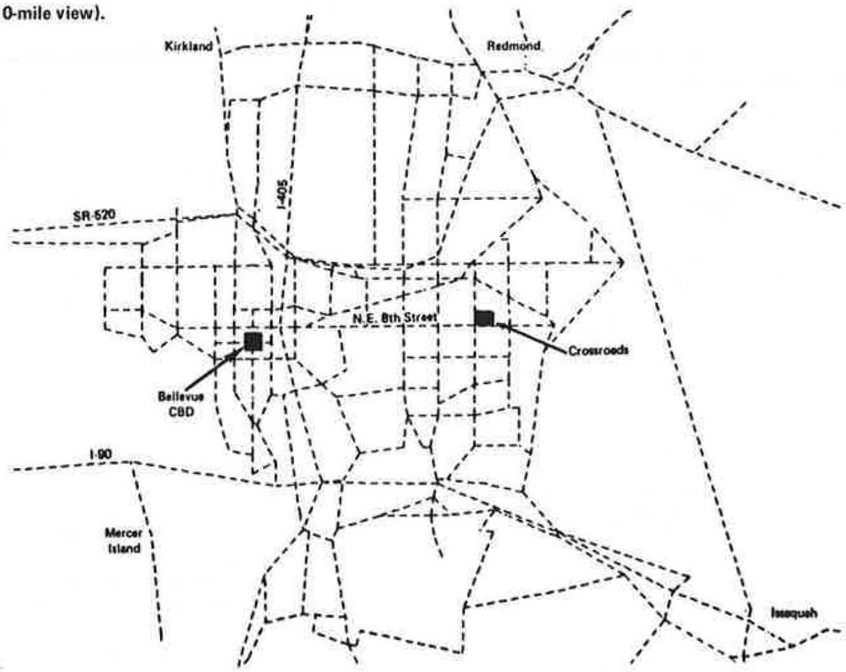


Figure 4. IGTDS network—Bellevue (10x10-mile view).



considered: (a) a standard 45-passenger coach that costs $18/h to operate, (b) a 72-passenger articulated vehicle that costs $21/h, and (c) an 11-passenger van that costs $15/h. These costs reflect Metro Transit's 1978 average costs.

Two sets of travel-demand data were developed, one for the purpose of comparing IGTDS with the Bellevue subarea study and one for the purpose of comparison with the regional transportation system study. Both travel-demand data sets were originally obtained from PSCOG. The destination zones selected were the Bellevue CBD and the Crossroads shopping center in Bellevue. Figure 5 shows a symbolic representation of Bellevue's forecast of peak-hour travel demand to the CBD in the year 1990. The triangles represent demand values in their original form by Bellevue micromodel transportation planning zones. The diamonds represent travel demand allocated to IGTDS network nodes after transformation by the zone-to-node data conversion system (ZONOCO) (6). (This figure was drawn off-line with a Calcomp plotter by using the IGTDS data base.)

IGTDS employs a multinomial logit model of modal choice to predict ridership for a transit system design alternative. To apply the model, the impedances of trips by each of three modes are calculated for each node of the network. The trip impedance is a linear function of trip components X(i,m), where the X's represent the various time and money costs of making a trip by each of the available modes.

The trip components are multiplied by the coefficients a, which represent the tripmakers' valuations of the time and money costs of the components of the trip. Mathematically, the impedance I(m) of a trip

Figure 5. 1990 travel demand to CBD—Bellevue forecast.

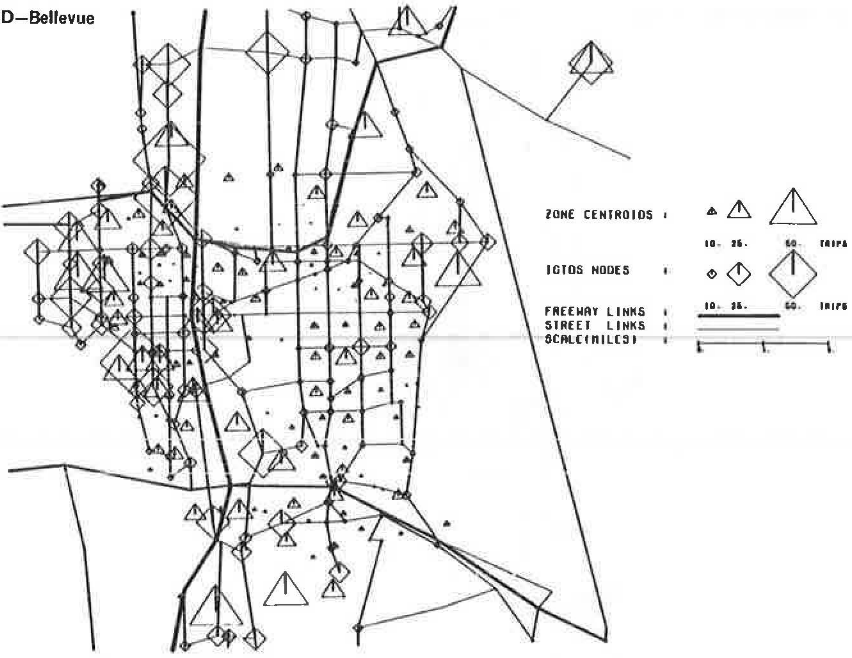
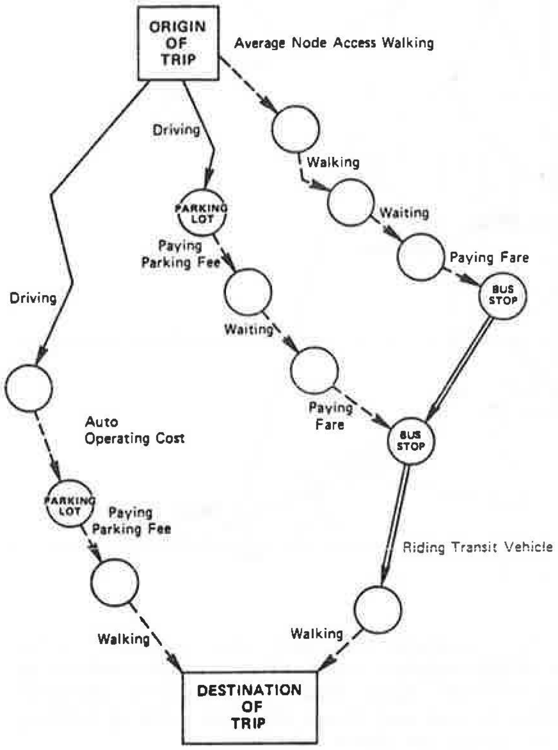


Figure 6. Trip components defining a trip from origin to destination via walk-and-ride, park-and-ride, and drive modes.



by mode m is written as follows:

$$I(m) = a(o) + \mathrm{sum}(i;i=1,k)\, a(i)*X(i,m) \qquad (1)$$

where k is the number of components that comprise a trip by the mode. The probability that a tripmaker will choose mode m is then given by the logit equation:

$$P(m) = \exp[-I(m)]/\mathrm{sum}(n;n=1,3)\, \exp[-I(n)] \qquad (2)$$

Table 1. IGTDS impedance coefficients.

Coefficient	Value	Remarks
W-R CONSTANT	216.0	Includes 6 min access and egress
P-R CONSTANT	199.0	Includes 4 min egress
DRIVE CONSTANT	0.0	
WALK	8.55	
DRIVE	5.62	Includes driving cost, assuming 30 mph average speed and $0.075/mile
WAIT	8.55	
RIDE 1	3.42	
RIDE 2	3.42	
RIDE 3	3.42	
STAND	6.84	
FARE	58.56	
FEE	29.28	Median income = $21 000; note, all coefficients above are divided by this value
EXPNT	150.00	

The IGTDS trip components for the three modes are shown in Figure 6.

The behavioral data required by IGTDS are values for the impedance coefficients. In order to facilitate the comparison of the application with the Metro TRANSITion studies, values for the impedance coefficients were chosen to make the IGTDS logit model as similar as possible to the Metro TRANSITion modal-choice models.

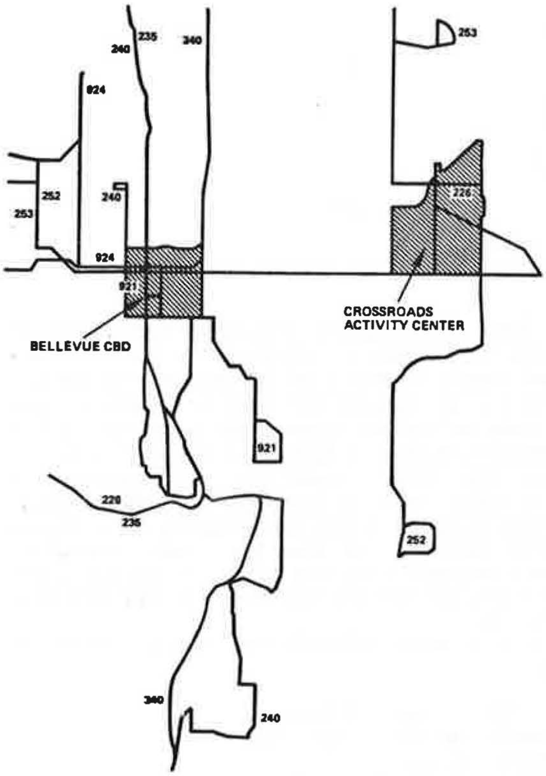
In order to compare the results of the IGTDS study with those of the Bellevue subarea study, the manual sketch-planning modal-choice model was used to estimate coefficients for the IGTDS logit model effected by means of an intuitive process based on the experience of the Bellevue, PSCOG, and Metro transportation planners who comprised the local planning group. For example, they felt that if the impedance of a trip by transit equaled the impedance of driving, then the probability of choosing transit should be 40 percent. Similarly, in comparing the IGTDS model with the one used in the regional transportation planning process, a number of simplifying assumptions were made in order to adapt the regional model coefficients for use with IGTDS.

The final coefficient values in the form required by IGTDS are shown in Table 1.

Table 2. Comparative performance of alternatives.

		Alternative						
Objective	Goal	Do-Nothing	Do-Nothing with Park-and-Ride	1990 System Plan	1990 System Plan with Park-and-Ride	CBD Circulator	Transit Mall	Crossroads
Transit riders (morning peak hour)	Maximum	167	490	290	850	783	337	65
Revenue/cost (%)	>60.0	22.0	30.5	57.8	45.1	53.6	60.0	15.3
Within 5-min walk of a stop (%)	>47	25	35	45	45	49	47	40
Within 5-min drive of a park-and-ride lot (%)	>52	-	60	-	40	18	-	-
No. of stops less than 15 min avg access	0	6	1	0	0	0	0	5
No. of park-and-ride lots less than 10 min avg access	0	0	0	-	0	1	-	-

Figure 7. Transit routes serving Bellevue CBD during morning peak hour in 1980.



TRANSIT SYSTEM ALTERNATIVES

To compare IGTDS with the manual sketch-planning method of the Bellevue subarea study, a series of case studies was designed to correspond to the five alternatives evaluated by the subarea study. Each case study illustrated the application of IGTDS to a particular service concept. The case studies were as follows:

1. Do-nothing alternative (maintain the existing transit service to the Bellevue CBD),
2. Do-nothing alternative with park-and-ride,
3. 1990 regional transit plan (includes high-occupancy-vehicle lanes on the Interstate freeway that serves Bellevue and on certain arterials, and assumes automobile travel times would double in selected arterial corridors, but transit travel times would remain at their 1980 levels),
4. 1990 regional transit plan with park-and-ride,
5. CBD circulator (the 1990 regional transit plan with a parking lot on the fringe of the CBD; distribution of travelers from the parking lot to be provided by a CBD circulator service),
6. CBD transit mall (a new transit-only access ramp would join the mall and the freeway), and
7. Service to the Crossroads shopping center.

HIGH-PERFORMANCE DESIGN OBJECTIVES

A high-performance transit system was designed for each of the above alternatives (except the do-nothing alternative) by using IGTDS. The following design objectives were established as the criteria for measuring achievement of the high-performance design concept:

1. Maximize transit ridership,
2. Maximize the ratio of transit system revenue to operating cost (the total revenue should be greater than 60 percent of the total operating cost),
3. Maximize accessibility to the transit system (for trips that originate within Bellevue, 90 percent of demand should be within a 5-min walk of a transit stop and 100 percent of demand should be within a 5-min drive of a park-and-ride lot), and
4. Minimize average access time at each stop (for service within Bellevue, average walk access time should be less than 15 min and average drive access time should be less than 10 min).

Table 2 compares the performance of each of the alternatives developed with IGTDS in relation to the performance objectives.

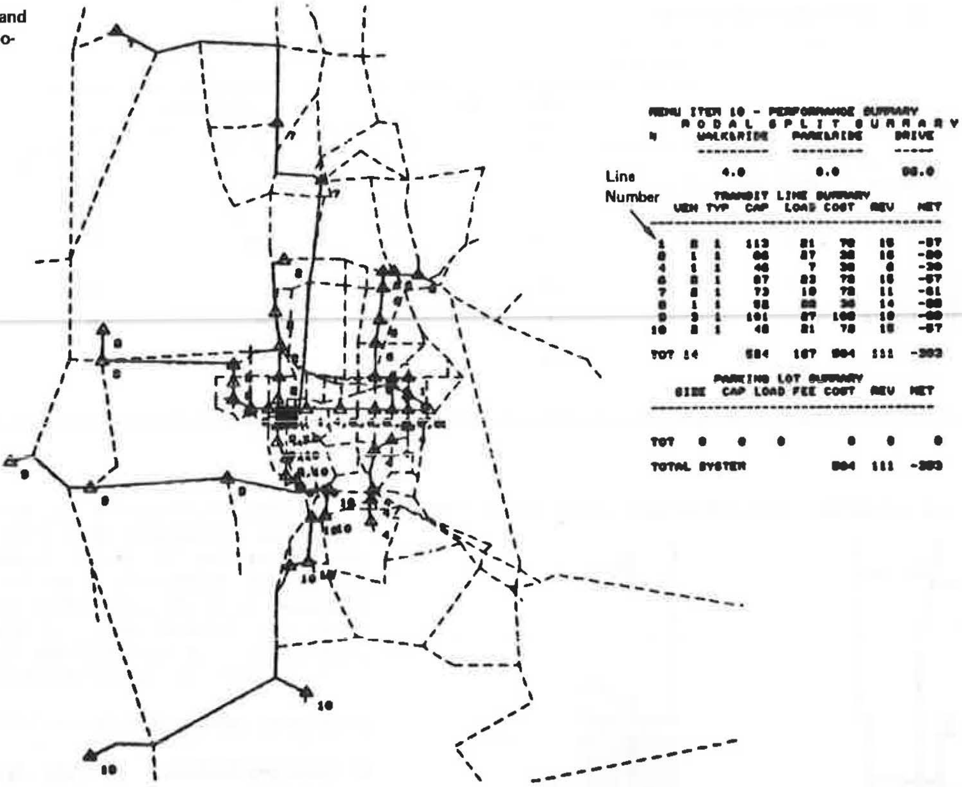
DESCRIPTION OF SELECTED DESIGN ALTERNATIVES

Do-Nothing Alternative

The objective for the do-nothing alternative was to replicate as closely as possible with IGTDS the transit system as planned for 1980. Figure 7 is a map of Bellevue that shows transit routes that provide service to the CBD during the morning peak hour.

The transit route network that was designed to model the do-nothing alternative and the performance summary report that IGTDS produced are shown in Figure 8. Parameters that were input to IGTDS were free parking at the destination and a 1-min destination walk time for the drive mode. Fares were set at $0.30 for all routes except numbers 8 and 9, which had a zone fare of $0.50 for trips from Seattle, which corresponded to Metro Transit's fare

Figure 8. Transit route network and IGTDS performance summary—do-nothing alternative.



```
MENU ITEM 10 - PERFORMANCE SUMMARY
     M O D A L   S P L I T   S U M M A R Y
%    WALK&RIDE    PARK&RIDE     DRIVE
     ---------    ---------     -----
        4.0          0.0         96.0

          TRANSIT LINE SUMMARY
  VEH TYP  CAP  LOAD COST  REV   NET
---------------------------------------
1  2   1   112   21   72   15   -57
3  1   1    66   17   36   16   -20
4  1   1    46    7   36    6   -30
6  2   1    87   23   72   15   -57
7  2   1    73   19   72   11   -61
8  1   1    58   22   36   14   -22
9  3   1   101   37  108   19   -89
10 2   1    48   21   72   15   -57

TOT 14         584  167  504  111  -393

       PARKING LOT SUMMARY
  SITE  CAP LOAD FEE COST  REV   NET
---------------------------------------

TOT  0     0    0        0    0     0

TOTAL SYSTEM            504  111  -393
```

structure in 1978. Standard 40-ft coaches (vehicle type 1) were assigned on all routes. The alternative required a total of 14 vehicles (to provide 28 peak-period hours of service daily) at a cost of $504. IGTDS predicted a ridership of 167, or 4 percent of the total demand, which yielded revenues of $111/day. (Figures 8 and 9 are graphic output displays produced by IGTDS; some annotations were added manually.)

1990 Regional Transit Plan with Park-and-Ride Service

The 1990 regional plan with park-and-ride service used the same transit route network as the 1990 system plan alternative, but it added five park-and-ride lots at the locations shown in Figure 9. Because of the additional vehicles and demand attracted by the park-and-ride service, vehicles with greater capacity were needed on routes that served the park-and-ride lots. The performance summary display for this alternative is also shown in Figure 9. The design required 23 vehicles at a daily cost of $846, plus $700 as the cost of providing 700 parking spaces. The system attracted 850 riders and had a revenue/cost ratio of 45.1 percent. This design was selected as the high-performance design because it very nearly achieves the goal of 20 percent transit ridership in 1990 that Bellevue's CBD Action Committee has established. This goal appears to be achievable, according to this IGTDS result, even if parking remains free in the CBD, provided that convenient park-and-ride service be established and that relatively low fares be charged. A charge for parking in the CBD would result in greater transit ridership as well as allowing for higher transit fares, which would permit a larger fraction of transit operating cost to be recovered.

SENSITIVITY ANALYSIS

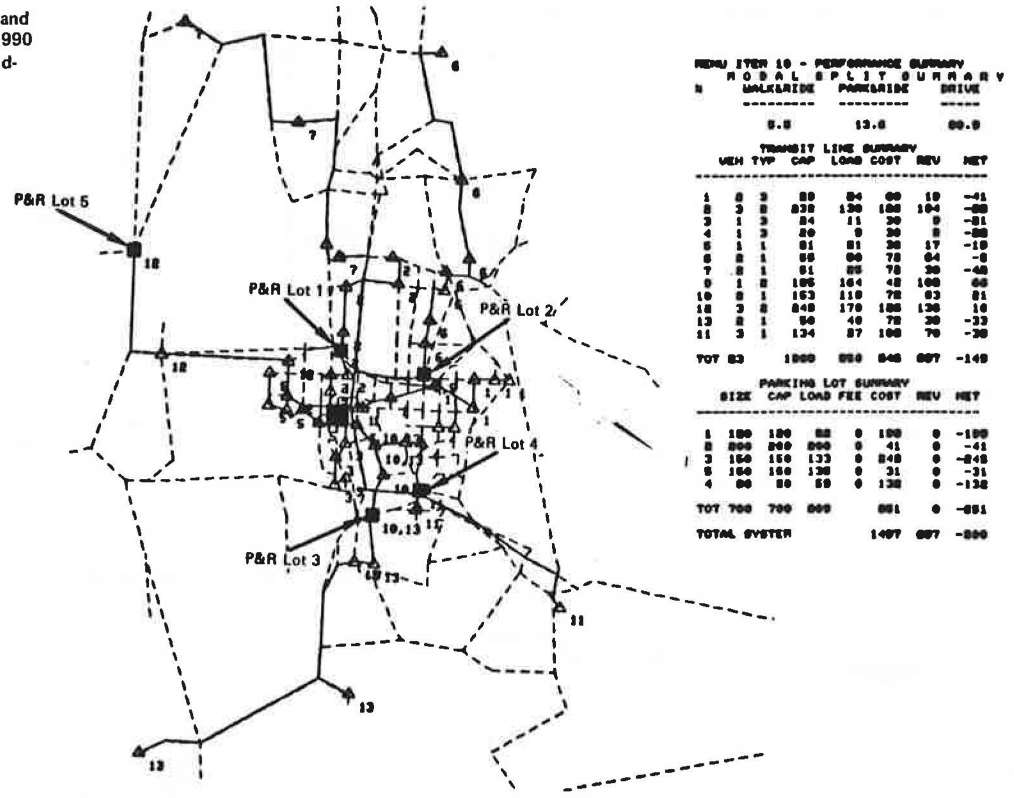
Because IGTDS was designed to permit interactive input of data and to allow immediate prediction of the effects of changes in input parameters, it can be applied easily to the task of sensitivity analyses. That is, an analyst can vary a parameter over some range of values and obtain ridership, cost, and revenue results within a span of a few minutes. To illustrate the IGTDS capability for sensitivity testing of variables that have policy implications, the planning group selected six analyses to perform. The city's interest in the particular variables chosen were derived from its need for policy guidance to achieve objectives set forth in the Bellevue CBD Action Plan.

The six variables selected for testing were the following:

1. Destination parking fee,
2. Transit service frequencies,
3. Transit fares,
4. Transit route and stop locations,
5. Travel-demand levels, and
6. Logit model coefficients.

The reference design used in all of these tests was the 1990 system plan alternative with park-and-ride service. Ideally, one variable would be identified and varied while all other variables were held constant in order to isolate the effects of the selected variable. In practice this is not always possible. For example, a policy that increases transit ridership must very quickly deal with capacity limits in the transit system. If capacity is increased in order to serve the added demand, then service improvements also generate additional ridership, i.e., a multiplier effect; ridership is increased beyond the effects of the original policy variable. Other variables, such as route locations, are not easily isolated as a single variable, which causes the sensitivity analysis to be performed in a more qualitative, descriptive manner than a quantitative, trade-off curve fashion.

Figure 9. Transit route network and IGTDS performance summary—1990 regional transit plan with park-and-ride service.



Destination Parking Fee

The average parking fee at the destination can be easily isolated, and it conveniently affects all who choose the drive mode. The effect on transit ridership of raising the all-day parking fee from $0 to $3 is shown in Figure 10. (The graphs shown in Figures 11 through 17 were made with the EASYGRAPH subset of the Tektronix PLOT-10 software.) This analysis has a feedback effect because transit service must be improved in order to accommodate the additional demand, and the service improvements themselves contribute to the added transit demand.

Figure 11 illustrates the total service increase in terms of cost. At parking fees less than $1, adequate capacity exists to handle the increase in transit ridership from about 20 to 22 percent of all trips. As the parking fee increases from $1 to $3, transit service costs rise from $1545 to $2148. The combined effect of the destination parking fee of $3/day and added transit service bring the transit ridership to 30 percent of the total.

Transit Service Frequencies

The testing of service frequencies is not as easy as it might be because IGTDS does not permit direct specification of headways. Instead, the number of vehicles is selected to serve each line and IGTDS calculates the headways. Thus, when several transit lines exist, the selection of numbers of vehicles for each line will result in a range of headway values. The curve for transit ridership and service frequency that results is then a region bounded by two envelope curves (Figure 12). The envelope curves represent the upper and lower headway limits for the set of transit lines for a specific number-of-vehicles allocation. The curves in the figure result from a total vehicle allocation that ranges from 15 vehicles (yielding between 1 and 2 vehicles/h on every line) to 72 vehicles (yielding between 6 and 7.5 vehicles/h on every line). A point of diminishing returns can be observed in the figure at somewhere between 4 and 6 vehicles/h, which corresponds to headways between 10 and 15 min. Beyond these service levels very little additional transit ridership is induced by the associated reductions in waiting times.

Transit Fares

Only flat-fare schedules were examined; that is, no zonal increments were included in the fare structures. One-way transit fares were varied from $0 to $1.50, and the transit ridership varied in a virtually linear relation from 22 to 14 percent of the total. The result is shown in Figure 13.

Travel-Demand Levels

The travel-demand data were derived from forecasts of population and employment growths. Because forecasting of growth is not an exact science, it is of interest to test the sensitivity of IGTDS results to demand levels in order to ascertain the effects of possible forecasting errors. In particular, in light of the rapid population growth that the Seattle metropolitan area has been experiencing, the Bellevue planners are concerned that the forecasts of growth may be too low. They were interested in the effects of greater-than-expected growth in selected areas. Accordingly, travel-demand values were selectively increased along certain arterial corridors within Bellevue. These corridors are shown in Figure 14. Three scenarios were developed in which a demand increase of 50, 100, and 150 percent, respectively, was proposed at each network node along these arterials over the forecast 1990 travel-demand values.

The 50 percent increase scenario was numerically 402 additional trips, which was an increase of 9 percent in the total number of trips to the Bellevue

Figure 10. Transit modal split as a function of destination parking fee.

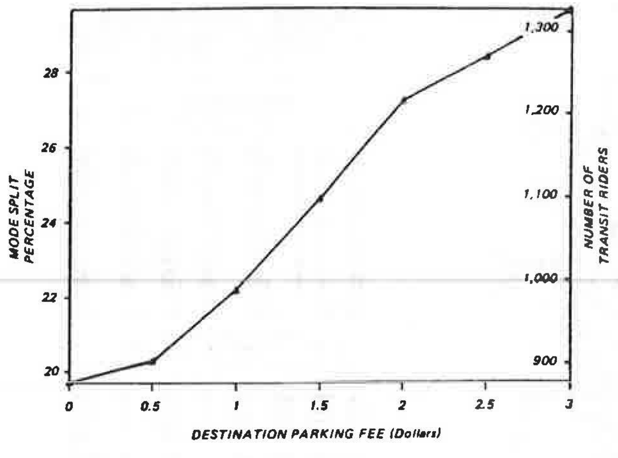


Figure 11. Operating cost and revenue versus destination parking fee.

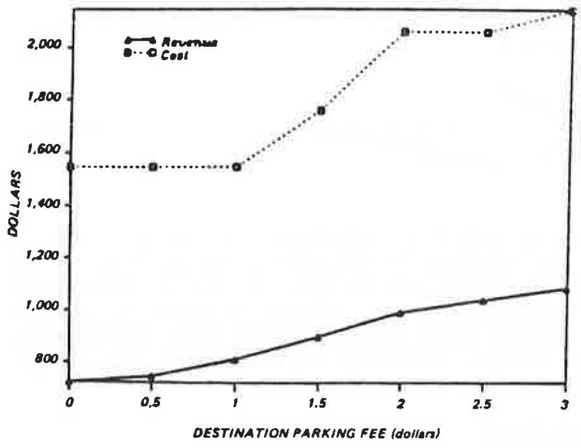


Figure 12. Transit modal split as function of service frequency.

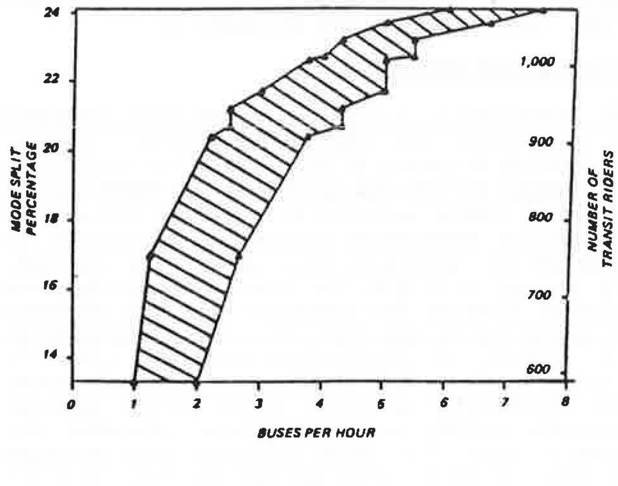


Figure 13. Transit modal split as function of transit fare.

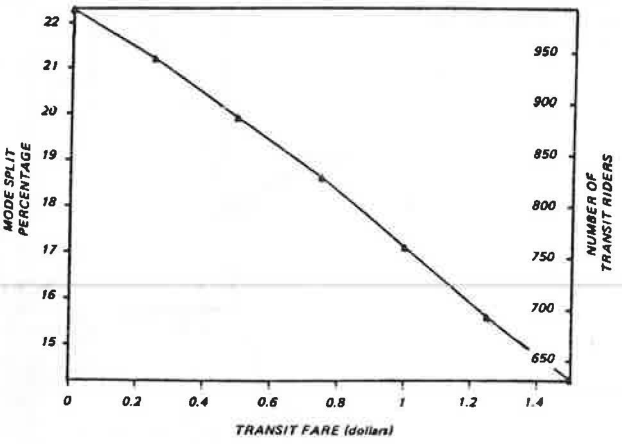


Figure 14. Selected arterial corridors with increased travel demand.

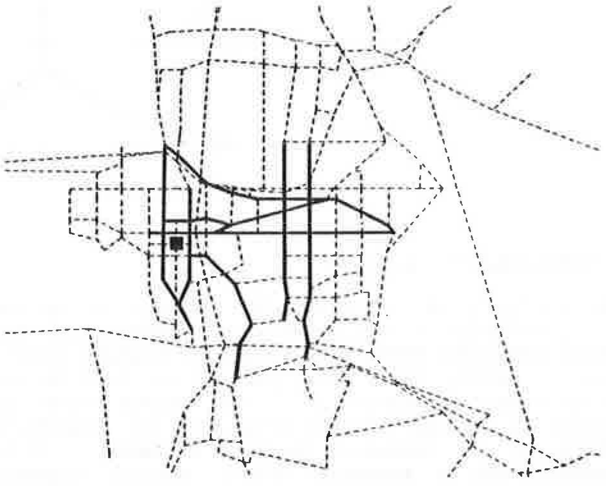


Figure 15. Increase in transit ridership and cost over 1990 forecast level for selected increases in demand.

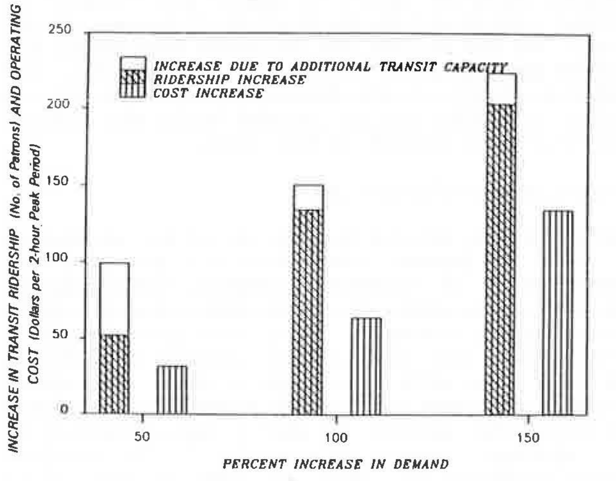
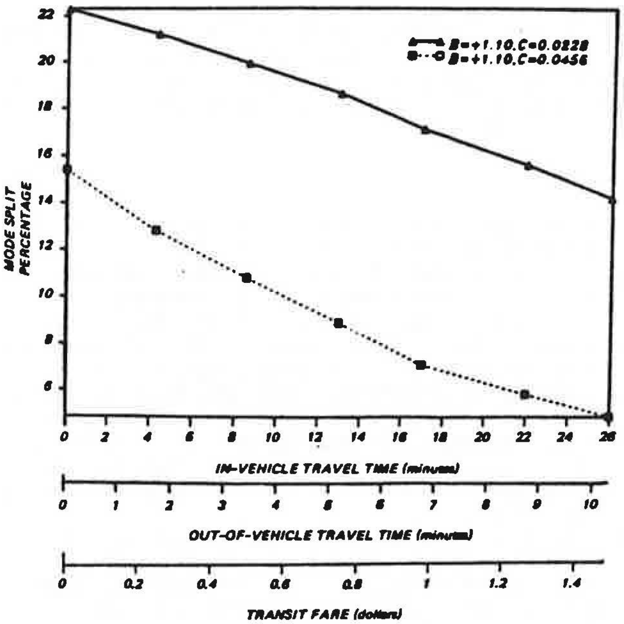


Figure 16. Transit modal split as a function of logit-model coefficients.



CBD. The 100 percent increase scenario was 781 actual trips, which was a 17 percent increase in total trips. (The 100 percent increase level is less than twice the 50 percent increase level due to rounding upwards of half trips in order to express all demand values as whole numbers.) The 150 percent increase scenario was 1183 actual trips, which was 26 percent of the forecast number of trips for 1990. The incremental transit ridership and cost results are shown in Figure 15.

In each demand-increase scenario, transit capacity limits were reached. In order to isolate the effects of the demand-increase effects as much as possible from the effects of adding transit capacity, the following procedure was used. After changing the demand, the modal-choice model was exercised to obtain an initial ridership prediction by using the transit service design with no changes from the previous case. The results were then examined to determine where capacity limitations had developed. Park-and-ride lots were increased in size where needed, and larger-capacity vehicles were added to transit lines that had reached capacity. The modal-choice model was then exercised again to determine the new ridership levels. For example, in the 50 percent increase scenario, ridership increased by 52 over the base case with no capacity increase. After increasing the size of one park-and-ride lot and changing from van service to standard coach service on one transit line at an incremental cost of $32, ridership increased by another 47. Figure 15 thus shows a total transit increase of 99 riders and a cost increase of $32 for the 50 percent increase scenario.

In all cases, the needed increases in capacity were obtained by changing to higher-capacity vehicles with no increase in service frequency. It is possible that the higher demand levels would support increased frequencies in a cost-effective manner. This latter situation is more complex to analyze, since IGTDS uses the service frequency in its modal-choice calculation. Increasing frequencies decreases waiting times, and IGTDS will predict increased modal splits as a result. Thus, potentially even greater ridership (at greater cost) increases could be obtained than are shown in Figure 15, but the multivariate type of analysis is beyond the scope of the simple sensitivity analysis that this exercise was intended to illustrate.

Logit Model Coefficients

Testing of sensitivity to values of the logit model coefficients was performed in conjunction with the development of their values, as explained in the data development section. Results of the sensitivity tests aided in the decision about values to be used for design development and other sensitivity studies. Illustrated in Figure 16 is the effect on transit modal split (walk-and-ride plus park-and-ride) of increasing transit impedance (equivalent axes are drawn to show how this increase could come from fare, in-vehicle time, or out-of-vehicle time) for two different levels of the impedance conversion coefficient C. The value B = 1.10 in the figure is the modal constant value that yields a modal split of 40 percent when all three modes have equal impedances.

EVALUATION

IGTDS demonstration study successfully achieved its primary objectives. Planners can greatly increase their productivity in designing and evaluating alternative transit systems by using this interactive graphic sketch-planning technique. Comparison of the IGTDS demonstration study with the Bellevue manual sketch-planning subarea study revealed that approximately one-half as much effort was required for the IGTDS method as for manual sketch planning.

The IGTDS evaluated approximately 300 transit service designs, each of which cost about $1 of computer expense and required an average of 5-7 min of elapsed time on the computer terminal. Preparing the initial data base (the network and travel demand) required approximately $1000 of computer expense and 8 person-weeks of effort. On the basis of the number of design alternatives considered per unit of design cost, using IGTDS to design transportation services for Bellevue showed an increase in design productivity over the manual method by a factor of 60 to 1.

The transit system designs produced by IGTDS were presented in graphical form at a high level of detail. We believe that the results obtained by using IGTDS were much more easily understood because of both the graphic presentation and rapid feedback of answers through interactive computing.

Predictions of transit ridership and cost by the two studies were comparable. The IGTDS results depended importantly on whether or not park-and-ride service was included. However, we believe that this reflects the design of the modal-choice model rather than the inherent nature of park-and-ride service. The two studies obtained very different vehicle requirement results due to quite different procedures for estimating vehicle requirements.

The comparison of IGTDS with the regional systems planning process was more difficult because of the great difference in the scope of the two projects. In terms of effort and cost involved, the Metro TRANSITion Phase IV planning effort was more than an order of magnitude greater. It was also difficult to compare forecasts of transit ridership because of the significantly different scales of analysis: The regional systems planning study was conducted with large zones and the IGTDS study performed a more detailed evaluation down to the level of individual transit lines and stops. At an aggregate level, however, the two models were found to predict very similar levels of total transit ridership to the Bellevue CBD.

The regional macrocomputer model was not designed for use in a small subarea of a metropolitan region nor does PSCOG have any plans to use the model to specifically design transportation services for Bellevue or any other suburban community. A much greater level of data preparation is required for running the macrocomputer model, as computer costs are higher and turnaround time is much greater than for IGTDS. IGTDS, on the other hand, is particularly well-suited to designing specialized transportation services for well-defined CBDs and activity centers in suburban communities like Bellevue.

The GM TSC project team and members of the local planning group for the demonstration study recommended that UMTA encourage the use of IGTDS as a useful new technique for transportation sketch planning. Recommendations for improvements to IGTDS, which would further enhance its capabilities for transportation sketch planning, are provided in the final report on IGTDS (1), which is also available from the Office of Planning Methods and Support, UMTA, U.S. Department of Transportation.

ACKNOWLEDGMENT

We would like to express our appreciation to members of the GM TSC project team who conducted the IGTDS demonstration study and the transportation planners who participated as members of the local planning group in Bellevue.

In addition, several people provided input data for and information related to the study, specifically William E. Popp of Wm. E. Popp Associates, Bellevue, Washington, consultant; Robert Berg of Parsons, Brinckerhoff, Quade and Douglas (Seattle office); and Gary Norris, Bellevue Public Works Department. Paul C. Atallah and John R. Augustin, both of GM TSC, prepared input data transformations required for IGTDS and digitized the network data base.

Members of Bellevue's local planning group who learned to use IGTDS and participated in the application and evaluation of the IGTDS system included Martha S. Lester of the TRANSPO Group, Bellevue, Washington, transportation planner; Labh Sachdev and Jack Meijsen of PSCOG; and Bob Simpson and Steve Grissom of Metro Transit.

REFERENCES

1. Interactive Graphic Transit Design System (IGTDS) Demonstration Study: Bellevue, Washington, Final Report, Volumes 1 and 2. GM Transportation Systems Center, General Motors Technical Center, Warren, MI, Rept. EP-80023A, June 1980.
2. Bellevue CBD Action Plan. City of Bellevue Planning Department, Bellevue, WA, June 1979.
3. Parsons, Brinckerhoff, Quade and Douglas. Bellevue CBD Transit Alternatives Analysis. Municipality of Metropolitan Seattle (Metro), Seattle, WA, Draft Final Rept. and Technical Appendix, Metro TRANSITion Phase IV Interim Discussion Paper, Oct. 1978.
4. Interactive Graphic Transit Design System (IGTDS): User's Manual for GM TSC Release No. 1. GM Transportation Systems Center, General Motors Technical Center, Warren MI, Rept. EP-79133, Dec. 1979.
5. GM TSC Network Editor: User's Guide for Prototype Version. GM Transportation Systems Center, General Motors Technical Center, Warren, MI, Rept. EP-78144, Aug. 1978.
6. Zone to Node Data Conversion System (ZONOCO): User's Manual for the IGTDS Version. GM Transportation Systems Center, General Motors Technical Center, Warren, MI, Rept. EP-80001, Jan. 1980.

Publication of this paper sponsored by Committee on Bus Transit Systems.