Volume 7, No. 1 2004

Journal of Public Transportation

Volume 7, No. 1 2004

Journal of Public Transportation

NCTR at CUTR

SUBSCRIPTIONS

Complimentary subscriptions can be obtained by contacting:

Laurel A. Land, *Managing Editor*
Center for Urban Transportation Research (CUTR) • University of South Florida
4202 East Fowler Avenue, CUT100 • Tampa, FL 33620-5375
Phone: 813•974•1446
Fax: 813•974•5168
Email: *land@cutr.usf.edu*
Web: *www.nctr.usf.edu/journal.htm*

SUBMISSION OF MANUSCRIPTS

The Journal of Public Transportation is a quarterly, international journal containing original research and case studies associated with various forms of public transportation and related transportation and policy issues. Topics are approached from a variety of academic disciplines, including economics, engineering, planning, and others, and include policy, methodological, technological, and financial aspects. Emphasis is placed on the identification of innovative solutions to transportation problems.

All articles should be approximately 4,000 words in length (18-20 double-spaced pages). Manuscripts not submitted according to the journal's style will be returned. Submission of the manuscript implies commitment to publish in the journal. Papers previously published or under review by other journals are unacceptable. All articles are subject to peer review. Factors considered in review include validity and significance of information, substantive contribution to the field of public transportation, and clarity and quality of presentation. Copyright is retained by the publisher, and, upon acceptance, contributions will be subject to editorial amendment. Authors will be provided with proofs for approval prior to publication.

All manuscripts must be submitted electronically, double-spaced in Word file format, containing only text and tables. If not created in Word, each table and/or chart must be submitted separately in Excel format. All supporting illustrations and photographs must be submitted separately in an image file format, such as TIF, JPG, AI or EPS, having a minimum 300 dpi. Each chart and table must have a title and each figure must have a caption.

All manuscripts should include sections in the following order, as specified:

Cover Page - title (12 words or less) and complete contact information for all authors
First Page of manuscript - title and abstract (up to 150 words)
Main Body - organized under section headings
References - Chicago Manual of Style, author-date format
Biographical Sketch - of each author

Be sure to include the author's complete contact information, including email address, mailing address, telephone, and fax number. Submit manuscripts to the Managing Editor, as indicated above.

JOURNAL OF

Public Transportation

Volume 7, No. 1, 2004
ISSN 1077-291X

The *Journal of Public Transportation* is published quarterly by

National Center for Transit Research
Center for Urban Transportation Research
University of South Florida • College of Engineering
4202 East Fowler Avenue, CUT100
Tampa, Florida 33620-5375
Phone: 813•974•3120
Fax: 813•974•5168
Email: *land@cutr.usf.edu*
Website: *www.nctr.usf.edu/journal*

Volume 7, No. 1, 2004
ISSN 1077-291X

CONTENTS

Our troubled planet can no longer afford the luxury of pursuits confined to an ivory tower. Scholarship has to prove its worth, not on its own terms, but by service to the nation and the world.
—Oscar Handlin

Estimation of Bus Arrival Times Using APC Data

Jayakrishna Patnaik, Steven Chien, and Athanassios Bladikas
New Jersey Institute of Technology

Abstract

Bus transit operations are influenced by stochastic variations in a number of factors (e.g., traffic congestion, ridership, intersection delays, and weather conditions) that can force buses to deviate from their predetermined schedule and headway, resulting in deterioration of service and the lengthening of passenger waiting times for buses. Providing passengers with accurate bus arrival information through Advanced Traveler Information Systems can assist passengers' decision-making (e.g., postpone departure time from home) and reduce average waiting time. This article develops a set of regression models that estimate arrival times for buses traveling between two points along a route. The data applied for developing the proposed model were collected by Automatic Passenger Counters installed on buses operated by a transit agency in the northeast region of the United States. The results obtained are promising, and indicate that the developed models could be used to estimate bus arrival times under various conditions.

Introduction

Public transportation planners and operators face increasing pressures to stimulate patronage by providing efficient and user-friendly service. Within the context of Intelligent Transportation Systems (ITS), Advanced Public Transportation Systems (APTS) and Advanced Traveler Information Systems (ATIS) are designed to collect, process, and disseminate real-time information to transit users via emerg-

ing navigation and communication technologies (Federal Transit Administration 1998). One of the key elements and requirements of APTS/ATIS is the ability to estimate transit vehicle arrival and/or departure times. With quickly expanding APTS-related technologies (e.g., Global Position Systems [GPS], Automatic Vehicle Location Systems [AVLS] and Automatic Passenger Counting [APC] systems), ATIS could provide timely vehicle arrival and/or departure information to en-route, wayside, and pretrip passengers for managing their journeys (Kalaputapu and Demetsky 1995; Abdelfattah and Khan 1998; Chien and Ding 1999; Dailey, Maclean, Cathey, and Wall 2001; Lin and Padmanabhan 2002).

To estimate vehicle arrival times, dynamic models may be developed using accurate data collected by new technologies (e.g., AVLS and APC). Since bus travel times between stops depend on a number of factors (e.g., geometric conditions, route length, number of intermediate stops and intersections, turning movements, incidents, etc.), stochastic traffic conditions along the route and ridership variation at stops further increase uncertainties. Thus, the goal of this study is the application of quantitative and qualitative data to develop creditable models for estimating reliable bus arrival times.

In this study, bus arrival time estimation models are developed on the basis of data collected by APC units installed in buses. One should be surprised if a new technology works exactly as intended and generates accurate data immediately after its deployment. APC systems should be no exception. Therefore, the purpose of this article is not only to develop models for estimating bus arrival times, but also to explore problems that could be encountered while processing data collected by the APC units.

Literature Review

Bus arrivals at stops in urban networks are difficult to estimate because travel times on links, dwell times at stops, and delays at intersections fluctuate spatially and temporally. The joint impact of these fluctuations may cause schedule and headway deviations as a bus moves farther from the starting terminal, thereby lengthening the average waiting time for transit users and consequently degrading the quality of service. A sound model, which could accurately estimate vehicle arrival times, would be capable of mitigating such impact to a large extent. However, developing such a model while considering the effects of time and space, varying traffic, ridership, and weather conditions is a challenging task.

AVLS, smart pager, and ATIS devices used by transit operators can provide useful information. However, these devices fall short when it comes to estimating the travel times between any two downstream stops and the arrival times at each downstream stop from the point of real-time observation. An arrival time estimation model at every downstream stop can be developed by establishing stop-to-stop travel times as a function of several significant variables (e.g., distance, number of intermediate stops, total intermediate bus halting time, and time of day) to supplement the services offered by ATIS devices (Abdelfattah and Khan 1998).

A variety of prediction models developed in previous studies were reviewed and they can be classified into univariate and multivariate forecasting models (Chien, Ding, and Wei 2002). Univariate forecasting models are designed to predict a dependent variable by describing the intrinsic relationship with its historical data mathematically. The commonly used univariate forecasting models include probabilistic estimation and time series models (Okutani and Stephanedes 1984; Stephanedes, Kwon, and Michalopoulos 1990; Delurgio 1998).

These methods usually have a short time lag while predicting in real-time. The accuracy of time series models highly relies on the similarity between real-time and historical traffic patterns. Variation of the historical average could cause significant inaccuracy in prediction results (Smith and Demesky 1995). Unlike univariate models, multivariate models can predict and explain a dependent variable on the basis of a mathematical function of a number of independent variables. The commonly-used multivariate models are regression models and state-space Kalman filtering models (Okutani and Stephanedes 1984).

Historically, regression models (both linear and nonlinear) have been popular because they are relatively easy to use, well established, comparable with other available procedures, and well suited for parameter estimation problems. Abdelfattah and Khan (1998) developed linear and nonlinear regression models with simulation data to predict bus delays and the simultaneous influence of various factors affecting delay. They obtained relatively promising results by using a microsimulation approach.

In this study, regression models were developed using data collected by APC units installed in buses to estimate vehicle arrival times at all downstream stops. These models are developed using path-based data (e.g., travel time between two stops along the route), and the travel times are defined as a function of ridership and other external independent factors. Nonetheless, regression is not the only pos-

sible estimation approach and other methods, such as artificial neural networks, have been explored (Chien, Ding, and Wei 2002).

Objective and Scope

The primary objective of this study is to develop multivariate linear regression models for estimating bus arrival times at major stops of a route in an urban network. The study examines the methodology for developing bus arrival time estimating models; the processing, analyzing, and refining of collected data; and the behavior and impact of the independent variables. The scope of this study encompasses model development and validation; analysis of variance and covariance and colinearity matrices of dependent and independent variables; and suggestions for future research on APC implementation that can benefit users and operators.

Data Collection

Previous studies (Abdelfattah and Khan 1998; Chien, Ding, and Wei 2002) indicated that bus travel times might be affected by a number of factors such as route length, ridership (which, in turn, depends on population density and major trip generators), the number of stops and intersections, and the geometry of the route. To develop a meaningful model, data collected from the study route should have substantial variability in the aforementioned factors.

In this study, data was collected from APC units installed on buses operated on a 30-mile (48 km) urban bus route by a transit agency in the northeast United States. Various data relating to trip information can be captured and recorded as the bus heads out for a trip until it reaches the final destination. After the bus reaches the garage/terminal, a centralized computer is engaged to transfer the trip data recorded by the APC to the transit agency's data center. Service along the studied route is provided by five different patterns per each direction (e.g., inbound and outbound) over different time periods. Patterns differ in terms of where the route originates/terminates, whether or not the bus visits specific locations, and the time the bus commences the trip at the origin. Because of data availability and sufficiency, only data collected from service patterns A and B were used for developing bus travel time estimation models. There are 105 intended stops in the outbound direction for each pattern. Pattern A crosses 134 intersections (89 of which are signalized) and has 24 right and 23 left turns. Twelve important stops (known as time points) have been chosen for the analysis. These time

points serve significant trip generators and are listed on the timetables distributed by the transit agency.

The study route operates 24 hours a day. Buses operating on different patterns may travel different portions of the route. The 12 time points are at identical physical locations. The scheduled run time for the route ranges from 92 to 119 minutes for the outbound trips and 78 to 113 minutes for the inbound trips. This study was based on data recorded from January through June 2002. The data contained a total of 311 trips (including 162 outbound and 149 inbound trips) and most of the data were collected during weekday operations (including 108 outbound and 96 inbound trips). In general, each trip serves more than 60 intended stops and 100 to 300 passengers. Data collected from outbound weekday trips were used to develop the proposed models for estimating bus arrival times. Table 1 illustrates the type of data collected from the APC system.

Table 1. Variables Description of APC Data

Variable	Description
Direction	Service direction (inbound or outbound)
Open Time	Recorded bus door opening time
Close Time	Recorded bus door closing time
Leg Time	Travel time between a pair of stops
Dwell Time	
On	Number of passengers boarding at a stop
Off	Number of passengers alighting at a stop
Stop Distance	Travel distance between any two consecutive stops
Distance	Cumulative distance from the origin
Pattern ID	A code associated with each pattern of the route
Stop Sequence	A unique number attached to each stop along the route
Transit Day	Date of the service
Week Day	Day of the week
Time of Day	Starting time of the trip

Data Preparation for Model Development

As mentioned previously, arrival times may be influenced by traffic conditions, ridership, number of intermediate stops, and weather condition, which, in turn, may be different depending on time of day, day of the week, and pattern ID. If one is to estimate travel times with regression models, sufficient observations (samples) should be available for developing creditable models to produce meaningful results. For example, if the 108 outbound trips were grouped by different days, time periods, and pattern IDs, the sample size in each group would not be sufficient. Furthermore, although the actual arrival time of a bus at each time point is needed, a bus may skip a stop due to the lack of demand in some time periods. Thus, the size of data in each group is further limited.

An attempt was made to include as many data as possible in the analysis, as will be described subsequently. If a door open time was available at a time point, this was the arrival time used in the analysis for that time point. The distance between each time point and the origin is assumed as fixed with respect to each pattern ID. This data was provided by the transit agency separately. The original data were further refined by generating interstop travel times, actual number of stops a bus made and the total dwell time, and number of alighting and boarding passengers between two consecutive time points where the bus actually halted during every single trip. Based on the departure time at the first time point, trips can be grouped by time period based on their dispatching time, as indicated in Table 2, where the classification and definition of the time periods and their break points were provided by the transit agency.

Table 2. Time Periods Defined by APC Data Provider

Time Period	Symbol	Description
Early Morning	Em	Trips take place between 4:00 AM - 6:59 AM
Morning Peak	Mp	Trips take place between 7:00 AM - 9:29 AM
Late Morning	Lm	Trips take place between 9:30 AM - 11:59 AM
Mid-Day	Md	Trips take place between 12:00 Noon - 12:59 PM
Early Afternoon	Ea	Trips take place between 1:00 PM - 3:29 PM
Afternoon Peak	Ap	Trips take place between 3:30 PM - 5:29 PM
Evening	Ev	Trips take place between 5:30 PM - 7:59 PM
Late Night	Ln	Trips take place after 8:00 PM or later

Buses departing from the first time point during different time periods may experience varying traffic congestion and ridership along the route and therefore deviate from their schedule. For example, during the midday, people are likely to use buses to do shopping or errands; thus, the buses may serve more stops. Also, most schools dismiss in the early afternoon, generating student ridership and school bus traffic, causing traffic congestion. On the other hand, early morning and late night trips are likely to experience the least traffic congestion. These facts signify that time period is a significant factor associated with the estimation of bus travel times.

Whenever one uses a large database, it is desirable to screen the data carefully for erroneous entries and inconsistencies, which can be generated by equipment malfunction, human errors, software bugs, and other causes. Corrections and adjustments were made to the problematic data. When a correction was impossible, erroneous records were excluded from the analysis. Data had to be corrected/eliminated primarily because of the following reasons:

1. The *Leg Time* was reported as zero. In cases where both the door open time at a subsequent stop and close time at the previous stop were available, the difference of those times was used to compute the leg time.
2. The *Stop Distance* was reported as zero. Since distance is fixed between each time point and the origin, such data were replaced by actual time point to time point distance.
3. The *Open Time* was blank. To get this time, the *Leg Time* was added to the *Close Time* of the immediately preceding stop.
4. The *Close Time* was blank. To get this time, the *Dwell Time* was added to the *Open Time* for that stop.
5. The *Stop Sequence* was reported as zero. To identify the *Stop Sequence* (and hence the time point), the cumulative distance traveled up to that stop was computed and compared with the known distance to the time points. If a time point could be identified, the record was kept; otherwise, it was dropped.
6. The *Open Time* at a subsequent stop was earlier than the *Close Time* at a previous stop. These records were dropped.
7. The *Cumulative Distance* from the origin to a particular stop was unusually longer than the average. These records were dropped.

8. Occasionally, the *Stop Distance* would be unusually high. These records were dropped.
9. Occasionally, the bus stops (there is *Dwell Time*), but there are no on or off passengers. These records were retained (particularly since *Dwell Time* is one of the independent variables used).
10. Occasionally, there is no *Dwell Time*, but there are boarding and alighting passengers. The *Dwell Time* was calculated by taking the difference between the *Door Open Time* and *Door Close Time* at that particular stop. If door time data were not available, the record was dropped.
11. *Trip-Status* (START and END) tags would show up somewhere in the middle of the trip. The tags were moved to their appropriate places.

The data were then augmented with weather information (precipitation, visibility, and wind speed) obtained from another source.

Selection of Independent Variables

The independent variables selected to develop path-based travel time estimation models were distance, number of stops, dwell times, boarding and alighting passengers, and weather descriptors. Furthermore, there was the option of generating classes of separate models for each factor (i.e., time of day, day of week, pattern ID) that can affect travel time or include that factor as an independent variable in an overall regression.

The SAS (Version 8.02) package was used to develop a set of regression models. The decision on whether a model was reasonable was based on the signs of the coefficients, values of the R-squares, t-values of the coefficients, correlation factors among the variables, and analysis of the residuals to indicate that the developed linear models would be appropriate.

The analysis of the regression results indicated that weather variables were not among the significant factors for estimating arrival times. This can be attributed to the fact that the weather data were not sufficiently detailed or that during the study period the weather variations were not significant enough to have an impact on arrival times. A general linear model was developed for the difference of actual and scheduled journey time with independent variables (e.g., week day, time period, weather) that were categorically chosen as class factors. To identify the statistical insignificance of these variables, Tukey's test (Montgomery 2001) was conducted. The p-value generated for day of the week was 0.4712, suggesting

that trips taking place on different days of the week do not contribute any measurable difference to the travel time. These results also suggest that day of the week is not significant as an independent variable. In addition, regression models generated separately for each day of the week did not exhibit differences that could be attributed to the day. On the contrary, time of day appeared to affect travel time significantly, having very small p-values (< 0.0001).

Demand-related variables (number of stops, dwell times, boarding and alighting passengers between time points) should definitely have an impact on bus travel times. However, it is obvious that they might be highly correlated to each other. For example, regressions were tested with different combinations of data, such as (1) stops, dwell time, boarding passengers, and alighting passengers; (2) stops, dwell times, and the sum of boarding and alighting passengers (i.e. number of passengers served); and (3) stops and boarding passengers. The correlation factor between number of passengers served and total dwell time within any pair of time points was as high as 0.93. Therefore, only one of these two variables was selected. Bus dwell time was chosen, as opposed to the total number of passengers served, because the count of total passengers served could be deceptive in the sense that two distinct activities (i.e., passengers boarding and alighting the bus) could be taking place simultaneously. Even so, dwell times at previous stops directly impact vehicle arrival times in further downstream stops. The regression that included all variables produced R-square values that are smaller than the ones of the model presented here. Besides distance and time period, number of stops and duration of dwell times were the most appropriate and significant independent variables with p-values of 0.15 or less. The proposed model has some independent variables that are highly correlated (e.g., dwell time and number of stops, distance and stops) and some of their coefficients do not have a very high statistical significance.

After reviewing the data, it was found that bus travel times exceed scheduled times during certain periods. The difference is greater if a bus was dispatched during the time periods of late morning, mid-day and early afternoon than during morning peak and afternoon peak. This may be due to the prohibition of street parking in the peak hours and the presence of construction activities during nonpeak periods. Due to these differences, variables associated with the time of day the trip took place (as described in Table 2), are treated as independent variables. Additionally, the pattern IDs show a unique subset of stops along the route. An analysis of numerous regression results indicated that it was best to develop separate models for each pattern.

Given the above, the general model used to estimate bus travel (and therefore arrival) time for pattern "p" from time point "i" to all downstream time points "j" is formulated as

$$T_{i,p}=b_0+b_1d_{i,j}+b_2t_{i,j}+b_3s_{i,j}+b_4Em+b_5M_p+b_6L_m+b_7M_d+b_8E_a+b_9A_p+b_{10}E_v+b_{11}L_n$$
$$\text{for } \forall\ i \text{ and } i+1 \le j \le 12$$

where:

$T_{i,p}$	is the estimated travel time from time point "i" to all downstream time points for bus pattern "p" (e.g., A, or B) (minutes)
$d_{i,j}$	is the distance between TP_i and TP_j (miles)
$t_{i,j}$	is the average of cumulative dwell time between TP_i and TP_j (minutes)
$s_{i,j}$	is the average of cumulative number of stops between TP_i and TP_j
E_m	is a binary variable that indicates Early Morning
M_p	is a binary variable that indicates Morning Peak
L_m	is a binary variable that indicates Late Morning
M_d	is a binary variable that indicates Mid-Day
E_a	is a binary variable that indicates Early Afternoon
A_p	is a binary variable that indicates Afternoon Peak
E_v	is a binary variable that indicates Evening
L_n	is a binary variable that indicates Late Night
b_0	is the intercept of the travel time estimation model
b_k	are the parameters for variables $d_{i,j}$, $t_{i,j}$, $s_{i,j}$, E_m, M_p, L_m, M_d, E_a, A_p, E_v and L_n, respectively, where k varies from 1 to 11
i	is the index of origin time points
j	is the index of destination time points

Given a pattern ID, origin time point, and time period, the proposed model can estimate the required time to travel the path to every downstream time point and thereby the vehicle arrival time at that time point. All time periods are assigned a

value of 1 if present (if the trip started in that time period), and 0 otherwise. Regressions were run both with and without intercepts. All variable notations and their associated coefficients are the same for both types of regression models. The only difference is that models having no intercepts would have their b_0 values equal to zero.

Analysis of Results

For each of the two patterns used here, it is possible to develop one path-based model to estimate bus travel time for all downstream time points from a given starting time point. It is not possible to present the results of all models in this article. A sample of path-based models with intercepts for all possible origins of Pattern A is shown in Table 3. Conversely, Table 4 presents all path-based models of Table 3 but with no intercepts. Using the same methodology, all potential models for Pattern B were also developed but are not shown here.

The models were developed using the stepwise regression method. Variables having significance level values more than 0.15 were considered to be insignificant and, hence, were not included in the model. As shown in Tables 3 and 4, the R-square values obtained ranged from 0.96 to 0.99 for all models that have intercepts and 0.99 for those that do not have any intercepts. The estimation of arrival times is largely dependent upon the travel distance between a pair of time points. This distance was provided by the transit agency and is constant for all trips. Consequently, this results in high R-square values for all models developed. The overall p-values obtained for all models of both Patterns A and B is <0.0001. The parameter estimates of morning peak, evening, and late nighttime periods are zero. This suggests that M_p, E_v, and L_n do not enter in any of the models.

Since the methodologies used to develop all models are the same, their final results are similar. Therefore, it is redundant to discuss each one of them individually and in detail. The plot of actual versus estimated bus travel time to all downstream stops for Model I from Table 4 is presented in Figure 1 and the scatterplot of the residuals in Figure 2. Both figures substantiate visually the linear relationship of the dependent variable with all independent variables that are used in the models. In addition, normal probability plots of the residuals (not shown here) indicate that the normality assumption for the distribution of residuals is not violated.

The overall model statistics for the same model (I from Table 4) are shown in the table. The stepwise selection of variables for this model was in the order of $d_{i,j}$, $s_{i,j}$, E_a, A_p, E_m, and $t_{i,j}$. Each of these independent variables as they entered into the

Figure 1. Estimated Versus Actual Travel Time (minutes)

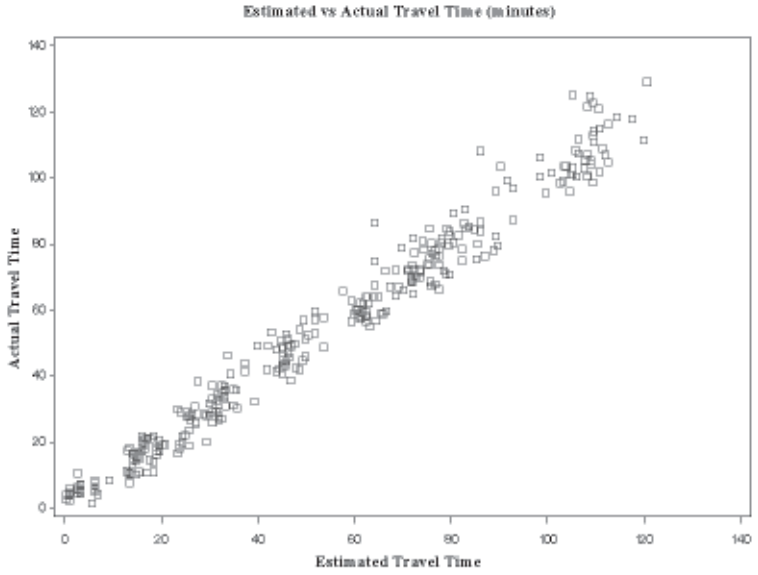
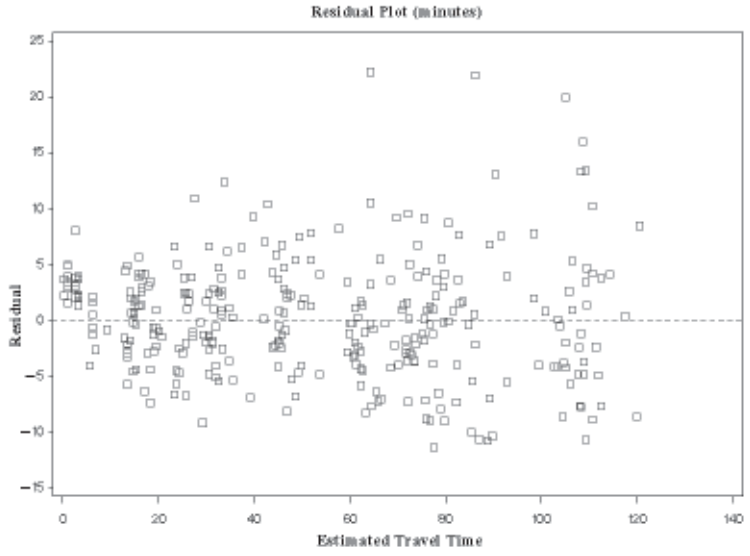


Figure 2. Residual Plot of Estimated Travel Time (minutes)



model retained their final p-values of <0.0001, 0.0914, <0.0001, 0.0084, 0.0002, and <0.0001, respectively. The summary statistics for each model are presented in Tables 3 and 4.

Table 3. Statistics of Bus Travel Time Estimation Models With Intercepts

Models	I	II	III	IV	V	VI	VII	VIII	IX
i	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9
p	A	A	A	A	A	A	A	A	A
b0	1.11	0.02	1.21	0.08	4.11	0.55	1.77	1.62	-2.86
b1	2.61	3.18	3.03	2.73	2.73	2.82	2.92	3.02	2.43
b2	0.21	0.95	0.41	0	0	0.46	0.47	0.77	5.64
b3	0.57	0	0.27	0.55	0.53	0.29	0.16	0	0
b4	-2.57	-4.39	-3.98	-3.99	-6.44	-2.78	-2.16	-2.74	0
b5	0	0	0	0	0	0	0	0	0
b6	0	-3.21	-2.37	0	-3.63	0	0	-0.77	0
b7	0	0	0	2.34	0	0	-1.13	-1.53	0
b8	2.59	0	0	5.61	0	2.61	0	0	1.98
b9	6.34	4.25	0	0	0	0	0	0	0
b10	0	0	0	0	0	0	0	0	0
b11	0	0	0	0	0	0	0	0	0
R-Sq	0.98	0.97	0.96	0.98	0.96	0.98	0.98	0.99	0.97
F value	2016.87	1372.38	1309.29	1230.32	810.44	1433.45	1123.77	858.41	307.6
RMSE	5.28	5.23	5.14	3.2	4.59	2.47	1.86	1.6	3.11
N	312	210	254	107	154	160	99	69	29
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 4. Statistics of Bus Travel Time Estimation Models Without Intercepts

Models	I	II	III	IV	V	VI	VII	VIII	IX
i	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9
p	A	A	A	A	A	A	A	A	A
b1	2.71	3.18	3.12	2.73	2.71	2.82	2.9	3.02	2.15
b2	0.24	0.95	0.5	0	0	0.49	0.4	0.76	5.6
b3	0.53	0	0.21	0.55	0.56	0.3	0.21	0	0
b4	-2	-4.39	-3.38	-3.94	-2.68	-2.59	-0.77	-1.11	0
b5	0	0	0	0	0	0	0	0	0
b6	0	-3.21	-1.91	0	0	0	1.31	0.86	0
b7	0	0	0	2.4	3.93	0	0	0	0
b8	3.01	0	0	5.67	3.67	2.54	1.67	1.53	0
b9	6.82	4.25	0	0	0	0	0	3.22	0
b10	0	0	0	0	0	0	0	0	0
b11	0	0	0	0	0	0	0	0	0
R-Sq	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
F value	7193.81	6142.3	6021.84	5177.83	3433.28	8314.9	4017.38	2921.03	1233.66
RMSE	5.3	5.21	5.14	3.19	4.6	2.47	1.86	1.59	3.24
N	313	211	255	108	155	161	100	70	30
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

As shown in Table 3, the travel time estimation model IX has a negative intercept of -2.86. However, this does not mean that the model will generate negative travel times. The models have positive values for the parameter estimates of variables that are reasonably significant contributors of the travel time estimation (e.g., $d_{i,j}$, $t_{i,j}$, and $s_{i,j}$), and these variables are always positive. This suggests that the estimated negative value of an intercept tends to act as an adjustor to the accuracy of a travel time estimate. Therefore, under no circumstance will a travel time estimation model generate negative travel times. Negative signs of parameter estimates for their associated indicator variables representing a specific time period can be explained similarly.

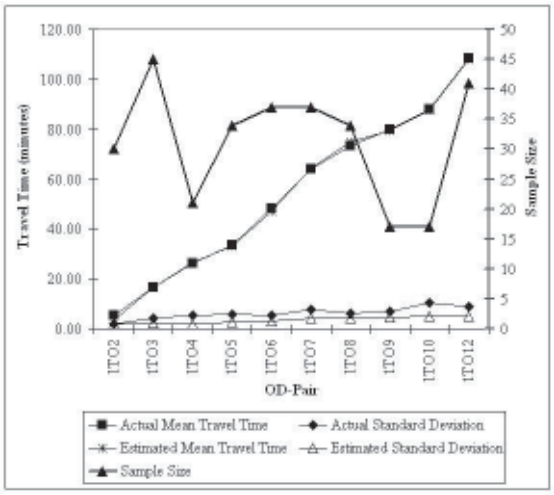
All models have a negative sign for some parameter estimate (e.g., b_4 value for variable E_m). This makes sense, because during early morning time periods, outbound buses are likely to experience less traffic congestion and, hence, shorter travel times. On the other hand, all models contained in both Tables 3 and 4 always have positive signs for parameter estimates (e.g., b_8 and b_9 for variables E_a and A_p). These results may be due to the fact that buses operating during the time periods of early afternoon and afternoon peak are expected to experience more traffic congestion and are more likely to be stopped at the signalized intersections, causing longer travel times. However, another interesting observation that can be made from these models is that some parameter estimates (e.g., b_5 for variable M_p) have either zero or negative values. This suggests that the morning peak time period either has a small or no contribution to the travel time estimation. This may be due to the fact that routes of Patterns A and B possibly experience less traffic congestion during the morning peak time period. This may be because buses are facing favorable signal timings and prohibition of street parking along the route during this time period.

A comparison of F-values of both sets of models shows that the ones that have intercepts generate smaller values than the ones that do not have any intercepts. This is consistent with the corresponding R-square values, which are a little smaller for models that have intercepts.

Data splitting or a cross validation approach (Snee 1977) is chosen for developing and then validating the models of Patterns A and B. These travel time estimation models were developed with 80 percent of the total available data for a sample size (N). The remaining 20 percent of the data were used to validate the model. Observations are chosen randomly for developing and validating the models.

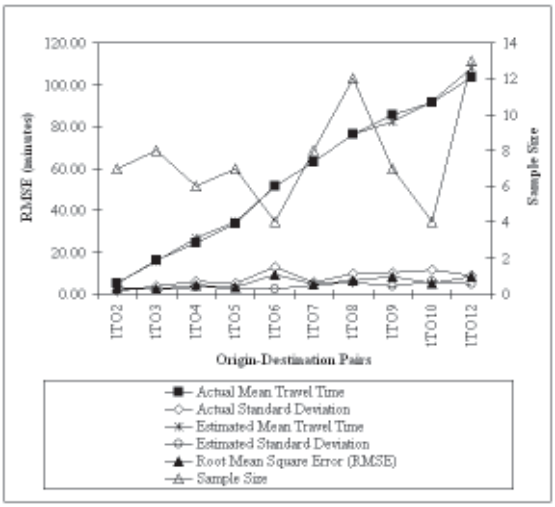
Figure 3 presents statistical descriptions of the model developed using the randomly-selected 80 percent of the total sample data available. On the other hand, Figure 4 illustrates how the 20 percent data best fits and validates the model developed by using the other 80 percent of data. The presented statistics are for the previously discussed Model I of Table 4. Means of actual versus estimated travel times for each OD pair were plotted to determine if there are any significant differences. Both Figures 3 and 4 point out that actual and estimated travel times are reasonably close to each other since the observations for model development (sample size N is equal to 313) and for model validation (sample size of 76) were randomly picked.

Figure 3. Model Development Statistics (80% of data)



As shown in Figure 4, for the OD pair TP_1-TP_6, the actual standard deviation is the highest, having a value of 12.88 minutes, while the corresponding mean actual travel time is 51.48 minutes. This may be attributed to the fact that the available sample size that was randomly chosen for this OD pair is very small and equal to 4. This explains why the root mean squared error for this OD pair is the highest (9.10) in spite of the fact that its estimated mean travel time is very close to the

Figure 4. Model Validation Statistics (20% of data)



actual mean travel time. The estimated standard deviation for this OD pair is 2.45 minutes.

The OD pair TP_1-TP_{10} has the minimum sample size of 4, as did TP_1-TP_6. But, its actual standard deviation is 11.53 minutes while its actual mean travel time is 91.49. Proportionally (as a percent of mean) this standard deviation is approximately half that of OD pair TP_1-TP_6. This can explain the smaller mean squared error value for TP_1-TP_{10} OD pair in comparison with the TP_1-TP_6 OD pair.

The OD pair TP_1-TP_{12} RMSE is 8.24 (the third highest in the sequence), in spite of its highest sample size of 13, and can be attributed to the fact that the estimated mean travel time is essentially about 5.36 minutes higher than the actual mean travel time. The estimated standard deviations of all OD pairs vary from 1.73 to 5.93 minutes, depending upon how close the downstream stops are and also what their overall sample size is. Sample size varies from 4 through 13 for all OD pairs as described.

Having mentioned all these facts, it can be concluded that the results of model validation using the 20 percent data are quite promising, suggesting that the model can be appropriately used to estimate travel times with a new set of data later. As indicated in the table and figures, the results generated by the models are

very reasonable. The plots of the estimated versus actual values indicate linear relationships. The coefficients have the anticipated signs and the adjusted R-squares are almost 0.99 for both Patterns A and B. Some models are better than others in terms of their R-squares and the statistical significance of their co-coefficients. In all cases, the mean travel time increases as we estimate travel times to farther downstream stops and so are their standard deviations. This makes sense, due to the fact that a bus is likely to encounter more and more stochastic traffic situations, causing delays as it moves farther away from the originating terminal.

On the basis of all developed models, a database can be generated that would contain parameter estimates and values of the dependent variables for the purpose of estimating the travel time at downstream stops. The transit operator would be required to input pattern ID, stop ID, and time period. Based on these inputs, the travel time estimation engine will select the appropriate model from the list of models developed to estimate the arrival times at each downstream stop. This portion of the research will commence after all models are finalized.

Conclusions and Future Research

One of the major stochastic characteristics in transit operations is that vehicle arrivals tend to deviate from the posted schedule. Poor schedule or headway adherence is undesirable for both users and operators, since it increases passenger wait/transfer times, discourages passengers from using the transit system, and degrades operating efficiency and productivity. This study developed regression models to predict bus arrival information on the basis of distance traveled, demand characteristics, and time of day. Although the available data were limited, some interpolations had to be made, and some data had to be corrected, there is no absolute certainty that some erroneous figures were not included. The initial results presented here appear to be reasonable and promising.

The methodology used for developing the travel time estimation model with APC data can be used for adjusting or planning timetables for existing or new transit routes, respectively. The developed model can be applied with ATIS to calculate and broadcast bus arrival time information at downstream stops to transit users. If a dynamic algorithm (e.g., Kalman filter) can be developed and integrated with the developed model, the accuracy of predicted bus arrival times can be greatly improved.

Another obvious comment that can be made as a result of this exercise is that one might not use indiscriminately data that are generated automatically, particularly if the system that generates them is complex and new. This is not surprising. It almost always happens, and the data quality and consistency improves rapidly with time. A good and well-known transit practitioners' example of this is the Section 15 database, which had substantial problems with the quality of its data during the first year of its release (Bladikas and Papadimitriou 1985). Therefore, the statement made here about the data quality is not meant as a criticism but as an illustration of the difficulties encountered when using new and large databases.

The data used for this study were relatively limited. The results and the models' predictive ability will certainly improve in the future when data of greater quantity and quality will be available. In the future, it may be possible to generate models for trips grouped by day, time of day, and pattern ID. Furthermore, as the ITS system deployment continues, the models could be expanded to include traffic condition variables, such as congestion and incidents, that can be automatically generated by these systems.

References

Abdelfattah, A. M., and A. M. Khan. 1998. Models for predicting bus delays. *Transportation Research Record 1623*, 8 15.

Bladikas, A., and C. Papadimitriou. 1985. A guided tour through the Section 15 maze. *Transportation Research Record 1013*, 20 27.

Chien, S., and Y. Ding. 1999. A dynamic headway control strategy for transit operations. *Conference Proceedings* (CD-ROM), 6th World Congress on ITS, Toronto, ITS Canada.

Chien, S., Y. Ding, and C. Wei. 2002. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering* 128 (5).

Dailey, D., S. Maclean, F. Cathey, F., and Z. Wall. 2001. Transit vehicle arrival prediction: Algorithm and large-scale implementation. *Transportation Research Record 1771*, 46 51.

DeLurgio, S. A. 1998. *Forecasting principles and applications.* New York: McGraw-Hill.

Federal Transit Administration. 1998. *Advanced Public Transportation Systems: The State of the Art, Update'98.* U.S. Department of Transportation, Washington DC.

Kalaputapu, R., and M. J. Demetsky. 1995. Application of Artificial Neural Networks and Automatic Vehicle Location Data for Bus Transit Schedule Behavior Modeling. *Transportation Research Record 1497*, 44 52.

Lin, W-H., and V. Padmanabhan. 2002. Simple procedure for creating digitized bus route information for Intelligent Transportation System applications. *Transportation Research Record 1791*, 78 84.

Montgomery, D. C. 2001. *Design and analysis of experiments.* 5th edition. John Wiley and Sons Inc.

Okutani, I. and Y. J. Stephanedes. 1984. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research* 18B(1), 1 11.

Smith, B. L., and M. J. Demesky. 1995. Short-term traffic flow prediction: Neural network approach. *Transportation Research Record 1453*, 98 104.

Snee, Ronald. D. 1977. Validation of regression models: Methods and examples. *Technometrics* 19 (4), 15 428.

Stephanedes, Y. J., E. Kwon, and P. Michalopoulos. 1990. On-line diversion prediction for dynamic control and vehicle guidance in freeway corridors. *Transportation Research Record 1287*, 11 19.

About the Authors

Jayakrishna Patnaik (*jp42@njit.edu*) is a research assistant and masters candidate in the Department of Industrial and Manufacturing Engineering at New Jersey Institute of Technology (NJIT). He earned his bachelor's degree in mechanical engineering from Orissa University of Agriculture and Technology, India. He is a member of IIE and Alpha Pi Mu, Industrial Engineering Honors Society.

Steven I-Jy Chien (*chien@adm.njit.edu*) is an associate professor of civil engineering and has a joint appointment with the Interdisciplinary Program in Transportation at NJIT. He earned his Ph.D. degree from the University of Maryland at College Park. He is a member of ASCE, ITE, and TRB.

Athanassios Bladikas (*bladikas@adm.njit.edu*) is an associate professor of industrial and manufacturing engineering, director of the interdisciplinary program in transportation and chairperson of the Industrial and Manufacturing Engineering Department at NJIT. He earned his Ph.D. from Polytechnic University of New York and an MBA from Columbia University. He is a member of ITE, IIE, and ASEE.

Determinants of Bus Dwell Time

Kenneth J. Dueker, Thomas J. Kimpel, James G. Strathman
Portland State University
Steve Callas, TriMet

Abstract

Bus dwell time data collection typically involves labor-intensive ride checks. This paper reports an analysis of bus dwell times that use archived automatic vehicle location (AVL)/automatic passenger counter (APC) data reported at the level of individual bus stops. The archived data provide a large number of observations that serve to better understand the determinants of dwells, including analysis of rare events, such as lift operations. The analysis of bus dwell times at bus stops is applicable to TriMet, the transit provider for the Portland metropolitan area, and transit agencies in general. The determinants of dwell time include passenger activity, lift operations, and other effects, such as low floor bus, time of day, and route type.

Introduction

Bus dwell time data collection typically involves labor-intensive ride checks. This paper reports an analysis of bus dwell times that use archived automatic vehicle location (AVL)/automatic passenger counter (APC) data reported at the level of individual bus stops. The archived AVL/APC data provides a rich set of dwell time observations to better understand the determinants of dwells. In addition, the large quantity of data allows analysis of rare events, such as lift operations. The analysis of bus dwell times at bus stops was originally used to estimate delay associated with bus lift use operations for passengers with disabilities in the Tri-County Metropolitan Transportation District of Oregon (TriMet), the transit provider for

the Portland metropolitan area (Dueker, et al. 2001). In addition, the analysis yielded useful information about dwell times that has applicability to transit agencies in general.

The estimated models provide a system-wide baseline. Stop-level, route-level, operator-specific, and passenger boarding-level analyses can follow. This paper includes examples of applying the model results to simulate dwell times for different times of day, route types, and various levels of passenger boardings and alightings. The effects of fare payment method and bicycle rack usage on dwell times was unable to be incorporated, but suggest how future research could extend the model.

Prior Work

Literature on bus dwell times is sparse, due to the cost and time required for manual data collection. Consequently, most prior analyses tend to be route-specific, focus on analyzing various issues causing bus delay, and are based on small samples. Previous studies on dwell time have used ordinary least squares (OLS) regression to relate dwell time to boardings and alightings, with separate equations estimated for different operating characteristics likely to affect dwell time. Kraft and Bergen (1974) found that passenger service time requirements for AM and PM peaks are similar, midday requirements are greater than those in peak periods, boarding times exceed alighting times, and rear door and front door alighting times are the same. They also found that dwell time is equal to 2 seconds plus 4.5 seconds per boarding passenger for cash and change fare structures, and 1.5 seconds plus 1.9 seconds for exact fare.

Levinson's (1983) landmark study of transit travel time performance reported that dwell time is equal to 5 seconds plus 2.75 seconds per boarding or alighting passenger. Guenthner and Sinha (1983) found a 10-20 second penalty for each stop plus a 3-5 second penalty for each passenger boarding or alighting. However, dwell time models based on small samples have low explanatory power, even when controlling for factors such as lift activity, fare structure, and number of doors. Guenthner and Hamet (1988) looked at the relationship between dwell time and fare structure, controlling for the amount of passenger activity. Lin and Wilson (1992) reviewed prior work and formulated a model of dwells as a function of boardings, alightings, and interference with standees, which was then applied to light rail transit dwells. Bertini and El-Geneidy (2004) modeled dwell time for a single inbound radial route in the morning peak period in their analysis of trip

level running time. They incorporated the results of the dwell time analysis directly into the trip time model by estimating parameters for number of dwells and number of boarding and alighting passengers.

Data Issues

Dwell time is defined as "the time in seconds that a transit vehicle is stopped for the purpose of serving passengers. It includes the total passenger service time plus the time needed to open and close doors" (HCM 1985).

In the past, dwell time data collection consisted of placing observers at highly utilized bus stops to measure passenger service times, and by ride checks or on-board observers for dwells at bus stops along routes. The ride check procedure as prescribed in the Transit Capacity and Quality of Service Manual consists of the following steps to collect field data for estimating passenger service times:

1. From a position on the transit vehicle, record the stop number or name at each stop.
2. Record the time that the vehicle comes to a complete stop.
3. Record the time that the doors have fully opened.
4. Count and record the number of passengers alighting and the number of passengers boarding. (The data collection form calls for front and rear door specific counts).
5. Record the time that the major passenger flows end.
6. When passenger flows stop, count the number of passengers remaining on board. (Note: If the seating capacity of the transit vehicle is known, the number of passengers on board may be estimated by counting the number of vacant seats or the number of standees).
7. Record time when doors have fully closed.
8. Record time when vehicle starts to move. (Note: Waits at timepoints or at signalized intersections where dwell is extended for cycle should be noted but not included in the dwell time. Delays at bus stops when a driver is responding to a passenger information request are everyday events and should be included in the calculation of dwell time. Time lost dealing with fare disputes, lost property or other events should not be included.)
9. Note any special circumstances. In particular, any wheelchair movement times should be noted. Whether this is included in the mean dwell time

> depends on the system. Dwell times due to infrequent wheelchair movements are often not built into the schedule but rely on the recovery time allowance at the end of each run. The observer must use judgment in certain cases. At nearside stops before signalized intersections the driver may wait with doors open as a courtesy to any late-arriving passengers. The doors will be closed prior to a green light. This additional waiting time should not be counted as dwell time but as intersection delay time. (TCRP 1999)

Automating the collection of dwell time data through the employment of AVL and APC technologies compromises the procedures outlined above. The dwell time is measured as specified, but the time the bus stops and starts is not recorded, nor is the starting and stopping of passenger flows. Our analysis deleted dwells of over 180 seconds (3 minutes). This censoring was done to purge the analysis of dwells that are abnormal. Also, TriMet's Automated Passenger Counters (APC) record total boardings and alightings rather than door-specific counts. Finally, there is no guarantee that operators will behave similarly in closing the doors while awaiting for traffic to clear or traffic signals to change. These compromises to the conventional measurement of dwell time are offset by the ability to collect data on large numbers of dwells, with any "special circumstances" included in the error term of OLS regression models.

Automating Collection of Dwell Time Data

Uses of Archived AVL/APC Data to Improve Transit Performance and Management (Furth, et al. forthcoming), identifies the bus stop as the appropriate spatial unit for data aggregation and integration. This integration of scheduled and actual arrival time at the level of the individual stop is crucial for research on bus operations and control strategies. Integrating data at the bus stop level supports real time applications, such as automated stop annunciation and next-stop arrival time information. Importantly, if bus stop data are archived, operations performance and monitoring analysis can also be supported (Furth, et al. forthcoming).

TriMet has automated the collection and recording of bus dwell time and passenger activity at the bus stop level, and archives the data consistent with the TCRP recommendations. TriMet operates 97 bus routes, 38 miles of light rail transit, and 5 miles of streetcar service within the tri-county Portland metropolitan region. TriMet's bus lines carry approximately 200,000 trips per day, serving a total popu-

lation of 1.3 million persons within an area of 1,530 square kilometers (590 square miles).

TriMet implemented an automated Bus Dispatch System (BDS) in 1997 as a part of an overall operation and monitoring control system upgrade.

The main components of the BDS include:

1. AVL based upon differential global positioning system (GPS) technology, supplemented by dead reckoning sensors
2. Voice and data communication system using radio and cellular digital packet data (CDPD) networks
3. On-board computer and control head displaying schedule adherence information to operators and showing dispatchers detection and reporting of schedule and route adherence
4. APCs on front and rear doors of 70% of vehicles in the bus fleet
5. Computer-aided dispatch (CAD) center

The BDS reports detail operating information in real time by polling bus location every 90 seconds, which facilitates a variety of control actions by dispatchers and field supervisors. In addition, the BDS collects detailed stop-level data that are downloaded from the bus at the end of each day for post-processing. The archived data provide the agency with a permanent record of bus operations for each bus in the system at every stop on a daily basis. These data include the actual stop time and the scheduled time, dwell time, and the number of boarding and alighting passengers. The BDS also logs time-at-location data for every stop in the system, whether or not the bus stops to serve passengers. This archived data forms a rich resource for planning and operational analysis as well as research.

The GPS-equipped buses calculate their position every second, with spatial accuracy of plus or minus 10 meters or better. Successive positions are weeded and corrected by odometer input. When the bus is within 30 meters of the known location of the next bus stop (which is stored on a data card along with the schedule), an arrival time is recorded. When the bus is no longer within 30 meters of the known bus stop location, a departure time is recorded. If the door opens to serve passengers, a dwell is recorded and the arrival time is overwritten by the time when the door opens. Dwell time (in seconds) is recorded as the total time that the door remains open.

When passenger activity occurs, the APCs count the number of boardings and alightings. The APCs are installed at both front and rear doors using infrared beams to detect passenger movements. The APCs are only activated if the door opens. The use of a lift for assisting passengers with disabilities is also recorded. TriMet has used on-board cameras to validate APC counts (Kimpel, et al. 2003). The validation procedures could be extended to include dwell time and the timing of passenger flows, and perhaps even fare payment if the video clips are not too choppy.

The archived AVL/APC data have been used in various studies of operations control and service reliability (Strathman et al. 1999; Strathman et al. 2000; Strathman et al. 2001a; Strathman et al. 2001b), for route-level passenger demand modeling (Kimpel 2001), for models of trip and dwell time (Bertini and El-Geneidy 2004), and for evaluating schedule efficiency and operator performance (Strathman, et al. 2002).

Dwell Time Data

The data are from a two-week time period in September 2001 for all of TriMet's regular service bus routes. For this analysis, dwell time (DWELL) is the duration in seconds the front door is open at a bus stop where passenger activity occurs. The data were purged of observations associated with the beginning and ending points of routes, layover points, and unusually long dwell time (greater than 180 seconds).[1] Observations with vehicle passenger loads (LOAD) of over 70 persons were also excluded, indicating the automatic passenger counter data were suspect. Two weeks of data generated over 350,000 dwell observations. Even though lift operations (LIFT) occur in less than one percent (0.7 %) of dwells, the number of lift operations is large enough for a robust estimation of separate model (N = 2,347).

Table 1 presents descriptive statistics for variables used in the full-sample dwell time model. The mean dwell time is 12.29 seconds, with a standard deviation of 13.47 seconds. On average, there were 1.22 boardings and 1.28 alightings per dwell. Also, 61% of the dwells involved low floor buses. Dwells by time of day (TOD) are 15% in morning peak period (6-9 AM) (TOD1), 41% in midday (9 AM -3 PM) (TOD2), 17% in afternoon peak period (3-6 PM) (TOD3), 21% in evening (6-10 PM) (TOD4), and 7% in late night and early morning (10 PM- 6 AM) (TOD5). The mix of dwells by route type is 71% for radial, 4% feeder, and 25% cross-town. Also, the average dwell occurs 2.36 minutes behind schedule (ONTIME).

Table 1. Bus Dwell Time Descriptive Statistics

Name	Mean	St. Dev.	Var.	Min.	Max.
DWELL	12.29	13.47	181.42	2.00	180.00
ONS	1.22	1.99	3.94	0.00	44.00
ONS2	5.43	25.79	664.92	0.00	1936.00
OFFS	1.28	1.90	3.63	0.00	47.00
OFFS2	5.26	25.22	636.04	0.00	2209.00
ONTIME	2.36	3.56	12.70	29.66	57.50
LIFT	0.007	0.081	0.007	0	1
LOW	0.61	0.49	0.24	0	1
TOD1	0.15	0.36	0.13	0	1
TOD2	0.41	0.49	0.24	0	1
TOD3	0.17	0.37	0.14	0	1
TOD4	0.21	0.40	0.16	0	1
TOD5	0.07	0.25	0.06	0	1
RAD	0.71	0.45	0.21	0	1
FEED	0.04	0.19	0.04	0	1
XTOWN	0.25	0.43	0.19	0	1
FRICTION	3.19	4.41	19.46	0	73

The analysis includes information derived from three separate but related samples: (1) a full sample consisting of all observations; (2) a lift operation-only sub-sample; and (3) a without lift operation only sub-sample.

Table 2 shows the effect of a lift operation on mean dwell time. Mean dwell times for the sub-sample without lift operation average 11.84 seconds, while mean dwell times for the sub-sample with lift operation average 80.70 seconds. The coefficient of variation for dwell time with lift operation is 46.4%, and 100.7% for without lift operation. An OLS model for the full sample of both lift and no lift operation had a coefficient of 62.07 for a dummy variable for lift operation (LIFT).[2] A Chow test indicated that a separate model was needed for dwells where lift operations occur.

Table 2. Bus Dwell Time Means

Dwell (seconds)	Mean Time	St. Dev.	N
Sub-sample with lift operation	80.70	37.44	2,347
Sub-sample without lift operation	11.84	11.92	353,552
Both (full sample)	12.29	13.47	355,899

Dwell Time Estimation

Table 3 presents results of the model of the sub-sample without lift operation. Dwell time is explained by boarding passengers (ONS), alighting passengers (OFFS), whether the bus is ahead or behind schedule (ONTIME), if the bus is a low floor bus (LOW), passenger friction (FRICTION),[1] time of day (TOD), and type of route feeder (FEED) and cross-town (XTOWN) as compared to radial (RAD). The estimation results indicate that each boarding passenger adds 3.48 seconds to the base dwell time of 5.14 seconds (CONST) and each alighting passenger adds 1.70 seconds. Square terms of passenger activity are used to account for diminishing marginal effects of additional boarding and alighting passengers on dwell time. Each additional boarding passenger is estimated to take 0.04 seconds less, while each additional alighting passenger takes 0.03 seconds less.[2] The negative coefficient of ONTIME indicates that dwell times tend to be less for late buses than for early buses[3]. The CONST value of 5.14 seconds reflects the basic opening and closing door process.

The other variables have small but significant effects. Time-of-day estimates are referenced to the morning peak period (TOD1). Midday dwells (TOD2) are 1.36 seconds longer than morning peak dwells; afternoon peak dwells (TOD3) are 0.92 seconds longer than morning peak dwells; and evening period dwells (TOD4) are 1.25 seconds longer than morning peak dwells, while late evening and early morning period dwells (TOD5) are not significantly different than morning peak dwells. The morning peak period is the most efficient in terms of serving passengers, perhaps due to regular riders and more directional traffic. Regular riders may tend to board using bus passes[6] and ask fewer questions. More directional traffic would reduce the mix of boardings and alightings at the same stop.

The type of route also affects dwell times. Feeder routes have 0.15 second longer dwells than radials, the reference route type, and cross-town routes have 0.39 second shorter dwells than buses operating on radial routes.

Table 3. Bus Dwell Time Model: Without Lift Operation

Name	Coeff.	Std. Err.	T-Ratio
ONS	3.481	0.015	231.90
ONS2	-0.040	0.001	-37.38
OFFS	1.701	0.015	113.00
OFFS2	-0.031	0.001	-29.11
ONTIME	-0.144	0.005	-30.68
LOW	-0.113	0.034	-3.30
FRICTION	0.069	0.005	12.92
TOD2	1.364	0.049	27.82
TOD3	0.924	0.059	15.77
TOD4	1.248	0.055	22.51
TOD5	0.069	0.076	0.91
FEED	0.145	0.086	1.70
XTOWN	-0.388	0.039	-9.99
CONST.	5.136	0.051	99.96
N	353,552		
ADJ. R2	0.3475		

Lift Operation Effects

The estimated effect of a lift operation on dwell time in a full-sample model is 62.07 seconds. This lift operation effect is examined more closely in a separate model of dwell times involving lift operations only.

Table 4 presents the results of the bus dwell time model for the sub-sample of lift operation-only. The mean dwell time for lift operation-only dwells is 80.70 seconds, and is explained by the same variables as the overall dwell time model, but they differ and are less significant. For example, a low-floor bus reduces the dwell time for lift operations by nearly 5 seconds. But the large CONST value of 68.86

seconds indicates that the majority of time is for the lift operation itself. Boarding activity is estimated to extend dwells at a diminishing marginal rate, while alighting passenger activity does not substantially affect dwell time. However, wheelchairs, walkers, and strollers may confound APCs. There are significant effects by time of day, but they are not easily explained. Lift operations during the morning peak (TOD1) take longer than lift operations at other times.

Table 4. Bus Dwell Time Model: With Lift Operation

Name	Coeff.	Std. Err.	T-Ratio
ONS	10.206	0.488	20.91
ONS2	-0.359	0.029	-12.31
OFFS	0.513	0.396	1.30
OFFS2	-0.022	0.017	-1.33
ONTIME	-0.037	0.176	-0.21
LOW	-4.741	1.388	-3.42
FRICTION	-0.234	0.208	-1.13
TOD2	-4.141	2.554	-1.62
TOD3	-6.271	2.869	-2.19
TOD4	-4.588	2.925	-1.57
TOD5	-14.447	4.542	-3.18
FEED	1.036	3.354	0.31
XTOWN	-1.675	1.519	-1.10
CONST.	68.861	2.706	25.45
N	2,347		
ADJ. R2	0.2848		

An estimate of delay associated with lift operation can be used to modify arrival time estimates provided to transit users at downstream stops. However, we have three choices of delay estimates for lift operation. One is 62.07 seconds, the coefficient on LIFT from the full model. Another is the difference between the mean of all dwell time with lift operations (80.70 seconds) and without lift operations (11.84 seconds). This difference is 68.86 seconds. The third choice is the effect of a lift operation on running time from an earlier study of route running times (Strathman, et al. 2001a). This third choice provides an estimate of the lift effect as 59.80 seconds. This smaller value indicates that before the end of their trip, operators make up some of the time lost due to lift operations.

We recommend the middle estimate of 62.07 seconds (the coefficient on the LIFT dummy variable from the full sample estimation) be selected as the delay estimate at the outset of the lift event and that it be updated with the actual dwell time less the mean dwell time without lift operation as the bus departs that stop. In this manner, next stop bus arrival time estimates could be refined when impacted by delays associated with lift operations. This would require a message from the bus to the dispatch center at the onset of the lift operation and another at its conclusion.

Low Floor Bus Effect

TriMet was also interested in the effect of low floor buses on dwells, particularly dwells with lift operations. The dwell time model for the without lift operation sub-sample yields an estimated effect of a low-floor bus of -0.11 seconds (-0.93%) per dwell. A typical TriMet route has 60 bus stops. On an average bus trip, buses actually stop at 60% of them. Thus, the 0.11 second reduction per dwell for a low floor bus translates into a 3.96 second savings in total running time per trip.

The low floor bus effect is examined in a model of dwell times involving lift operations only. The mean dwell time for stops where the lift is operated is 80.70 seconds. A low-floor bus reduces dwell time for lift operations by nearly 5 seconds (4.74 or 5.8 %). The impact of low floor buses on dwell time with lift operation is more substantial.

Simulation

Models can be used to simulate dwell times. The coefficients are multiplied by assumed values of the variables that represent operating conditions of interest. Table 5 presents simulated dwell times for various operating conditions. Although the simulation produces precise dwell time estimates, the results should be viewed in relative terms, because of large coefficients of variation in dwell time and the explanatory power of the models are low (adjusted R2 values of 0.35 for without lift operation and 0.28 for with lift operation).

The first condition simulated is a radial route in the AM peak period. Five boardings (ONS) are assumed to load at a stop and there are no alightings (OFFS). The bus is operating two minutes late. This simulation yields a dwell time estimate of 21.15 seconds. The second simulation is of a radial route in the PM peak operating with five OFFS and no ONS. It also has 10 standees. The dwell time estimate is 13.99 seconds. In comparing the two estimates, a greater time associated with boardings as compared to alightings is quantified. The third simulation is for a cross-town

Table 5. Simulation of Bus Dwell Times

Name	Coeff.	Radial AM Inbound		Radial PM Outbound		Cross-Town Midday	
ONS	3.481	5	17.41		0.00	2	6.96
ONS2	-0.040	25	-0.99		0.00	4	-0.16
OFFS	1.701		0.00	5	8.50	2	3.40
OFFS2	-0.031	0	0.00	25	-0.78	4	-0.12
ONTIME	-0.144	2	-0.29	5	-0.72	2.5	-0.36
LOW	-0.113	1	-0.11	1	-0.11		0.00
FRICTION	0.069	0	0.00	10	1.04		0.00
TOD2	1.364		0.00		0.00	1	1.36
TOD3	0.924		0.00	1	0.92		0.00
TOD4	1.248		0.00		0.00		0.00
TOD5	0.069		0.00		0.00		0.00
FEED	0.145		0.00		0.00		0.00
CTOWN	0.145		0.00		0.00	1	0.15
CONST.	5.136	1	5.14	1	5.14	1	5.14
DWELL EST.			21.15		13.99		16.37

Name	Lift Specific Model (w/lift only)			Full Model (w/lift dummy)		
	Coeff.	Midday Feeder Service		Coeff.	Midday Feeder Service	
ONS	10.206	2	20.41	3.551	2	7.10
ONS2	-0.359	4	-1.43	-0.042	4	-0.17
OFFS	0.513	1	0.51	1.703	1	1.70
OFFS2	-0.022	1	-0.02	-0.033	1	-0.03
ONTIME	-0.037	-1	0.04	-0.145	-1	0.14
LOW	-4.741		0.00	-0.143		62.07
LIFT	..	..	..	62.07	1	0.00
FRICTION	-0.234		0.00	0.067		0.00
TOD2	-4.141	1	-4.14	1.352	1	1.35
TOD3	-6.271		0.00	0.902		0.00
TOD4	-4.588		0.00	1.231		0.00
TOD5	-14.447		0.00	-0.013		0.00
FEED	1.036	1	1.04	0.148	1	0.15
CTOWN	-1.675		0.00	-0.390		0.00
CONST.	68.861	1	68.86	5.117	1	5.12
DWELL EST.			85.26			77.43

route in the midday at a stop with two ONS and two OFFS and running 2.5 minutes late. This produces an estimated dwell time of 16.36 seconds.

Table 5 also contains two simulations of a lift operation with two ONS and two OFFS on a feeder line in the midday period with a bus that is running one minute early. This condition is estimated using the lift specific model (dwell estimate of 85.26 seconds) and using coefficients from the full-sample model with a lift dummy variable (77.43 seconds). The difference in estimates is less than the standard deviations of either sample.

For a better understanding of boarding and alighting passenger activity, two additional sub-samples were drawn. Both are for radial routes with no lift operation. One was AM peak period dwells with boardings but no alightings, and the other was PM peak period dwells with alightings but no boardings. This allows the estimation of parameters for boardings and alightings that are not confounded by a mixture of boardings and alightings. Table 6 is the dwell time model for boardings only and Table 7 the model for alightings only. The parameter for boardings is 3.83 seconds per boarding passenger and the parameter for alightings is 1.57 seconds per alighting passenger. Again, both parameters have a significant square term that indicates a declining rate for each additional passenger. Simulations for 1, 2, 5, 10, and 15 boarding passengers are contained in Table 8, and simulations for alighting passengers are contained in Table 9. Both simulations assumed an average lateness (ONTIME) value of 1.56 minutes for the boarding passenger sub-sample and 4.46 minutes for the alighting passenger sub-sample. Both simulations also assumed a low floor bus and a bus load of less than 85 percent of capacity. The simulations calculate dwell time in seconds for various boarding and alighting passengers. For instance, dwell time for five boarding passengers is estimated to be 21.01 seconds (from Table 8) and is estimated to be 12.75 seconds for five alighting passengers (from Table 9). These two simulations illustrate the benefit of working with large amounts of data that is made possible by automated data collection. We were able to select route type, time of day, and dwells with boardings or alightings, but not both.

Comparison of the simulation of five boarding passengers in Tables 5 and 8 yield results that are within a second. Focusing on just the boarding passengers, parameters for the basic stop (CONST) is 4.05 seconds versus 5.14, 19.12 seconds versus 17.41 to board five passengers, and -1.45 versus -0.99 seconds for the diminishing effect of the multiple of five passengers. Similarly, the comparison of five alighting

passengers in Tables 5 and 9 yield results that are within a second when comparing only the alighting times and the constant.

Again, the results of the simulation should be used in comparing scenarios and not be used for precise estimates of dwells.

Table 6. Bus Dwell Time Model: Boardings Only - AM Peak Period

Name	Coeff.	Std. Err.	T-Ratio
ONS	3.825	0.063	61.000
ONS2	-0.058	0.005	-11.340
FRICTION	0.040	0.014	2.845
ONTIME	-0.164	0.020	-8.021
LOW	-0.464	0.103	-4.483
CONST.	4.054	0.126	32.230
N	16,509		
ADJ. R2	0.3819		

Table 7. Bus Dwell Time Model: Alightings Only - PM Peak Period

Name	Coeff.	Std. Err.	T-Ratio
OFFS	1.566	0.057	27.610
OFFS2	-0.016	0.006	-2.703
FRICTION	0.119	0.012	10.150
ONTIME	-0.046	0.008	-5.971
LOW	0.523	0.079	6.651
CONST.	5.001	0.100	49.850
N	18,098		
ADJ. R2	0.1616		

Table 8. Simulation of Bus Dwell Times by Number of Boardings AM Peak Period

		Boardings				
Name	**Coeff.**	**1**	**2**	**5**	**10**	**15**
ONS	3.825	3.82	7.65	19.12	38.25	57.37
ONS2	-0.058	-0.06	-0.23	-1.45	-5.80	-13.04
FRICTION	0.040					
ONTIME	-0.164	-0.26	-0.26	-0.26	-0.26	-0.26
LOW	-0.464	-0.46	-0.46	-0.46	-0.46	-0.46
CONST.	4.054	4.05	4.05	4.05	4.05	4.05
TOTAL DWELL		7.10	10.75	21.01	35.79	47.67

Table 9. Simulation of Bus Dwell Times by Number of Alightings PM Peak Period

		Alightings				
Name	**Coeff.**	**1**	**2**	**5**	**10**	**15**
ONS	1.566	1.57	3.13	7.83	15.66	23.49
ONS2	-0.016	-0.02	-0.06	-0.39	-1.58	-3.55
FRICTION	0.119					
ONTIME	-0.046	-0.21	-0.21	-0.21	-0.21	-0.21
LOW	0.523	0.52	0.52	0.52	0.52	0.52
CONST.	5.001	5.00	5.00	5.00	5.00	5.00
TOTAL DWELL		6.87	8.39	12.75	19.40	25.26

Discussion

The original purpose of this research was to identify the effects of delay that occur at unexpected times, such excess dwell time resulting from bus lift operations. Our research provides an estimate of delay at the time of initiation of the occurrence, which needs to be updated with the actual time of delay at the ending time of the occurrence. This research provides a basis for shifting from predicting transit bus arrival times for customers based on normal operating conditions to one that predicts transit vehicle arrival time when operating conditions are not normal (Dueker, et al. 2001).

An ancillary benefit of this research identified the general determinants of bus dwell time. As expected, passenger activity is an important determinant. In addition, the archived AVL/APC data provided a large sample size that allowed examination of determinants, such as low floor buses, time of day, and route type effects, and allowed estimation of a separate model for dwells with lift operation only.

Automation of dwell time data collection results in a tradeoff of labor-intensive direct observation but small sample data with the large samples of more consistent data. While directly observing door-specific passenger activity, fare payment method, and unproductive door opening time, as called for in the Transit Capacity and Quality of Service Manual, improvements in automated data collection may be able to deal with these issues. For example, integration of farebox and bicycle rack with a BDS data collection system is possible in the future. This would deal with the effect of fare payment method and use of the bicycle rack on dwell time. In addition, validation of dwell time data is needed. TriMet has validated its APC data by viewing on-board video camera data. This procedure could be extended to record the time of passenger activity to the door opening time from the automated data.[7] This would provide evidence to determine a better cutoff value for maximum dwell time. The current value of 180 seconds is too arbitrary; it needs to be replaced with a value that includes most passenger activity and reduces the amount of unneeded door opening time. In addition, the validation procedure could include observation of fare payment method and bicycle rack use prior to integration at the hardware level.

Acknowledgements

The authors gratefully acknowledge support provided by TriMet and the US Department of Transportation (USDOT) University Transportation Centers Program, Region X (TransNow). The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the views or policies of TriMet or the USDOT.

Endnotes

[1] Long dwells are likely to be associated with vehicle holding actions or operator shift changes, and thus should be excluded from the analysis.

[2] Table 5 contains coefficients of the full-sample dwell time model.

[3] A passenger friction factor was constructed to account for passenger activity on buses with standees. It was posited that heavily loaded buses have greater dwell times. A proxy variable was constructed by adding ONS, OFFS, and STANDEES. STANDEES are the number of passengers when LOAD minus 85% of bus capacity is positive. LOAD is an APC calculated number that keeps a running count of ONS and OFFS.

[4] Kraft and Deutschman (1977) did not find any difference in the average service time for each successive passenger to board.

[5] Operators tend to hurry to regain schedule adherence.

[6] The farebox is not integrated with the BDS, so we do not know the proportion of cash paying boarding passengers at the stop level.

[7] Kraft and Deutschman (1977) used photographic studies of passenger movements through bus doors.

References

Bertini, R. L. and A. M. El-Geneidy. 2004. Modeling transit trip time using archived bus dispatch system data. *Journal of Transportation Engineering*, 130 (1): 56-67.

Dueker, K. J., T. J. Kimpel, J. G. Strathman, R. L. Gerhart, K. Turner, and S. Callas. 2001. *Development of a statistical algorithm for real-time prediction of transit vehicle arrival times under adverse conditions*. Portland, OR: Project Report 123, Center for Urban Studies, Portland State University.

Furth, P. G., B. Hemily, T. H. Mueller, J., Strathman. Forthcoming. *Uses of archived AVL-APC data to improve transit performance and management: Review and potential*. TRCP Project H-28. Washington, DC: Transportation Research Board.

Guenthner, R. P. and K. Hamat. 1988. Transit dwell time under complex fare structure. *Transportation Engineering Journal of ASCE*, 114 (3): 367-379.

Guenthner, R. P. and K. C. Sinha. 1983. Modeling bus delays due to passenger boardings and alightings. *Transportation Research Record*, 915: 7-13.

Highway Capacity Manual (HCM), Special Report 209, Transportation Research Board, 1985, p. 12-3.

Kimpel, T. J. 2001. *Time point level analysis for transit service reliability and passenger demand*. Portland, OR: Special Report 36, Center for Urban Studies, Portland State University.

Kimpel, T., J. Strathman, S. Callas, D. Griffin and R. Gerhart. 2003. Automatic Passenger Counter Evaluation: Implications for National Transit Database Reporting. *Transportation Research Record,* 1835: 109-119.

Kraft W. and T. Bergen. 1974. Evaluation of passenger service times for street transit systems. *Transportation Research Record,* 505: 13-20.

Kraft, W. and H. Deutschman. 1977. Bus passenger service-time distribution. *Transportation Research Record,* 625: 37-43.

Levinson, H. S. 1983. Analyzing transit travel time performance. *Transportation Research Record,* 915: 1-6.

Lin T. and N. H. Wilson. 1992. Dwell time relationships for light rail systems. *Transportation Research Record,* 1361: 287-295.

Strathman, J.G., T.J. Kimpel., and K.J. Dueker. 1999. Automated Bus Dispatching, Operations Control and Service Reliability. *Transportation Research Record*, 1666: 28-36.

Strathman, J.G., T.J. Kimpel, K.J. Dueker, R.L. Gerhart, K. Turner, G. Taylor, and D. Griffin. 2000. Service reliability impacts of computer-aided dispatching and automatic vehicle location technology: a case TriMet case study. *Transportation Quarterly*, 54 (3): 85-102.

Strathman, J.G., T.J. Kimpel, K.J. Dueker. 2001a. *Evaluation of transit operations: Data applications of TriMet's automated bus dispatching system.* Portland, OR: Project Report 120, Center for Urban Studies, Portland State University.

Strathman, J.G., T.J. Kimpel, K.J. Dueker, R.L Gerhart, K. Turner, and D. Griffin. (2001b). Bus transit operation control: review and an experiment involving TriMet's automated bus dispatching system. *Journal of Public Transportation*, 4 (1): 1-26.

Transit Capacity and Quality of Service Manual, First Edition. 1999. Transit Cooperative Research Program, Transportation Research Board, Web Document 6, Part 2: 2 90 – 2-91, available at http://gulliver.trb.org/publications/tcrp/tcrp_webdoc_6-a.pdf .

About the Authors

Kenneth J. Dueker (*duekerk@pdx.edu*) is emeritus professor of urban studies and planning, and research associate of the Center for Urban Studies at Portland State University. He formerly directed both the Center for Urban Studies and the Transportation Studies Center. His interests are in transportation and land use interactions, travel and parking behavior, and geographic information systems – transportation.

Thomas J. Kimpel (*kimpelt@pdx.edu*) is a research associate in the Center for Urban Studies at Portland State University. His interests are in transit operations, service reliability, passenger demand analysis, and geographic information systems.

James G. Strathman (*strathmanj@pdx.edu*) is a professor of urban studies and planning and director of the Center for Urban Studies at Portland State University. His research focuses on transit operations and planning and travel behavior.

Steve Callas (*callasc@TriMet.org*) is coordinator of service and performance analysis at TriMet in Portland Oregon, where he is responsible for the agency's performance statistics. He has a master's of urban and regional planning from the University of Iowa.

Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data

Amer Shalaby, University of Toronto
Ali Farhan, City of Calgary

Abstract

The emphasis of this research effort was on using AVL and APC dynamic data to develop a bus travel time model capable of providing real-time information on bus arrival and departure times to passengers (via traveler information services) and to transit controllers for the application of proactive control strategies. The developed model is comprised of two Kalman filter algorithms for the prediction of running times and dwell times alternately in an integrated framework. The AVL and APC data used were obtained for a specific bus route in Downtown Toronto. The performance of the developed prediction model was tested using "hold out" data and other data from a microsimulation model representing different scenarios of bus operation along the investigated route using the VISSIM microsimulation software package. The Kalman filter-based model outperformed other conventional models in terms of accuracy, demonstrating the dynamic ability to update itself based on new data that reflected the changing characteristics of the transit-operating environment.

A user-interactive system was developed to provide continuous information on the expected arrival and departure times of buses at downstream stops, hence the expected deviations from schedule. The system enables the user to assess in real time

transit stop-based control actions to avoid *such deviations before their occurrence, hence allowing for* proactive *control, as opposed to the traditional* reactive *control, which attempts to recover the schedule* after *deviations occur.*

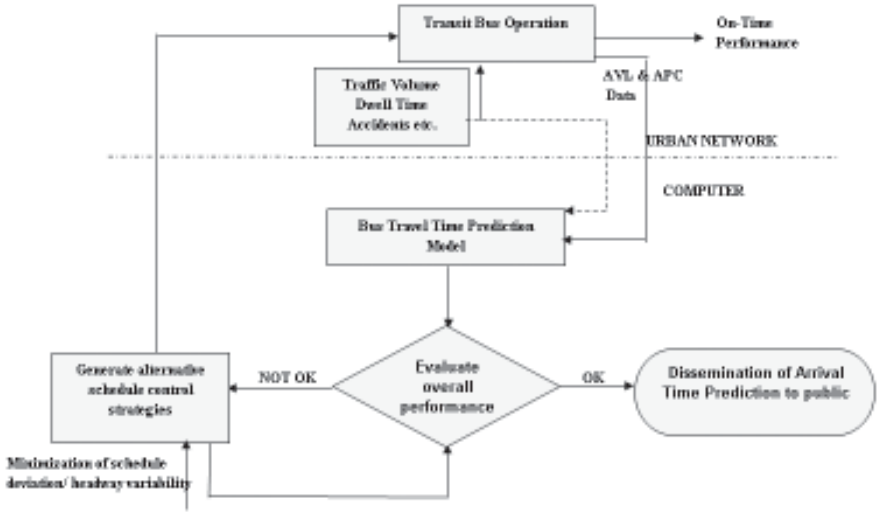
Introduction

Recently, a growing interest has been developing in various Advanced Public Transportation Systems (APTS) solutions that mainly aim at maximizing transit system efficiency and productivity using emerging technologies. Examples of such advanced technologies include Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) systems.

Several researchers (Kalaputapu and Demetsky 1995; Lin and Zeng 1999; Wall and Dailey 1999) have used AVL (and less often APC) data to develop models specifically for bus travel time prediction. The motivation for developing these models was mostly for providing information to transit riders on expected bus arrival times with virtually no sensitivity of such models to operations control strategies. Thus, these models included very simple independent variables such as historical link travel times, upstream schedule deviations, and headway distributions, in addition to the current location of the next bus.

This study develops a dynamic bus arrival/departure time prediction model, using AVL and APC information, for dynamic operations control and dissemination of real-time transit information. The study is part of a larger project that aims at developing an integrated system for dynamic operations control and real-time transit information. Currently, almost all transit operators implement control strategies, such as bus holding and expressing, after detecting schedule/headway deviations in the system, hence *reactive* in nature. The proposed system (shown in Figure 1) takes a *proactive* approach to operations control that would enable the controller to implement preventive strategies before the actual occurrence of deviations. This innovative approach requires the use of arrival/departure time models sensitive to the considered control strategies (mainly stop-based strategies). This research study focuses on developing a model of such characteristics.

Figure 1. Integrated Operations Control and Information-Dissemination System



Data

The data used for this study were collected from bus route Number 5 in the Downtown Toronto area in May 2001. The route length is approximately 6.5 km, spanning 27 bus stops in each direction, 6 of which are time-point stops located at points of high passenger demand (e.g., major intersections). The route starts at the Eglinton subway station stop in the north and ends at the Treasury stop in the south during the morning peak period. At the other times of the day, the route ends further south at the Elm stop. There are 21 signalized intersections along the route, 10 of which are actuated SCOOT system signals. The schedule headway during the AM and PM peak periods is 12 minutes, increasing to 30 minutes during off peak. For the duration of the study (five weekdays in May 2001), the Toronto Transit Commission (TTC) assigned 4 buses to the route, each fitted with a GPS (Global Positioning System) receiver and an APC (Automatic Passenger Counter). Each time the bus stopped, the bus location was recorded using the GPS receiver. Also, the numbers of passengers boarding and alighting at bus stops were recorded using the APC. The route was segmented into 5 links in each direction, with each link starting and ending at two consecutive time-point stops. The

links range from 0.40 to 1.7 km in length, depending on the spacing between the time-point stops, and may include 2 to 8 minor bus stops. This study focused on modeling travel times along those links for the morning peak-hour bus operation.

Approach

As implied earlier, most of the models found in the literature (e.g., Kalaputapu and Demetsky 1995; Lin and Zeng 1999; Wall and Dailey 1999; Farhan et al. 2002) have included bus dwell times along any link in the travel time of that link (i.e., link travel time includes running plus dwell times). As such, these models cannot consider explicitly the effect of late or early bus arrivals at bus stops on the dwell times at those stops and *vice versa*. Ignoring such relationship yields these models insensitive to the effects of variations at upstream bus stops, such as demand surge, bus holding strategy, and bus expressing strategy, etc., on downstream bus arrivals and subsequent dwell times. The approach taken in this study addresses this issue.

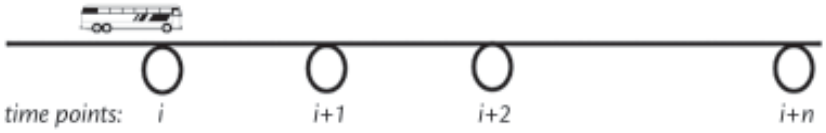
Conceptual Framework

The link running time and bus dwell time are modeled separately in this study but in a consistent single modeling framework. It is assumed that real-time information on vehicle location, numbers of boarding and alighting passengers at bus stops, and bus arrival and departure times is known from the AVL and APC systems. The prediction modeling system consists of two separate algorithms, each based on the Kalman filter method. To predict the bus running time along a particular link at instant k+1, the first algorithm, *Link Running Time Prediction Algorithm*, makes use of the last three-day historical data of the bus link running time for the instant of prediction k+1, as well as the bus link running time for the previous bus on the current day at the instant k. The study used data for the previous three days only as this was deemed practical, given the limited historical data available for the study. Obviously, in real-world applications, the algorithm can make use of longer ranges of historical data. The second algorithm, *Passenger Arrival Rate Prediction Algorithm*, employs similar historical data of passenger arrival rate. To predict the dwell time at a particular stop, the predicted arrival rate is simply multiplied by the predicted headway (i.e., the actual arrival time of the last bus minus the predicted arrival time of the next bus) and by the passenger boarding time (assumed 2.5 seconds per passenger).

Separating the bus dwell time prediction from bus running time prediction in this modeling framework enhances the model's ability to capture the effects of lateness or earliness of bus arrivals at stops on the bus dwell time at those stops, and hence the bus departure from such stops. This is simply because bus dwell time at a stop is affected by the actual arrival time of the bus at that stop, particularly for high frequency transit routes where passengers are expected to arrive at a nearly constant rate (i.e., the later the bus, the longer the dwell time and *vice versa*). In addition, since the model treats dwell time separately, it is sensitive to stop-based control strategies such as bus holding and expressing.

In order to better understand the prediction-modeling framework, Figure 2 shows a schematic of a hypothetical transit route. The route is divided into a number of links between bus stops. When the transit bus n leaves stop i, the actual departure time is known from the AVL system. At this instant, the Kalman filter prediction algorithm for link running times will predict the next link running time $RT_{n\ (i,i+1)}$. Subsequently, the predicted arrival time of the bus at the downstream bus stop i+1 can be determined.

Figure 2. Schematic of a Bus Route with Several Stops



Assuming that bus n is currently at stop i

$$AT_{n\ (i+1)} = DT_{n\ (i)} + RT_{n\ (i,i+1)} \quad (1)$$

Where:

$AT_{n\ (i+1)}$ is the predicted arrival time of bus n at stop i+1

$RT_{n\ (i,i+1)}$ is the predicted running time between i and i+1 from Kalman Filter prediction algorithm

$DT_{n\ (i)}$ is the actual departure time of bus n from stop i

This predicted arrival time $AT_{n\ (i+1)}$ is used to predict the dwell time for bus n at stop i+1 based on the passenger arrival rate and the average passenger boarding time at stop i+1.

$$DWT_{n\ (i+1)} = \beta_{(i+1)} * [AT_{n\ (i+1)} - AT_{n-1\ \ (i+1)}] * \delta_{avg\ (i+1)} \quad (2)$$

Where:

$DWT_{n\ (i+1)}$ is the predicted dwell time for bus n at stop i+1

$\beta_{(i+1)}$ represents the predicted passenger arrival rate at stop i+1 from Kalman filter prediction algorithm

$AT_{n-1\ (i+1)}$ is the actual arrival time of the previous bus n-1 at stop i+1

$[AT_{n\ (i+1)} - AT_{n-1\ (i+1)}]$ is the predicted headway for bus n at stop i+1

$\delta_{avg\ (i+1)}$ represents average passenger boarding time at stop i+1, assumed to be 2.5 seconds/passenger.

In equation (2), the assumption is that the boarding passengers at each bus stop have a significant effect on bus dwell time at that stop, compared with alighting passengers. The time points used in this study, for which equation (2) applies, are located at high demand spots (i.e., subway station and major intersections) where stop skipping because of no passenger demand is extremely rare. If stop skipping at a particular time point were frequent, equation (2) would need to be modified to address this problem.

Having the arrival time and dwell time for bus n at stop i+1 predicted, it is now easy to calculate the predicted departure time for bus n from stop i+1 by adding the predicted arrival time to the predicted dwell time at stop i+1.

$$DT_{n\ (i+1)} = AT_{n\ (i+1)} + DWT_{n\ (i+1)} \quad (3)$$

Where:

$DT_{n\ (i+1)}$ is the predicted departure time for bus n from stop i+1

This departure time prediction $DT_{n\ (i+1)}$ is a function of both arrival time prediction and dwell time prediction at stop i+1. Hence, the effect of any variations in bus operation (i.e., bus early or late) could be captured in this stop and will be reflected on all downstream bus stops.

Similarly, predictions of arrival times and departure times at all downstream stops can be computed while the bus is still at stop i. This process is updated every time the bus leaves or arrives at a new time-point stop.

Kalman-Filter Prediction Algorithms

As mentioned above, the prediction modeling system consists of two Kalman flter algorithms. In general, the Kalman flter is a linear recursive predictive update algorithm used to estimate the parameters of a process model. Starting with initial estimates, the Kalman flter allows the parameters of the model to be predicted and adjusted with each new measurement. Its ability to combine the effects of noise of both the process and measurements, in addition to its easy computational algorithms, has made it very popular in many research fields and applications, particularly in the area of autonomous and assisted navigation (for further information on Kalman filters, see Maybeck 1979).

The main assumption used in developing the Kalman filters is that the patterns of the link running time and passenger arrival rate at stops are cyclic for a specific period of day. In other words, knowledge of the link travel time and number of passengers waiting for a specific bus in a certain period of time will allow one to predict these variables for the next bus during the same period, so long as conditions remain steady. When conditions change (e.g., demand surge at a stop and/or an incident occurred at a link), the model can update the effect of the new conditions on the predictions, so long as the new conditions persist for a sufficient length of time (at least one headway length).

The Kalman filter algorithm works conceptually as follows. The last three-day historical data of actual running times along a particular link at the instant k+1 plus the last running time observation at the instant k on the current day are used to predict the bus running time at the instant k+1. Similarly, passenger arrival rates of the previous three days at the instant k+1 plus the passenger arrival rate at the instant k on the current day are used to predict the passenger arrival rate at the instant k+1. The historical passenger arrival rate is obtained from the APC data as in this fashion: The number of on-passengers at a bus stop is divided by the most recent headway (i.e., the arrival time of the previous bus minus the arrival time of

the current bus). Below are the equations used for the Kalman filter prediction algorithms.

Link Running Time Prediction Algorithm

Generally, a Kalman filter algorithm for bus link running time has the following structure (Reinhoudt and Velastin 1997):

$$g(k+1) = \frac{e(k) + VAR[data_{out}]}{VAR(data_{in}) + VAR[data_{out}] + e(k)} \quad (4)$$

$$a(k+1) = 1 - g(k+1) \quad (5)$$

$$e(k+1) = VAR[data_{in}] \times g(k+1) \quad (6)$$

$$P(k+1) = a(k+1) \times art(k) + g(k+1) \times art_1(k+1) \quad (7)$$

where:

g	equals the filter gain
a	is the loop gain
e	represents filter error
p	equals prediction
art(k)	is actual running time of the previous bus at instant (k)
$art_1(k+1)$	is actual running time of the previous day at instant (k+1)
$VAR[data_{out}]$	equals the prediction variance
$VAR[data_{in}]$	is the last three days "art3(k+1), art2(k+1) and art1(k+1)" variance

The input variance VAR[$data_{in}$] is calculated at each instant k + 1 using the actual running time values for the last three days: $art_1(k + 1)$, $art_2(k + 1)$ and $art_3(k + 1)$:

$$VAR[data_{in}] = VAR[art_1(k + 1), art_2(k + 1), art_3(k + 1)] \quad (8)$$

where:

$art_1(k+1)$	is the actual running time of the bus at instant (k+1) on the previous day
$art_2(k+1)$	is the actual running time of the bus at instant (k+1) two days ago
$art_3(k+1)$	is the actual running time of the bus at instant (k+1) three days ago

The definition of the variance for a random variable is:

$$VAR[X] = E[(X - E[X])^2] \quad (9)$$

$$E(X) = Avg(art) = \frac{art_1(k+1) + art_2(k+1) + art_3(k+1)}{3} \quad (10)$$

Now the variance can be calculated as shown in equation (14):

$$\Delta_1 = [art_1(k+1) - avg(art)]^2 \quad (11)$$

$$\Delta_2 = [art_2(k+1) - avg(art)]^2 \quad (12)$$

$$\Delta_1 = [art_3(k+1) - avg(art)]^2 \quad (13)$$

$$VAR[data_{in}] = \frac{\Delta_1 + \Delta_2 + \Delta_3}{3} \quad (14)$$

VAR[$data_{out}$] is based on the model prediction output and the corresponding future observation. Both pieces of data are not available, since the prediction is not made yet and the future trip has not been made either. Ideally, VAR[$data_{out}$] should be equal to VAR[$data_{in}$] for good prediction performance (Maybeck 1979). Now, a new variance is introduced VAR[$local_{data}$]. It is equal to the variance of the input and output data.

$$VAR[local_{data}] = VAR[data_{in}] = VAR[data_{out}] \quad (15)$$

and equations (4) and (6) become:

$$g(k+1) = \frac{e(k) + VAR[local_{data}]}{e(k) + 2 \cdot VAR[local_{data}]} \quad (16)$$

$$e(k+1) = VAR[local_{data}] \cdot g(k+1) \quad (17)$$

Now it becomes easy to implement the actual Kalman filter algorithm to predict the bus running times along the links. The order of applying the equations should be (16), (5), (17), and (7).

Passenger Arrival Rate Prediction Algorithm

A Kalman filter algorithm was also developed to predict the passenger arrival rate using data from the APC and AVL systems. At the prediction instant k+1, the historical passenger arrival rates (at instant k+1 of the previous three days and instant k of the current day) at a particular stop are computed based on the number of corresponding boardings divided by the actual previous headway. Similar equations to those of the running time Kalman filter were developed and used to predict the passenger arrival rate at instant k+1.

Model Performance Evaluation

In order to assess the predictive performance of the Kalman filter model, it is compared with three previously developed models for the same route. They include historical average model, regression model, and an artificial neural network (specifically, Time Lag Recurrent Network [TLRN]). A more detailed exposition of the models can be found elsewhere (Farhan et al. 2002). Similar to most models found in the literature, the regression and TLRN models predict individual link travel times, which include running plus dwell times. Another distinct feature of those models is their static nature, in that the model parameters are not updated with new available data. As mentioned earlier, the AVL and APC data for the study route were available for five consecutive days. The regression and TLRN models were developed using data of four days only, with the fifth day's data held out for testing.

Four testing datasets were used for the comparison; the first set includes the hold out data of the fifth day of observations, while the remaining three sets include artificial data collected from three different bus operation scenarios representing: normal-condition bus operation, special-event scenario where there is a demand surge on transit service, and lane-closure scenario where one lane on a specific link is blocked (e.g., due to an accident or construction work). In contrast to the hold out data and normal condition scenario, the lane closure and special event scenarios represent atypical conditions. Because real-world data of such conditions are hard to obtain, the VISSIM traffic microsimulation software was used to simulate these scenarios. Simulation of the entire corridor, calibration results and simulation of the scenarios are described elsewhere (Farhan 2002). After each simulation run, all the necessary data required for model testing was extracted and analyzed. Three prediction error measurements were computed for all developed models to test the model performance (Okutani and Stephanedes 1984). These error indices include:

Mean relative error ($\mathbf{g}_{mean}$), which indicates the expected error as a fraction of the measurement

$$\mathcal{E}mean = \frac{1}{N}\sum_{t}\left|\left\{\frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)}\right\}\right| \qquad (18)$$

Root squared relative error (ε_{rs}), which captures large prediction errors

$$\varepsilon_{rs} = \sqrt{\frac{1}{\sum_t X_{true}(t)} \sum_t \left\{ \frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)} \right\}^2 X_{true}(t)} \quad (19)$$

Maximum relative error (ε_{max}), which captures the maximum prediction error

$$\varepsilon_{max} = \max \left| \left\{ \frac{X_{true}(t) - X_{pred}(t)}{X_{true}(t)} \right\} \right| \quad (20)$$

where:

N is the number of samples (here N=50) $X_{true}(t)$ = measured value at time t

$X_{pred}(t)$ is the predicted value at time t.

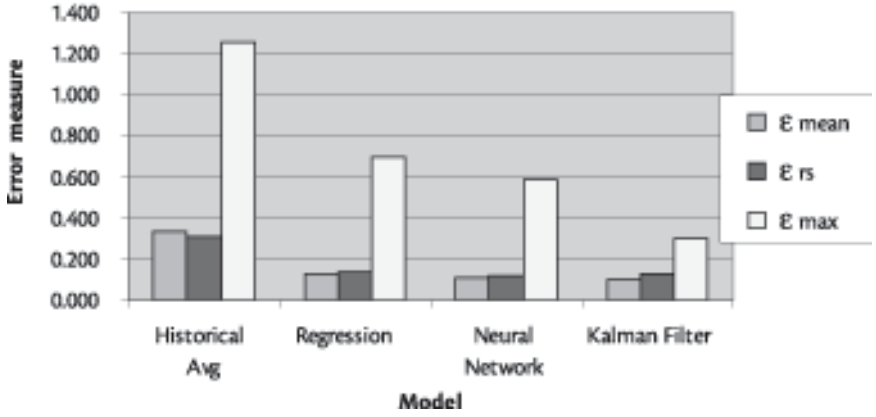
Table 1. Relative Error Results of the Prediction Models Using "Hold Out" Data

Link	Error Indices	Historical Avg	Regression	NN	Kalman Filter
Link 1	ε_{mean}	0.268	0.099	0.076	0.069
	ε_{rs}	0.332	0.115	0.094	0.074
	ε_{max}	0.900	0.192	0.168	0.117
Link 2	ε_{mean}	0.537	0.134	0.063	0.028
	ε_{rs}	0.603	0.149	0.075	0.036
	ε_{max}	3.172	0.224	0.131	0.077
Link 3	ε_{mean}	0.369	0.200	0.165	0.076
	ε_{rs}	0.543	0.220	0.166	0.093
	ε_{max}	0.801	0.466	0.342	0.227
Link 4	ε_{mean}	0.130	0.079	0.071	0.087
	ε_{rs}	0.208	0.092	0.088	0.109
	ε_{max}	0.502	0.228	0.153	0.232
Link 5	ε_{mean}	0.168	0.084	0.104	0.044
	ε_{rs}	0.181	0.076	0.120	0.055
	ε_{max}	0.502	0.228	0.240	0.122

Table 1 shows the three error measures ε_{mean}, ε_{rs}, ε_{max} for the hold out dataset, while Figure 3 (a, b, c) summarize the performance of the three prediction models for each simulated scenario. Obviously, the lower the error, the better the model performance.

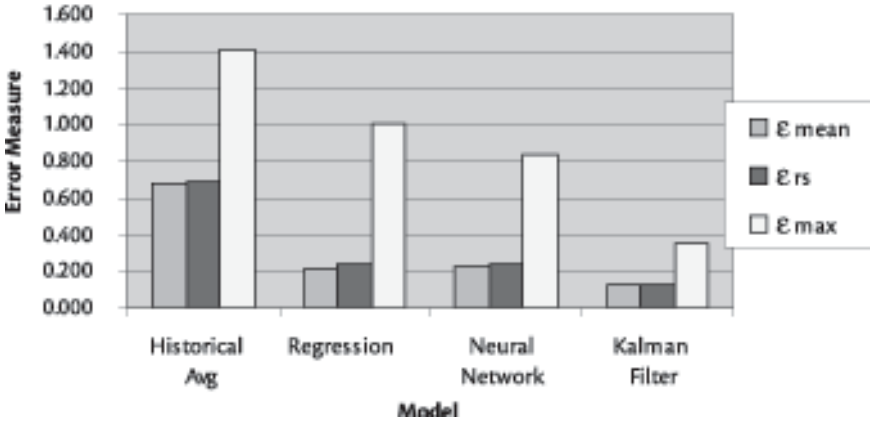
Figure 3. Relative Error Results of the Prediction Models Using Artificial Data (a, b, c)

(a) **Normal Condition Scenario**



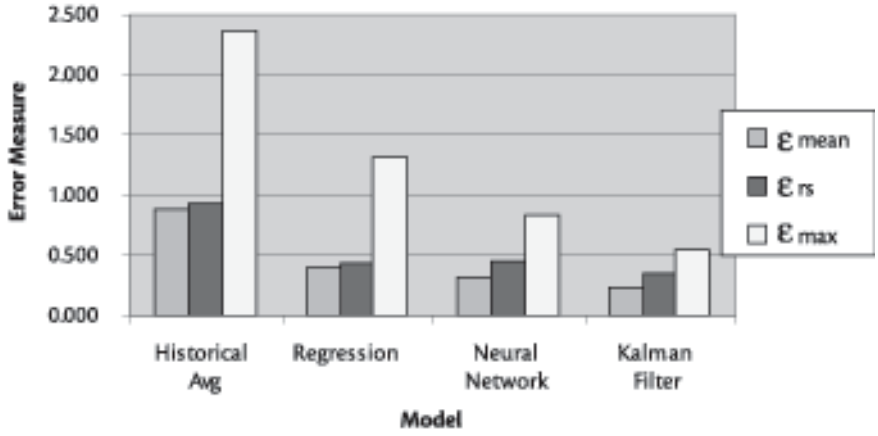
Normal Condition Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ε_{mean}	0.330	0.124	0.111	0.097
ε_{rs}	0.305	0.132	0.120	0.124
ε_{max}	1.253	0.695	0.584	0.299

(b) **Special Event Scenario**



Special Event Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ε_{mean}	0.679	0.218	0.220	0.123
ε_{rs}	0.685	0.240	0.239	0.127
ε_{max}	1.411	0.998	0.830	0.349

(c) **Lane Closure Scenario**



Lane Closure Scenario				
Total	Historical Avg	Regression	Neural Network	Kalman Filter
ε_{mean}	0.881	0.392	0.316	0.232
ε_{rs}	0.933	0.428	0.442	0.349
ε_{max}	2.362	1.324	0.830	0.547

Discussion

The results summarized in Table 1 show that for all links, the Kalman flter model provides the minimum value for the error measures $\mathbf{g}_{mean}$, $\mathbf{g}_{rs}$, $\mathbf{g}_{max}$, pointing to the fact that its performance was the best compared with the other prediction models, except for link #4 where the regression and TLRN models perform slightly better than the Kalman flter model.

Table 1 and Figure 3 (a, b, c) shows there is no significant difference in the performance of the regression and artificial neural network models for the three different scenarios. Both models give similar performance results for each scenario; their accuracy performance decreased for the special event and lane closure scenarios. But in general, the artificial neural network model always gives lower values for the relative error indices.

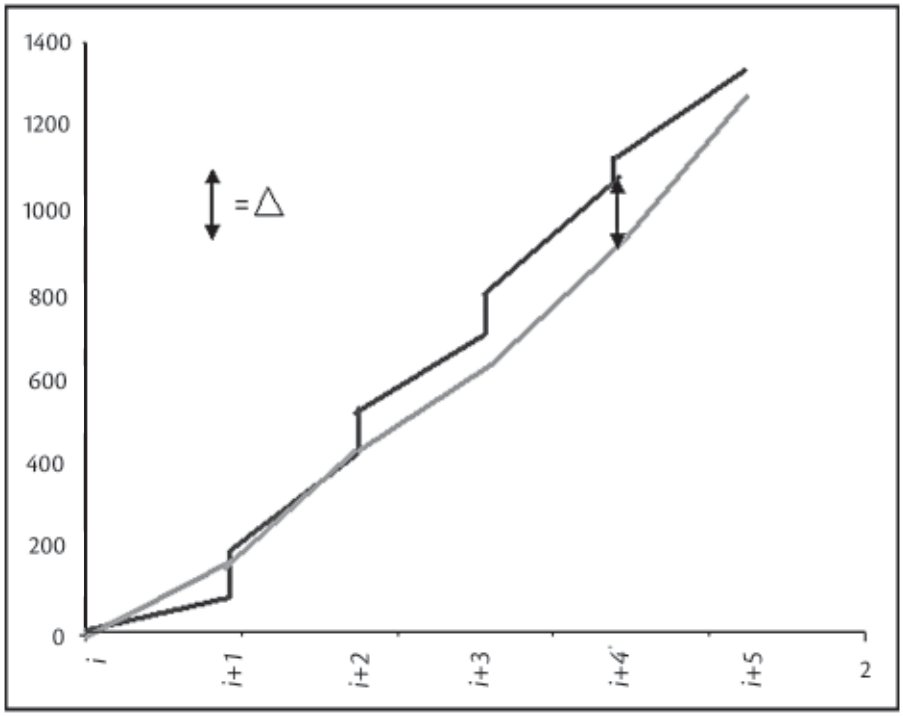
The Kalman filter model shows the best prediction performance in the simulated scenarios. In particular, it showed superior performance to the other models in the special event and lane closure scenarios. These results show the superior performance of the Kalman filter model compared with other prediction models in terms of relative error. The results also demonstrate how this model can capture dynamic changes due to different bus operation characteristics.

In addition to its highly accurate performance in dynamic environments, the model has the advantage of capturing the effects of control strategies, such as holding and expressing at upstream bus stops. For example, if the bus is currently at a time point where it will be held for an additional one minute, the model appropriately captures the effect of this extra time on the arrival time at the next bus stop and the dwell time at the next stop (which is function of number of passengers waiting at that bus stop when the bus is predicted to arrive) and so forth for the prediction of arrival and dwell times at subsequent stops.

User-Interactive Decision Support System

The developed arrival/departure time prediction model was used to build a DSS (Decision Support System) for dynamic control. This system simply uses the timetable of the time points to develop a time profile for each scheduled trip along the route (i.e., schedule travel time profile; see Figure 4), which is done for each bus at the start of its journey. Another prediction travel time profile is constructed using the Kalman filter prediction model. The prediction travel time profile is updated dynamically every time the bus arrives and departs from a time point.

Figure 4. Schedule Travel Time Profile

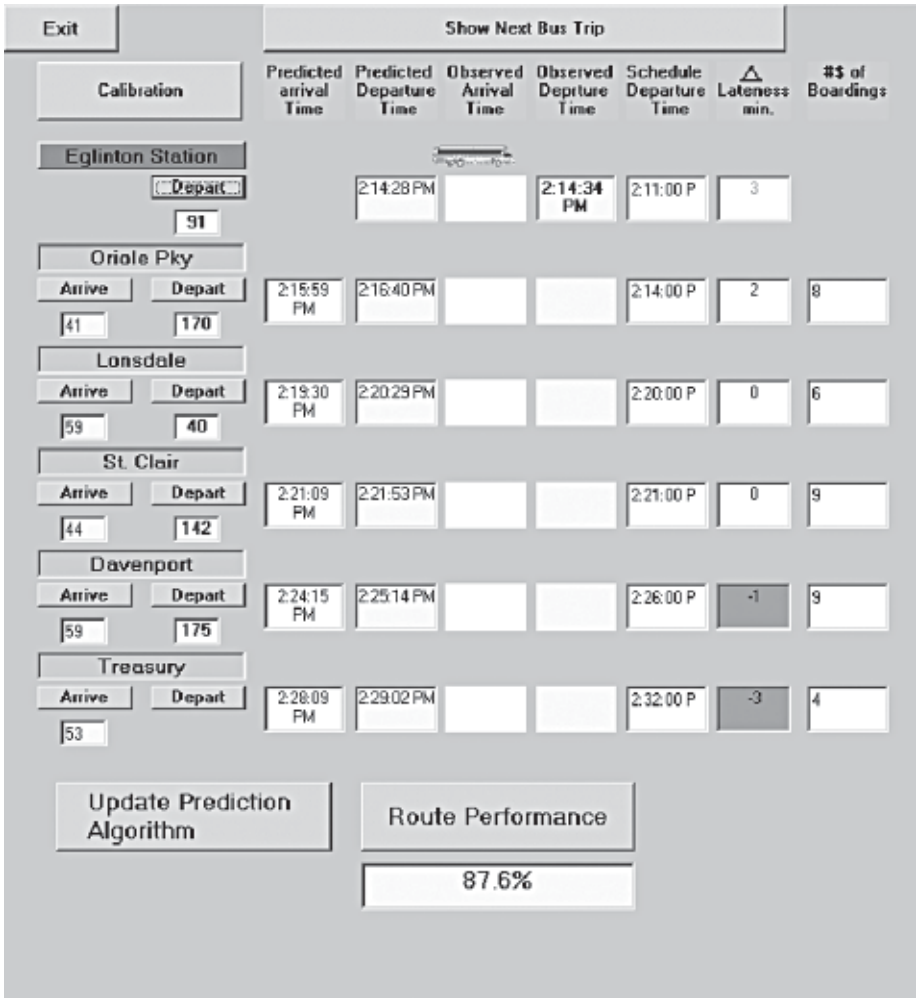


By using these two travel-time profiles, we are able to predict if the bus is running early or late at each time point. This is shown in Figure 4 as Δ. A positive value of Δ means that the prediction profile is currently lying above the scheduled one (bus will reach the downstream time points late), while a negative value of Δ occurs when the prediction profile lies below the schedule (bus will be ahead of schedule). In these cases, implementation of a corrective proactive control strategy is required to prevent expected schedule deviation downstream. A value of 0 refers to the compliance of the bus to the schedule. The value of Δ is the key factor for deciding what type of control strategy to implement. If Δ is positive, an expressing type of control is required to be applied, while a negative value of Δ indicates implementation of some type of holding strategy.

System Design and Architecture

The proposed system, shown in Figure 5, is an interactive program developed using the Visual Basic programming language. This program effectively utilizes AVL and APC data for dynamic bus arrival/departure information and performance analysis at downstream bus stops for the purpose of applying real-time, proactive control strategies.

Figure 5. Illustration of the Interactive Decision Support System



The program records the dynamic actual location and time of the transit vehicle when it arrives and leaves all time points along the transit route based on the AVL data. Also, it records the number of passengers alighting and departing the bus at each stop. Real time AVL and APC information is transferred instantly from the server to the program. For a specific bus trip, at the departure instant of bus from the terminal station, the prediction algorithms will automatically be activated to predict arrival and departure times of the bus at all downstream time points (prediction profile). At the same time, the associated bus departure schedule for all time points along the route is also displayed (schedule profile). The difference between the predicted and the schedule departure (Δ information) is automatically computed for all time points. If the value of the predicted bus time deviation Δ is within an accepted range (e.g., 0 to 2 minutes), the predicted departure times are within the schedule, and no control strategies are required to be implemented. In such a case, the program will display black font color with white background for Δ value labels. On the other hand, when Δ values are more than 2 minutes (i.e., bus expected to arrive/depart late at downstream time points), the font color will display red. The transit controller can interact with the program to assess the effect of bus expressing at one or more downstream time points (by setting dwell times at 0 for such time points) so as to as reduce predicted deviations. If the Δ value is less than 0 (i.e., bus expected to be ahead of schedule), then the Δ value label background is displayed red and the font color is black. The controller can assess the effect of bus holding at one or more downstream time points (by increasing the dwell times at those time points). The interaction with the program could be done by setting values under the "arrive" or "depart" button of the selected current/downstream time point (as shown on the left hand side of Figure 5), then clicking on the corresponding "arrive" or "depart" button to update model predictions.

The prediction algorithms of the system will be dynamically updated based on the AVL and APC data. As soon as the bus arrives or departs a new time point, new arrival and departure time predictions and new Δ values for the remaining time points downstream will be processed.

At the end of each trip, the system records the observed AVL arrival and departure times as well as the real APC data regarding the number of passenger boardings and alightings at each bus stop. These data are used to update the system historical database (link running time and passenger arrival rate at time points) to be used for the Kalman filter predictions of future trips.

In addition, the system computes the on-time route performance for each trip by comparing the actual bus arrival/departure with the schedule arrival/departure times for all time points. The average on-time performance is automatically calculated and displayed on the screen for all trips. This feature provides the transit management with an easy tool to evaluate the route level of service in terms of on-time performance. The information and analysis provided by this system could possibly be used for updating and adjusting schedules.

Conclusion

An innovative model was developed for dynamic bus arrival and departure time prediction. The model is based on two Kalman filter algorithms for the prediction of running times and dwell times alternately in an integrated framework. As such, the model can capture the interaction between the two variables (i.e., the effect of one on the other). The model was shown to outperform other traditional models (regression and neural network models) in terms of predictive ability when tested on "hold out" real-world data. More importantly, the superiority of the model was even more prominent when tested on two simulated scenarios representing passenger demand surge (e.g., because of a special event) and lane closure (e.g., because of an incident). This is primarily due to continuous updating of the model parameters based on dynamic real-time data, as opposed to traditional models which are typically calibrated using historical data, with infrequent recalibration of the model, if any.

Because dwell time is predicted separately and its effect on bus arrival times at downstream stops is accounted for, the model can be used for assessing transit stop-based dynamic control actions (e.g., bus holding, bus expressing). A user-interactive DSS was developed to provide continuous information on the expected arrival and departure times of buses at downstream stops; hence the expected deviations from schedule. The system enables the user to assess in real time transit stop-based control actions to *avoid* such deviations before their occurrence, allowing for *proactive* control, as opposed to the traditional *reactive* control which attempts to recover the schedule *after* deviations occur.

The model developed here was based on data from one bus route in downtown Toronto. However, the same modeling approach is applicable to other medium- to low-frequency routes where schedule control and dissemination of expected arrival times are relevant.

Further work can improve the model developed here in several ways. For example, better representative distributions of passenger arrivals at bus stops could be attempted instead of the implied uniform distribution assumed here. Also, further investigation is required to develop predictive models for overlapping routes that serve the same bus stops. In such cases, a special consideration should be given to dwell time prediction. Finally, the assessment of the model developed here would be greatly enhanced if tested in the field under both normal and atypical conditions.

Acknowledgments

The research reported in this article was funded by the Geomatics for Informed Decisions (GEOIDE) network, a Network of Centres of Excellence funded by the Government of Canada.

References

Farhan, A., A. Shalaby, and T. Sayed. 2002. Bus travel time prediction using GPS and APC. ASCE 7th International Conference on Applications of Advanced Technology in Transportation, Cambridge, Massachusetts (August).

Farhan, Ali. Bus arrival time prediction for dynamic operations control and passenger information systems. 2002. Unpublished Thesis of Masters of Applied Science, Department of Civil Engineering, University of Toronto.

Kalaputapu, Ravi, and Michael J. Demetsky. 1995. Modeling schedule deviations of buses using Automatic Vehicle Location data and artificial neural networks. *Transportation Research Record 1497*, 44-52.

Lin, Wei-Hua, and Jian Zeng. 1999. Experimental study of real-time bus arrival time prediction with GPS data. *Transportation Research Record 1666*, 101-109.

Maybeck, Peter S. 1979. *Stochastic models, estimation and control.* Vol. 1, Academic Press.

Okutani, Iwao, and Yorgos J. Stephanedes. 1984. Dynamic prediction of traffic volume through Kalman filter. Transportation Research Board 18B(1), 1-11.

Reinhoudt, Edwin M., and S. A. Velastin. 1997. A dynamic predicting algorithm for estimating bus arrival time. 8th IFAC/IFIP/IFORS Symposium, Vol. 3.,1295-8. Crete, Greece (June 16-18).

Wall, Z., and D. J. Dailey. 1999. An algorithm for predicting the arrival time of mass transit vehicle using Automatic Vehicle Location data. Transportation Research Board Paper No. 990870. 78th Annual Meeting, Washington, D.C.

About the Authors

Amer Shalaby (*amer@ecf.utoronto.ca*) received his master and Ph.D. degrees from the Department of Civil Engineering, University of Toronto (U of T) in 1991 and 1997, respectively. He was a postdoctoral fellow at U of T's Joint Program of Transportation in 1996-1997 and an NSERC Industrial Research Fellow at IBI Group in 1997-1998. Between 1998 and 2000, he was an assistant professor at Ryerson University, where he also directed the Vehicle Safety Research Centre. Since January 2001, Dr. Shalaby has been an assistant professor at the University of Toronto. He is specialized in the area of Intelligent Transportation Systems, with emphasis on transit applications. His research interests also include transportation system planning and road safety.

Ali Farhan *(alimfarhan@hotmail.com)* received his master's degree from the Department of Civil Engineering, University of Toronto (U of T) in 2002. Subsequently, he worked as a research associate at the Joint Program in Transportation at the University of Toronto, then as a transportation engineer with Maricopa Association of Governments, Phoenix, Arizona. Currently, Mr. Farhan works for the City of Calgary.

Transit Network Optimization – Minimizing Transfers and Optimizing Route Directness

Fang Zhao, Florida International University
Ike Ubaka, Florida Department of Transportation

Abstract

This paper presents a mathematical methodology for transit route network optimization. The goal is to provide an effective computational tool for the optimization of large-scale transit route networks. The objectives are to minimize transfers and optimize route directness while maximizing service coverage. The formulation of the methodology consists of three parts: (1) representation of transit route network solution space; (2) representation of transit route and network constraints; and (3) solution search schemes. The methodology has been implemented as a computer program and has been tested using previously published results. Results of these tests and results from the application of the methodology to a large-scale realistic network optimization problem in Miami-Dade County, Florida are presented.

Introduction

Transit route network (TRN) design is an important component in the transit planning process, which also includes transit network schedule (TNS) design. A TRN optimization process attempts to find the route network structure with optimal transfer, route directness, and ridership coverage. Unfortunately, TRN design optimization processes suffer from combinatorial intractability, and thus far for practical transit network problems of large scales, TRN designs seem to be

limited to the use of various heuristic approaches where the solution search schemes are based on a collection of design guidelines, criteria established from past experiences, and cost and feasibility constraints. A systematic mathematical methodology applicable to large-scale transit networks for TRN optimization design seems to be missing.

The quality of a TRN may be evaluated in terms of a number of network parameters, such as route directness, service coverage, network efficiency, and number of transfers required. Route directness refers to the difference between the trip lengths,[1] if the trip is to be made by transit or by a car following the shortest path. Service coverage refers to the percentage of the total estimated demand (measured by transit trips) that potentially can be satisfied by the transit services based on a given transit route network. In this study, if the origin and destination of a potential transit trip are within walking distance of a transit stop and are connected by transit routes, the trip is considered served by the network or "covered." Network efficiency reflects the cost of providing transit services within a given network, other things being equal. Transfers are a result of the inability of a given network to provide direct service between all pairs of origins and destinations. Stern (1996) conducted a survey of various transit agencies in the United States, and about 58% of the respondents believed that transit riders were only willing to transfer once per trip. This suggests that the ridership of a transit system may be increased by merely reducing required transfers through the optimization of a TRN configuration. In addition to increasing ridership, an improved TRN configuration may also reduce transit operating cost and allow more services to be provided.

For transit systems with small bus route networks, a seasoned planner may be able to obtain near optimal bus route network results based on personal knowledge, experience, and certain guidelines. For large transit systems, intuition, experiences, and simple guidelines may be insufficient to produce even near-optimal transit route network configurations, due to the problem complexity. Therefore, systematic methodologies are needed to obtain better TRN configurations. This paper presents a methodology for TRN structure optimization based on a mathematical approach with the objectives of minimizing transfers, optimizing route directness, and maximizing service coverage (Zhao 2003). The methodology has been implemented as a computer program and has been tested using previously published results and a large-scale realistic network optimization problem in Miami, Florida.

Formulation of A TRN Optimization Problem

A TRN optimization problem may be stated as the determination of a set of transit routes, given a transit demand distribution in a transit service area and subject to a set of feasibility constraints, to achieve objectives that optimize the overall quality of a TRN. Mathematically, a typical network optimization process may be stated as: optimize an objective function $f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{O}) \forall \boldsymbol{x} \in \boldsymbol{X}$ and $\boldsymbol{y} \in \boldsymbol{Y}$, subject to certain constraints, where $\boldsymbol{x}$ is a real vector, $\boldsymbol{y}$ is an integer vector (or a set of vectors), and $\boldsymbol{O}$ is a matrix defined on the network's node set. $\boldsymbol{X}$ is a space of real vectors, and $\boldsymbol{Y}$ is a set of integer vectors

$$\boldsymbol{Y} = \boldsymbol{Y}\left\{\boldsymbol{y}\left(i_1, i_2, \cdots, i_s\right) \middle| \; i_j \in \boldsymbol{N},\; j = 1, 2, \cdots, s\right\}$$

where $\boldsymbol{N}$ is an integer set. A combinatorial optimization problem is a special case of integer optimization problems and refers to an integer optimization problem where the integer vector's component set in vector $\boldsymbol{y}(i_1, i_2,..., i_s)$ is an ordered subset of a larger integer base set $\boldsymbol{N}\{n_1, n_2,..., n_n\}$, i.e., $(i_1, i_2,..., i_s)$ d $\boldsymbol{N}\{n_1, n_2, ..., n_n\}$ and $n \geq s$ (in this paper, an ordered set is enclosed in parentheses while an unordered set is enclosed in brackets). TRN design is a typical combinatorial optimization problem, where the base set $\boldsymbol{N}\{n_1, n_2,..., n_n\}$ is the set of all street nodes suitable to serve as transit stops, and the combinatorial set $\boldsymbol{P}_N$ is the set of all paths in the street network suitable for transit vehicle operations. The matrix $\boldsymbol{O} = \boldsymbol{O}(o_{ij})$ represents the transit demand at street nodes and is the OD matrix as o_{ij} represents the number of transit trips between street node n_i and n_j. This study deals with fixed transit demand problems. $\boldsymbol{O}$ is assumed to be constant, representing transit demand for a given period of time of day, and does not change with transit supply. It should be recognized that, in reality, transit demand may depend on transit supply, thus TRN optimization ideally should be carried out in an iterative manner in a cycle of demand estimation and route network design. A transit route may be represented by an integer vector $\boldsymbol{r}(i_1, i_2,..., i_s)$ with its component set $(i_1, i_2,..., i_s)$ representing the sequence of a transit route's stops. A transit route network consisting of l routes may be represented by a set of integer vectors,

$$\boldsymbol{T}^{(l)} = \boldsymbol{T}^{(l)}\{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_l\}, \boldsymbol{r}_j = \boldsymbol{r}(n_{j1}, n_{j2}, \ldots, n_{js(j)}), (j = 1, 2, \ldots, l) \quad (1)$$

where $s(j)$ is the number of transit stops on transit route $\boldsymbol{r}_j$. •A transit route vector is a member of the combinatorial space $\boldsymbol{P}_N$, and a transit route network is a subset of $\boldsymbol{P}_N$. Based on the above definitions and notations, a fixed demand TRN design optimization problem may be stated as follows:

Maximize/minimize:

$$f(\boldsymbol{x}, \boldsymbol{T}^{(1)}, \boldsymbol{O}) \; \forall \, \boldsymbol{x} \in \boldsymbol{X} \text{ and } \boldsymbol{T}^{(1)} \, \mathrm{d} \, \boldsymbol{P}_N \quad (2)$$

Subject to:

$$p_i(\boldsymbol{x}, \boldsymbol{T}^{(l)}) = 0, (i = 1, 2, \ldots, i_p) \text{ and } q_i(\boldsymbol{x}, \boldsymbol{T}^{(l)}) \leq 0, (i = 1, 2, \ldots, i_q) \quad (3)$$

where the real vector $\boldsymbol{x}$ represents any continuous variables in the optimization process, $\boldsymbol{O}$ is the OD matrix, and expressions in (3) represent various constraints in a TRN design process. Solving the TRN optimization problem, defined above, involves the search for an optimal set of feasible transit routes with unknown topology/geometry. It is difficult to solve problems with a large number of integer variables, since the associated solution procedure involves discrete optimization, which usually requires the search for optimal solutions from an intractable search space (Garey and Johnson 1979).

Literature on TRN Optimization

A great deal of research has been conducted in the area of transit network optimization. The methods in the literature may be roughly grouped into two categories: mathematical approaches and heuristic approaches. However, there are no clear boundaries between these approaches. We consider an approach to be *mathematical* if the problem is formulated as an optimization problem over a relatively complete solution search space. Generic solution search methods are then employed to obtain solutions. Examples of such algorithms include various greedy type algorithms, hill climbing algorithms, simulated annealing approaches, etc. References and descriptions of various mathematical search algorithms may be found (e.g., Bertsekas 1998). We consider an approach to be *heuristic* if domain specific heuristics, guidelines, or criteria are first introduced to establish a solution strategy framework. Mathematical programming or other techniques are then employed to obtain the best results. The main difference between these two approaches is that the mathematical approach formulates a problem on a solution space with certain completeness that, theoretically, should include optimal solu-

tions. In contrast, the heuristic approach formulates a problem directly on solution sub-spaces defined based on domain specified heuristic guidelines.

Table 1 provides the main features of some of the approaches reported in the literature, where MATH represents mathematical optimization, and H&M (heuristic and mathematical) means that the author(s) established a solution based on a heuristic framework, but employed certain mathematical optimization methods at some solution stages. Most of the studies introduced some heuristics or certain simplifying assumptions to limit the solution search space or to reduce optimization objectives to a particular network structure or a few design parameters, e.g., route spacing, route length, stop spacing, bus size, or service frequency. (Detailed information and reviews of various mathematical optimization approaches may be found in Zhao 2003, among others.)

Table 1. Main Features of Some Approaches Used in Transit Network Design

Year	Author	Optimization Objectives	Design Variables	Solution Approaches	Constraints
1979	Mandl	Generalized time	Route	H&M	Coverage & directness
1991	Baaj & Mahmassani	Multi-objects	Route & frequency	AI/heuristic	Multi-constraints (heuristic guidelines)
1992	Bookbinder et al.	Disutility function-transfer inconvenience	Timetable/headway (offset time)	Math	Heuristic guidelines
1994	Shih & Mahmassani	Multi-objects	Route & frequency	H&M	Multi-constraints (heuristic guidelines)
1998	Bruno et al.	Generalized access cost to transit line	Location of a rapid transit line	Math	Route connectivity, demand coverage, etc.
1999	Soehodo & Koshi	Generalized social cost	Route & frequency	H&M	-
2003	Chien et al.	Total operator and user cost	Route shape & headway	Math	Route length, waiting time, load factors, etc.

The advantage of heuristic approaches is that they are always able to provide feasible solutions to problems of any size while the main disadvantage is that their results are almost certainly do not provide global or even local optimal solutions. This may be because heuristic search schemes are usually ad hoc procedures based on computer simulations of human design processes guided by heuristic rules. The corresponding search spaces are usually not clearly defined and search results are likely to be biased toward existing systems or any systems on which the set of design heuristics are based.

Compared with other methods in transit network design, mathematical approaches usually have more rigorous problem statements. A major disadvantage of mathematical approaches in TRN design is the computational intractability due to the need to search for optimal solutions in a large search space made up of all possible solutions. The resultant mathematical optimization systems derived from realistic combinatorial TRN problems are usually NP-hard, which refers to problems for which the number of elementary numerical operations is not likely to be expressed or bounded by a function of polynomial form (Garey and Johnson 1979). For this reason, existing mathematical optimization solution approaches to TRN problems are usually applied to relatively small and idealized networks for small urban areas or medium-sized urban areas with coarse networks. The route network structures may also be limited to certain particular configurations.

Solution Methodology

Methodology was developed based on the following considerations: (a) the method should be generally applicable to the design and optimization of a wide range of TRN problems in practice; (b) the solution method should be as generic as possible and should not favor particular transit network configurations; and (c) solutions obtained from this method should give fairly good results in a reasonable amount of time, as permitted by the current computer power affordable to most transit agencies. Reliability of results should improve as the computer resource or power increases, and should approach the global optimum when there is no computer resource limitation.

Representation of Transit Service Area, Routes, and Route Network

A transit service area is represented by a street network, which consists of a set of street nodes that are connected to each other by a set of street segments. A street segment, $\boldsymbol{a}(n_1, n_2)$, may be defined by its two end nodes n_1 and n_2. In a directed network, segments $\boldsymbol{a}(n_1, n_2)$ and $\boldsymbol{a}(n_2, n_1)$ may be different as in the case of one-way streets or when travel impedance on the same link is different in the two opposite directions. In this study, only undirected network is considered (i.e., $\boldsymbol{a}(n_1, n_2)$ and $\boldsymbol{a}(n_2, n_1)$ are considered the same), but the methodology can be easily extended to directed networks. It is also assumed that the street network is connected; thus, any two nodes in the street network are connected by at lease one path.

The following is the mathematical representation of a street network. Denote $\boldsymbol{N}^{(n)} = \boldsymbol{N}^{(n)}\{n_1, n_2, \ldots, n_n\}$ as the set of n street nodes in a transit service area, then a street network consisting of m street segments may be written as $\boldsymbol{A}^{(m)} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_m\}$,

where $\boldsymbol{a}_i = \boldsymbol{a}_i(n_{i1}, n_{i2})$ and $n_{i1}, n_{i2} \in \boldsymbol{N}^{(n)}$ (i = 1, 2, …, m). A path/route between any two nodes is defined as a sequence of non-reoccurring nodes, or $\boldsymbol{p} = \boldsymbol{p}(n_1, n_2, \ldots, n_k)$, and there is one street segment, i.e., $\mathbf{a}(n_j, n_{j+1}) \in \boldsymbol{A}^{(m)}$ (j = 1, 2, …, k-1), that connects any two neighboring nodes. A street network may also be represented through an adjacency list of street nodes. For a given node, called the *master* node of the list, its associated nodal adjacency list consists of all the neighboring nodes that can be connected to the master node with one street segment. The set of all nodal adjacency lists of a street network may be expressed as

$$\boldsymbol{L}(k) = \boldsymbol{L}(k)\{k_1, k_2, \ldots, k_{m(k)}\},\ k=1, 2,\ldots, n \qquad (4)$$

where:

$\boldsymbol{L}(k)$ is the nodal adjacency list of the street node k

k_j is the street node number of the j^{th} neighboring node in the list

$m(k)$ is the number of nodes in the list

The TRN $\boldsymbol{T}^{(l)}$ in (2) may also be expressed as a TRN matrix.

$$\boldsymbol{T} = \boldsymbol{T}[t_{ij}],\ t_{ij} = \begin{cases} 1, \text{ if node } j \text{ is on route } i, & i=1,2,\ldots,l \\ 0, \text{ if node } j \text{ is not on route } i, & j=1,2,\ldots,n \end{cases} \qquad (5)$$

In this study, for the purpose of representation uniqueness, it is assumed that the transit route stop set and the corresponding street node subset are the same.

Representation of Search Spaces for Transit Routes and Route Network

The solution search spaces in this study are locally and iteratively defined, and the size of a local search space may be flexible based on available computing resources. A local path space consists of three components: (1) a master path; (2) a key-node representation of the master path; and (3) a set of paths that are in the neighborhood of the master path. A master path is a path from which a local path space will be generated. Key nodes are a set of nodes on a master path selected to defined paths in the local path space. A local path space is derived from the local node spaces of the key nodes on the master path. An i^{th} order local node space, denoted as $\boldsymbol{N}_{(i)}(k)$, of a master node k is defined as the set of nodes that can be connected to the master node with i or fewer street segments. The order of a local node space provides a measurement of the degree of localization. Figure 1 illustrates a three-key-node (nodes n_1, n_2, and n_3) representation of a master path (solid line) and the

three first order local node spaces, $\boldsymbol{N}_{(1)}(n_1) = \{n_1, n_{11}, n_{12}, n_{13}, n_{14}\}$, $\boldsymbol{N}_{(1)}(n_2) = \{n_2, n_{21}, n_{22}, n_{23}, n_{24}\}$, and $\boldsymbol{N}_{(1)}(n_3) = \{n_3, n_{31}, n_{32}, n_{33}, n_{34}\}$.

Denote the $(i\text{-}1)^{th}$ order local node space of a master node k as $\boldsymbol{N}_{(i-1)}(k) = \bullet\{k_1, k_2, \ldots, k_{q(k)}\}$, where $q(k)$ is the number of nodes in this local node space, then

$$\boldsymbol{N}_{(i)}(k) = \{k_1, k_2, \ldots, k_{q(k)}\} \subset \boldsymbol{L}(k_1) \subset \boldsymbol{L}(k_2) \subset \ldots \subset \boldsymbol{L}(k_{q(k)}) \qquad (6)$$

where $\boldsymbol{L}(k_j)$ is the nodal adjacency list of node k_j. A local node space is a subspace of the street node space $\boldsymbol{N}_{(i)}(k) \in \boldsymbol{N}^{(n)}$. As the order i increases, it will approach to the original street node space $\boldsymbol{N}^{(n)}$. The procedure to generate a local path space from a master path has three steps: (1) Select s key-nodes from the node set of the master path $\boldsymbol{p} = \boldsymbol{p}(n_1, n_2, \ldots, n_r)$, i.e., $\{m_1, m_2, \ldots, m_s\} \in \{n_1, n_2, \ldots, n_r\}$; (2) Generate a sequence of local node spaces from these key-nodes, $\left(\boldsymbol{N}_{(i)}(m_1), \boldsymbol{N}_{(i)}(m_2), \ldots, \boldsymbol{N}_{(i)}(m_s)\right)$; and (3) Define the local path space as the set of paths consisting of piecewise shortest path segments that start from nodes in the first local node space $\boldsymbol{N}_{(i)}(m_1)$, sequentially pass the nodes in each of the intermediate local node space $\boldsymbol{N}_{(i)}(m_j)$ (j = 2, 3, …, s-1), and end at nodes in the last local node space $\boldsymbol{N}_{(i)}(m_s)$. The shortest path segments used to connect nodes in neighboring local node spaces are from a k-level shortest path space $\boldsymbol{P}_S^{(k)}$ that consists of all the first k shortest paths between any two nodes in the street node space $\boldsymbol{N}^{(n)}$. (References on algorithms of finding a k-level shortest path space may be found in Zhao 2003.) The resultant path space, denoted as, $\boldsymbol{P}_{(i)}^{(k)}\left(p^{(s)}\right)$, will be referred to as the local path space based on the s-nodes representation of the master path $\boldsymbol{p}$, or simply the local path space of path $\boldsymbol{p}$.

The local network search spaces of a transit network $\boldsymbol{T}^{(l)} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_l\}$ is defined as

$$\boldsymbol{T}_{(i)}^{(k)}\left(\boldsymbol{T}^{(l)}\right)=\boldsymbol{T}_{(i)}^{(k)}\left\{\boldsymbol{P}_{(i)}^{(k)}\left(\boldsymbol{r}_1^{(s_1)}\right), \boldsymbol{P}_{(i)}^{(k)}\left(\boldsymbol{r}_2^{(s_2)}\right), \cdots, \boldsymbol{P}_{(i)}^{(k)}\left(\boldsymbol{r}_l^{(s_l)}\right)\right\} \tag{7}$$

where $\boldsymbol{P}_{(i)}^{(k)}\left(\boldsymbol{r}_j^{(s_j)}\right)$ is a local path space of s_j-node representation for master path $\boldsymbol{r}_j$. It may be seen that as the two numbers *i* and *k* increase, a local path space of any master path will approach to the combinatorial path space $\boldsymbol{P}_N$.

$\boldsymbol{P}_{(i)}^{(k)}\left(\boldsymbol{r}_j^{(s_j)}\right)$ will be the path search space of the corresponding transit route $\boldsymbol{r}_j$. In general, routes derived from smaller numbers of key nodes will result in better route directness and smaller local path search space, but their flexibility is also limited. Routes with larger numbers of key nodes are relatively more flexible to reach more neighboring nodes, thus may cover more trips. However, this will also result in larger local path search spaces, requiring more computing resources.

Integer Constraints for Transit Route Network

Integer constraints in this study include the following: (a) fixed route constraints prescribing fixed guideway lines or bus routes that are specified by transit planners to meet certain planning goals, which will remain unchanged during the optimization process; (b) constraints prescribing starting, ending, or in-between areas through which transit routes must pass, which may include major activity centers or transfer points; (c) route length constraints for individual transit lines or for the entire system; and (d) constraints on the number of transit stops on individual routes.

Route Directness Constraints

Route directness used in this study is defined as follows:

$$d\left(\boldsymbol{r}\right)=\sum_{i=1}^{s-1}\sum_{j=i+1}^{s} w_{ij}\left(d_{ij}^{(r)} / d_{ij}^{(S)}\right), \tag{8}$$

where:

- s is the number of nodes on route $\boldsymbol{r} = \boldsymbol{r}(n_1, n_2, \ldots, n_s)$
- $d_{ij}^{(r)}$ is the distance between nodes n_i and n_j measured along the transit route

$d_{ij}^{(S)}$ is the shortest network distance between nodes n_i and n_j

w_{ij} are weighting factors

For geometry based route directness, $w_{ij} = w_{ij}^{G} \equiv 2/(s^2 - s)$, and for ridership based route directness,

$$w_{ij} = w_{ij}^{R} \equiv \left(o_{ij} + o_{ji}\right) \Big/ \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \left(o_{ij} + o_{ji}\right)$$

where o_{ij} and o_{ji} are coefficients of the OD matrix. The geometry based route directness, denoted as $d^G(\boldsymbol{r})$, reflects the average ratio of the two travel distances, $d_{ij}^{(r)}$ and $d_{ij}^{(S)}$, between each node pair on route ***r***. A value of $d^G(\boldsymbol{r}) = 1$ indicates that, on average, transit vehicles on route ***r*** travel along the shortest paths between route stops. The ridership based route directness, $d^R(\boldsymbol{r})$, represents the average ratio of the distance a person travels between OD points along transit route ***r*** to the distance traveled along the shortest path. A value of $d^R(\boldsymbol{r}) = 1$ indicates that, on average, passengers on transit route ***r*** travel along the shortest paths between OD points. Route directness constraints used in this study may be expressed as $d^G(\mathbf{r}_i) \le d_r^G$ or $d^R(\mathbf{r}_i) \le d_r^R$ (i =1, 2,...,I), where d_r^G and d_r^R are the two travel directness constraint parameters. In general, smaller d_r^G and d_r^R imply better services, but may result in higher transit operating cost. Large d_r^G and d_r^R mean that some potential transit riders may be turned away and that existing transit riders may be forced to look for other alternatives, thus leading to loss of ridership and, eventually, higher operation cost.

Network Directness Constraints

Transit network directness has a physical meaning similar to that of the route directness, except that the directness measurement is based on geometry or ridership characteristics of the entire route network, instead of individual transit routes.

Out-of-Direction (OOD) Constraints

The OOD constraint used in this study is derived from the formulation given by Welch et al. (1991). Denote $d_{ij}^{(O)}(\boldsymbol{r})$ as the OOD impact index for travel between nodes i and j on transit route ***r***, then

$$d_{ij}^{(O)}(\boldsymbol{r}) = r_{ij}^{(1)}(\boldsymbol{r})[l_{ij}(\boldsymbol{r}) - d_{ij}]/r_{ij}^{(2)}(\boldsymbol{r})$$

where:

$r_{ij}^{(1)}(\boldsymbol{r})$ is the *through* ridership, or the number of trips on route $\boldsymbol{r}$ that pass through nodes i and j without boarding or alighting in between

$r_{ij}^{(2)}(\boldsymbol{r})$ is the OOD ridership, which is the number of trips on route $\boldsymbol{r}$ that involve either boarding or alighting or both at nodes between nodes i and j

$l_{ij}(\boldsymbol{r})$ is the distance between nodes i and j along route $\boldsymbol{r}$

d_{ij} is the distance along the shortest path between these two nodes in the street network

$d_{ij}^{(O)}(\boldsymbol{r})$ represents the extra travel distance that incurs to each *through* passenger in order to serve an OOD passenger.

Optimization Objective Functions

Objective functions considered in this study are various trip coverage functions or their combinations. The goal is to obtain a TRN structure with minimum transfers, while optimizing service coverage. If a trip between an OD pair requires no transfers, the trip is called a zero-transfer trip, while a trip between an OD pair that requires k or fewer transfers will be called a k-or-less transfer trip. A k-or-less transfer trip coverage function, or simply a k-or-less transfer function, is defined as the total number of OD trips that can be accomplished with k or fewer transfers in a transit network service area. The following is a description of various transfer coverage functions used in this study. Denote f_k as a k-or-less transfer function, then

$$f_k = f_k(\boldsymbol{T}, \boldsymbol{O}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\left(o_{ij} + o_{ji} \right) h\left(a_{ij}^{(k)} \right) \right], \; k = 0, 1, 2, \qquad (9)$$

where

$\boldsymbol{T}$ is the TRN matrix defined in (5)

$\boldsymbol{O}$ is the OD matrix

h is a step function that has the property:

$h(x) = 1$ for $x > 0$, and $h(x) = 0$ for $x \leq 0$

Coefficients $\alpha_{ij}^{(k)}$ in (9) are defined as $\alpha_{ij}^{(0)} = \sum_{k=1}^{I} t_{ki}\, t_{kj}$, $\alpha_{ij}^{(1)} = \sum_{k=1}^{I}\sum_{m=k}^{I}\left(t_{ki}\, t_{mj}\, \alpha_{km}\right)$, $\alpha_{ij}^{(2)} = \sum_{k=1}^{I}\sum_{m=k}^{I}\left(t_{ki}\, t_{mj}\, \beta_{km}\right)$, $\alpha_{km} = \sum_{j=1}^{n} t_{kj}\, t_{mj}$, and $\boldsymbol{\beta}_{km} = \sum_{i=1}^{I} \alpha_{ki}\, \alpha_{mi}$, where t_{ki}, t_{kj}, and t_{mj} are coefficients of matrix $\boldsymbol{T}$. It may be seen that calculation of transfer objective function, $f_{2,}$ is computational intensive, compared with functions f_0 and $f_{1,}$ due to the great number of arithmetic operations involved to obtain all the required coefficients.

The use of any of the transfer functions alone as the objective function may result in the optimization of one TRN parameter at the cost of others. The following are two objective functions that combine multiple coverage functions, thus giving more balanced results.

$$t_1(\boldsymbol{T},\boldsymbol{O}) = [f_0 + 2(f_1 - f_0) + \alpha(f_T - f_1)]/f_T \tag{10}$$

$$t_2(\boldsymbol{T},\boldsymbol{O}) = [f_0 + 2(f_1 - f_0) + 3(f_2 - f_1) + \alpha(f_T - f_2)]/f_T, \tag{11}$$

where

- α is a weighting coefficient to penalize uncovered trips during the optimization process of the TRN system
- f_T is the total number of trips in the transit network service area

The physical meaning of the objective function t_2 is the average number of vehicle boardings that a transit rider has to make to accomplish an OD trip. The optimal value of t_2 is 1.0, indicating that all trips are zero-transfer trips. Uncovered trips ($f_T - f_2$) are penalized by α. The value of α needs to be determined by transit planners. For example, by setting α = 4, each of the uncovered trips is considered as four vehicle boardings. In general, the larger the value of α, the greater relative importance is given to service coverage. The physical meaning of t_1 is similar to that of t_2.

Algorithm 1—Basic Greedy Search Method

The basic assumption of Basic Greedy Search (BGS) is that the demand distribution in a TRN service area has certain continuity. In other words, nodes with certain transit demands are probably close to nodes with similar demands. In such cases, it will be more effective in searching for a better solution by evaluating paths that are near nodes or areas with higher trip distributions. (Detailed description of various search algorithms used in this study can be found in Zhao 2003.) Assume

that during a solution search process, an intermediate TRN result $\boldsymbol{T}^{(l)} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots \boldsymbol{r}_l\}$ has been obtained. The solution search procedure for the next stage of BGS method involves the following steps:

1. Select key nodes from route $\boldsymbol{r}_j$. For illustration, assume the three-node representation of route $\boldsymbol{r}_j$ is used (see Figure 1), which is denoted as $\boldsymbol{r}_j^{(3)}$.
2. From the three key nodes n_1, n_2, and n_3, generate three first order local node spaces

   $\boldsymbol{N}_{(1)}(n_1) = \{n_1, n_{11}, n_{12}, n_{13}, n_{14}\}$

   $\boldsymbol{N}_{(1)}(n_2) = \{n_2, n_{21}, n_{22}, n_{23}, n_{24}\}$

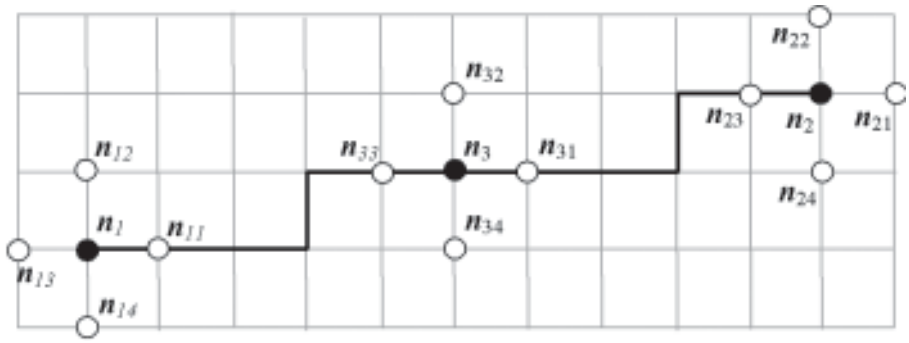
   $\boldsymbol{N}_{(1)}(n_3) = \{n_3, n_{31}, n_{32}, n_{33}, n_{34}\}$

   There are five nodes in each of the three local node spaces.
3. Connect nodes in node spaces $\boldsymbol{N}_{(1)}(n_1)$ and $\boldsymbol{N}_{(1)}(n_3)$ with the shortest paths in space $\boldsymbol{P}_S^{(1)}$, to obtain 5 × 5 = 25 shortest path segments. These shortest path segments are then extended with shortest paths to nodes in node space $\boldsymbol{N}_{(1)}(n_2)$ to obtain 25 × 5 = 125 paths. These 125 paths form the local path space of route $\boldsymbol{r}_j$, based on three-node representation $\boldsymbol{P}_{(1)}^{(1)}\left(\boldsymbol{r}_j^{(3)}\right)$.
4. Replace route $\boldsymbol{r}_j$ in the existing TRN $\boldsymbol{T}^{(l)}$ with a path $\boldsymbol{r}_{jk} 0 \boldsymbol{P}_{(1)}^{(1)}\left(\boldsymbol{r}_j^{(3)}\right)$ to obtain $\boldsymbol{T}^{(l)} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{j-1}, \boldsymbol{r}_{jk}, \boldsymbol{r}_{j+1}, \ldots, \boldsymbol{r}_l\}$, and perform function evaluation for k = 1, 2, …, 125. If a better result is obtained, replace $\boldsymbol{r}_j$ with $\boldsymbol{r}_{jk}$, and go to Step 1 to start a new search. If no better result is found from all the 125 paths $\boldsymbol{r}_{jk} 0 \boldsymbol{P}_{(1)}^{(1)}\left(\boldsymbol{r}_j^{(3)}\right)$, go to Step 5.
5. Select the next route from the transit route network, e.g., route $\boldsymbol{r}_{(j+1)}$ and go to Step 1 to start a new local search for route $\boldsymbol{r}_{(j+1)}$.
6. The search process will be considered converged if no better results can be found from the local path search spaces of all the individual routes.

Algorithm 2—Fast Hill Climb Search Method

Conceptually, the Fast Hill Climb (FHC) method is similar to the deepest decent method in continuous research fields. First, *l* new solutions are formed by replacing one route at a time in the network, with the best route from its local search space. These *l* best routes from the local search spaces also make up a new solu-

Figure 1. Three-Key-Node Representation of Transit Route



tion. These *l* + 1 solutions are compared and the best one is chosen as the current solution. Note that the computation process to obtain the *l* best routes is independent to each other, making it suitable for parallel computing.

Numerical Experiments

The first test experiment was based on a real network in Switzerland (Mandl 1979). This problem was also used by Shih and Mahmassani (1994) and Baaj and Mahmassani (1991) as a benchmark problem to test their approaches to TRN and TNS design optimization. Mandl problem consisted of a street network of 15 nodes with a total demand of 15,570 trips per day. For this particular problem, the length of a street segment was defined in terms of in-vehicle travel time in minutes. In Table 2, the first row identifies the source of the solutions to the benchmark problem. The second row indicates solutions to the benchmark problem with different numbers of routes, total route length, and/or search methods. The methods used to obtain the results are indicated in the third row. For each solution, the unshaded column provides the statistics for the layout produced in the original studies, and the shaded column gives the statistics for the results produced from the FHC method developed in this study.

It may seem that the percentages of zero transfer trips were higher for all solutions produced in this study. Except for Mandl's original results, all solutions provided 100% trip coverage with zero or one (one-or-less) transfer involved in each trip.

The second experiment involved a large-scale TRN optimization problem based on the service area of the Miami-Dade Transit Agency (MDTA), encompassing a

Table 2. Comparison of Results from Different Methods

Problem Source	Mandl[1]		Baaj & Mahmassani						Shih & Mahmassani			
Route layout case	1		1		2		3		1		2	
Search Method	Mandl	FHC[2]	B&M[3]	FHC	B&M	FHC	B&M	FHC	S&M[4]	FHC	S&M	FHC
0-transfer trips (%)	69.94	76.43	78.61	82.34	79.96	86.64	80.99	82.98	82.59	84.84	87.73	91.78
1-transfer trips (%)	29.93	23.57	21.39	17.66	20.04	13.36	19.01	17.02	17.41	15.16	12.27	8.22
2-transfer trips (%)	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total route length	82	82	126	125	144	144	106	105	124	124	151	152
Number of routes	4	4	6	6	8	8	7	7	6	6	8	8
Average Transfers	1.30	1.24	1.21	1.18	1.20	1.13	1.19	1.17	1.17	1.15	1.12	1.08

[1]Mandl's method
[2]Fast Hill Climb method
[3]Baaj and Mahmassani's methods
[4]Shih and Mahmassani's methods

region of about 300 square miles with a population of about 2.3 million. MDTA is ranked the 16th largest transit agency in the United States. At the time of this research, MDTA operated 83 transit routes, including a rail rapid transit system of 22.5 route miles (Metrorail), a 4.5-mile downtown automated circulation system (Metromover), and 81 bus routes with about 4,500 transit stops. The street network used in this experiment consisted of 4,300 street segments and 2,804 street nodes. In the optimization process, Metrorail and Metromover alignments were fixed and the longest and shortest bus routes were about 32 miles and 4 miles, respectively. The total length of the transit system was about 1,300 route miles, omitting some small loops at the ends of some routes or in shopping centers. The OD matrix was generated from the 1999 validated Miami-Dade travel demand model, which provided the daily number of passenger trips between each pair of traffic analysis zone centroids. These were manually distributed to the surrounding street network nodes with considerations given to land use patterns and street network connectivity. The total demand was 161,944 daily transit trips. All the numerical results were obtained on a personal computer with a 2.8GHz CPU and 1GB RAM memory. Table 3 presents the results from the BGS and FHC methods. There were two sets of results produced by each method, one based on an initial guess network that was the existing route network and the other based on a program generated initial guess network. The constraints were that the total

route length of the network should not exceed that of the existing system by more than 10%, and that the total number of transit lines remained the same as the existing system. Two objective functions were used, one maximizing zero-transfer trips (f_0) and the other maximizing one-or-less transfer trips (f_1). The values of the objective functions are given in the shaded cells.

Compared to the existing network, the FHC method with objective function f_0 gave the best zero-transfer trip coverage, with an improvement of 85% (from14.28% to 26.41%), while the BGS search method yielded an improvement of 84%. For objective function f_1, the FHC method again gave the best one-or-less transfer trip coverage, with a 48% improvement (from 55.13% to 81.57%). These improvements were achieved with a small increase of 5% in total network route mileage. Assuming most transit riders may be only willing to transfer once per trip (Stern 1996), the one-or-less trip coverage shown in the fourth row would be the actual total trip coverage of the corresponding route networks. The remaining trip demand either required two or more transfers or were not satisfied.

Table 3. Comparison of Results with the Existing Network

Network Parameters	Existing Results	Zero transfer objective function f_0				One-or-less transfer objective function f_1			
Solution Method	-	BGS		FHC		BGS		FHC	
Initial Guess Network	-	Exist[1]	Prog[2]	Exist	Prog	Exist	Prog	Exist	Prog
0-transfer trips (%)	14.28	22.25	26.29	23.43	26.41	18.01	20.14	18.9	19.47
1-or-less transfer trips (%)	55.13	69.71	72.71	71.37	73.77	77.86	80.74	80.45	81.57
2-or-less transfer trips (%)	65.2	77.69	74.91	77.75	76.25	86.59	83.58	87.36	84.23
Total covered trips (%)	65.66	77.86	74.92	77.89	76.25	86.72	83.58	87.48	84.24
Total route mileage	1,278	1,330	1,340	1,340	1,348	1,304	1,341	1,345	1,346
Trips per route mile	83	95	91	94	92	108	101	105	101
Average Transfers	1.94	1.82	1.68	1.78	1.69	1.89	1.79	1.86	1.78
CPU Time (hours)	-	0.39	0.39	0.79	0.65	3.79	4.68	15.76	12.93

[1]Existing network as initial guess network
[2]Program generated initial guess network

The number of covered trips per route mile shown in Table 3 was defined as

$$R_{T/L}^{(2)} = f_2 / l_T$$

where

f_2 was the number of trips accomplished with two-or-less transfers

l_T was the total length of the TRN

As a network efficiency indicator, the best $R_{T/L}^{(2)}$ value was given by the BGS method with f_1 as the objective function. The average transfers were defined as $[f_0 + 2(f_1 - f_0) + 3(f_2 - f_1)]/f_2$, which was the average number of boardings per transit rider who could complete a trip with two or fewer transfers. For the same objective function, the FHC method produced slightly better results than the BGS method. It may be seen that the differences in results produced by the BGS and FHC search methods were insignificant, but the BGS method was significantly faster than the FHC method.

Table 4 presents results obtained from composite trip coverage functions t_1 and t_2 described in (10) and (11), with the shaded cells indicating the objective function values. The penalty a was set at 4 in both functions t_1 and t_2. It may be seen that improvements in various trip coverage functions were consistent instead of being achieved at the cost of each other, as in the case of single trip coverage function shown in Table 2. Overall, FHC produced slightly better results than those from method BGS, but at a higher computational cost.

Conclusion

The methodology developed from this work has a systematic mathematical statement of TRN problems, including the definition of various objective functions, solution search spaces, and constraint conditions commonly used in transit planning fields, and a systematic scheme that flexibly defines solution search spaces based on available computing resources and/or optimization problem sizes. Two local search schemes have been developed to obtain results for large-scale practical problems in a reasonable amount of time.

The feasibility of the proposed method has been tested through practical TRN optimization problems of realistic sizes. Numerical results showed that the methodology developed in this work was capable of tackling large-scale transit network design optimization problems. Further improvements may include development

Table 4. Comparison of Results with Existing Networks

Network Parameters	Existing Networks	Average number of boardings function *t1*		Average number of boardings function *t2*	
		BGF	FHC	BGF	FHC
Objective Function *t1*	2.76	2.14	2.13	2.19	2.15
Objective Function *t2*	2.65	2.12	2.1	2.15	2.11
0-transfer trips (%)	14.28	24.22	24.55	22.4	24.88
1-or-less transfer trips (%)	55.13	80.7	81.31	79.45	80.23
2-or-less transfer trips (%)	65.2	83.54	84.37	83.67	83.61
Total covered trips (%)	65.66	83.54	84.38	83.68	83.62
Total route mileage	1,278	1,350	1,358	1,336	1,377
Trips per route mile	83	100	101	102	98
Average Transfers	1.94	1.74	1.76	1.79	1.74
CPU Time (hours)	-	7.62	9.85	90.41	114.59

of TRN optimization methods that consider dynamic transit demand, demand and travel time in different time period of a day, and waiting and transfer penalties.

Acknowledgements

This research was sponsored by a grant from the Florida Department of Transportation (FDOT). The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of the FDOT. The authors would like to thank the reviewers for their useful comments and suggestions.

Endnote

[1]Depending on particular applications, length/distance may refer to either geometric length/distance or travel time.

References

Baaj, M. H and H. S. Mahmassani. 1991. An AI-based approach for transit route system planning and design. *Journal of Advanced Transportation* 25(2): 187-210.

Bertsekas, D. P. 1998. *Network optimization: continuous and discrete models*. Belmont, MA: Athena Scientific.

Bookbinder, H. J. and A. Désilets. 1992. Transfer optimization in a transit network.

Transportation Science (26)2: 106-118.

Chien, S., B., V. Dimitrijevic, and L. N. Spasovic. 2003. Optimization of bus route planning in urban commuter networks. *Journal of Public Transportation* 6(1): 53-80.

Garey, M. R. and D. S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness.* Freeman, NY.

Mandl, C. E. 1979. Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research* (5): 396-404.

Shih, M. C. and H. S. Mahmassani. 1994. *A design methodology for bus transit networks with coordinated operations*. SWUTC/94/60016-1. Austin, TX: Center for Transportation, University of Texas at Austin.

Soehodo, S. and M. Koshi. 1999. Design of public transit network in urban area with elastic demand. *Journal of Advanced Transportation* 33(3): 335-369.

Stern, R. 1996. Passenger transfer system review. *Synthesis of Transit Practice* 19, Transportation Research Board.

Welch, W., R. Chisholm, D. Schumacher, and S. Mundle. 1991. Methodology for evaluating out-of-direction bus route segments. *Transportation Research Record* 1308: 43-50.

Zhao, F. 2003. *Optimization of transit network to minimize transfers.* Draft technical report prepared for the Florida Department of Transportation. Miami: Lehman Center for Transportation Research, Florida International University.

About the Authors

Fang Zhao (*zhaof@fiu.edu*) received her Ph.D. in civil engineering from Carnegie Mellon University. She teaches in the civil and environmental engineering department at Florida International University, and is associate professor and deputy director of the Lehman Center for Transportation Research. Her main research interests include public transportation, travel demand modeling, geographic information systems, and transportation data modeling. Dr. Zhao is a registered professional engineer in the state of Florida and member of the Transportation Research Board's Transit Planning and Development Committee and New Technologies and Systems Committee.

Ike Ubaka (*ike.ubaka@dot.state.fl.us*) is program manager for transit systems planning at the Florida Department of Transportation, where he is responsible for statewide transit systems planning. He has over 18 years of experience in transit planning, including service with the Los Angeles County Metropolitan Transportation Authority and the Orange County (California) Transportation Authority. Mr. Ubaka holds a master's degree in urban and regional planning from UCLA and a master of administration degree from the University of California, Riverside. He is a member of the American Institute of Certified Planners.

Vehicle Selection for BRT: Issues and Options

Samuel L. Zimmerman, DMJM+HARRIS
Herbert Levinson, Transportation Consultant

Abstract

Bus rapid transit (BRT) is a flexible, high performance rapid transit mode that combines facilities, equipment, service and intelligent transportation system (ITS) elements into a permanently integrated system with a quality image and unique identity. Vehicles are an extremely important component of BRT systems, because they not only contribute significantly to BRT's image and identity, but also play a strong role in achieving measurable performance success.

Vehicle-related planning and design issues confront BRT planners in seven basic areas:

1. *Capacity, External Dimensions*
2. *Internal Layout*
3. *Doors*
4. *Floor Height*
5. *Propulsion Systems*
6. *Vehicle Guidance*
7. *Aesthetics, Identity and Branding*

This paper draws heavily on 26 case studies documented in TCRP Project A-23 (Levinson, Zimmerman, et al. 2003). It also reflects experience from newer BRT

systems and concludes with a series of general principles and guidelines for vehicle design, selection, and use in BRT applications.

Introduction

BRT is a flexible, high performance, rapid transit mode that combines facility, equipment, service and ITS elements into a permanently integrated system with a quality image and unique identity. Its constituent elements include:

1. Running ways
2. Stops, stations and terminals
3. Vehicles
4. Services
5. Intelligent transportation systems
6. Fare collection

BRT must be planned as an integrated system ideally suited for the markets served and the application's physical environment. Having a quality image and a unique identity distinct from the rest of the transit (i.e., local bus) system are also important BRT attributes.

Vehicles may be the most important element to user and non-user perceptions of a BRT system's quality. Vehicles also play a strong role in determining real performance in terms of speed, reliability, and cost. They are critical from the perspective of customers, the community as a whole, and the operating entity for a number of reasons. First, vehicles have a strong effect on every aspect of measurable system performance.

- Propulsion systems impact revenue service times (thus, ridership and revenue), emissions (air pollutant and noise) and operating and maintenance (OM) costs.
- Seating, floor height, floor plan, and door configurations impact stop dwell times, hence, revenue service times and reliability.
- Physical size, aisle width, number of doors and their width and position, and seating numbers and configuration are important determinants of BRT system capacity.

Second, since potential new transit customers as well as existing ones are exposed to BRT vehicles, their design impacts community and customer perceptions of the quality of the entire system. This perception is primarily visual and aesthetic, but it also relates to impacts such as noise and air emissions. Although not as important as time and cost in effecting mode choice, image and brand influence the willingness of new customers to try BRT. This willingness to ride transit translates into additional ridership, revenue and other related benefits, as do performance factors such as travel time and reliability.

One of the major products of TCRP Project A-23 (Levinson, Zimmerman, et al. 2003) was the documentation of 26 case studies of BRT systems around the world and the results of their assessments into a number of summary observations. The synthesis showed that the proliferation of BRT systems has accelerated the trend toward more rubber-tired transit vehicle specialization, away from the one-size-fits-all (i.e., 40-foot [12 meter]) bus to perform all surface transit functions. More attention is being paid to the nature of the markets being served, service offered, and customer and non-customer perception of vehicle quality.

The discussion below provides planners with information that can help them make better vehicle choices. It summarizes observations relating to BRT vehicles from the case studies and other, more recent BRT applications, as well as the TCRP BRT guidelines. It is organized around seven basic themes:

1. Capacity, External Dimensions
2. Internal Layout
3. Doors
4. Floor Height
5. Propulsion Systems
6. Guidance
7. Aesthetics, Identity and Branding

Capacity and External Dimensions

In nearly all of the 26 case studies, demand was heavy, ranging up to 20,000 or more passengers per hour. That utilization of high capacity (e.g., articulated buses) vehicles with a total capacity (standing + seated) of at least 65 places was essential for either system capacity and/or OM cost reasons. In the case of Los Angeles

MTA's MetroRapidBus and Boston MBTA's Silver Line, BRT services were initiated with 40-foot (12 meter) vehicles, because long procurement times for larger (60-foot [18 meter]) vehicles would have delayed the start of service. Both the LA and Boston systems planned to have 60-foot vehicles. Eaarly on, demand had nearly outstripped the capacity of the 40-foot vehicles.

Less than one year after opening, some of the originally planned 60-foot (18 meter), low-floor (Neoplan CNG) vehicles are in operation on the Silver Line. In LA, 60-foot low-floor (NABI CNG) vehicles are on order after approximately three years of operation. Several BRT applications in South America and Europe, such as Curitiba and Sao Paulo, Brazil; Nancy, Nice and Caen, France; and Utrecht, Netherlands, operate double articulated vehicles of up to 83 feet (25 meters) in length, having a capacity of over 120 total places (at North American loading standards).

Given the high demand nature of many BRT routes and services, transit operators are increasingly using large (over 40-foot) vehicles. The use of unusually large (for the given community) rolling stock adds to BRT's distinct identity, while the extra capacity is helpful for financial, service, and operational efficiencies.

Table 1 shows the external dimensions and capacities (computed for a standee density of 3 standees per square meter) for typical vehicles used in BRT applications.

Table 1. Dimensions and Capacities of Typical U.S. and Canadian BRT Vehicles

Length (Feet)	**Width**	**Floor Height (Inches)**	**No. of Door Channels**	**No. of Seats** (including seats in wheelchair tiedown areas)	**Maximum Capacity*** seated plus standing
40 (12.2 m)	96-102 (2.45-2.6m)	13-36 (33-92 Cm)	2-5	35-44	50-60
45 (13.8 m)	96-102 (2.45-2.6m)	13-36 (33-92 Cm)	2-5	35-52	60-70
60 (18 m)	98-102 (2.5-2.6m)	13-36 (33-92 Cm)	4-7	31-65	80-90
80 (24 m)	98-102 (2.5-2.6m)	13-36 (33-92 Cm)	7-9	40-70	110-130

* Computed at Standee density of 3/mtr^2

Interior Configuration

The interior configuration of BRT vehicles influences both passenger capacity and comfort. As noted, the overall capacity of transit systems is influenced by a number of vehicle-related factors, and the interior configuration is one of the more important factors. Easy and rapid passenger boarding, alighting and internal circulation can minimize dwell times. BRT vehicle interior layouts usually include large standing/circulation areas around doors. These aid boarding, alighting, and circulation and can also function as storage areas for baby carriages, bicycles, and wheelchairs, explicitly supporting the mobility needs of the entire community.

Aisle width also influences vehicle capacity. Most conventional low-floor vehicles, even those with a step-up to the rear portion of the vehicle, have a minimum aisle width between the rear wheel wells (second and third axle on articulated vehicles) of about 24 inches (60 cm). The constraint on aisle width here is the need to accommodate tires and mechanical components; however, some specialized BRT vehicles have independently-suspended single, extra-wide, extra-strength tires with electric motor and gearboxes inside. This allows a wider aisle (maximum width of about 34 inches (87 cm), permitting easier in-vehicle circulation, lower passenger service and stop dwell times. Irrespective of the running gear utilized, where there is 2+2 perpendicular seating, the required width of seat banks and the wall of the vehicles will constrain aisle widths to no greater than approximately 24 inches (60 cm).

In rapid turnover markets with relatively short trip lengths (e.g., various European applications, Las Vegas Blvd., Denver Mall), planners have elected to maximize capacity and ease of circulation rather than maximizing the number of seats. Because many transit operators have policies that no customer should have to stand in excess of 20-30 minutes, for longer average trip length markets (e.g., suburb to urban corridors like Pittsburgh's busways and Ottawa's Transitways), interiors are usually configured to maximize seating.

The interior of the Irisbus Civis, used on the Rouen, France TEOR system (Figure 1), illustrates the trade-off between the number of seats, standee area, and aisle width when serving a dense urban corridor with significant passenger turnover.

Figure 1. Interior, Irisbus Civis Specialized BRT Vehicle, TEOR, Rouen, France

Doors

Number, Width

Irrespective of how fares are collected, a large number of wide doors will lower passenger service/stop dwell times. Wider doors provide lower friction than narrow doors and if wide enough, can support either multiple stream boarding or alighting, or simultaneous boarding and alighting. Multiple doors can also result in a better distribution of passengers within the vehicle, thus taking full advantage of available capacity. However, a given vehicle cannot have the maximum number of double stream doors (e.g., up to three on a 40-foot [12 meter] vehicle and up to four on a 60-foot [18 meter] vehicle) and still have the maximum number of seats, since seats are always tied to the outside wall of a vehicle.

A commonly used rule of thumb for the number of boarding and alighting channels in the U.S. is to have at least one channel per 10 feet of BRT vehicle length for typical radial, suburb - CBD corridors, assuming off-board fare collection. For dense corridors where significant, simultaneous boarding and alighting take place, an even larger number of passenger service streams in the same vehicle length may be warranted. For an express operation where virtually all customers alight in the AM peaks and board in the PM peaks at a limited number of all boarding or all alighting stops, fewer channels may be appropriate.

The Van Hool A300 60-foot (18 meter) articulated bus (Figure 2) operated by RTL from the south shore of the St. Lawrence River to Montreal, illustrates door number and placement for a conventional articulated bus used in a BRT-like service. Note the three double stream doors compared to the two narrower doors normally found on buses of the same size used for local service in the U.S.

Figure 2. Door Arrangement, Van Hool A300 60-Foot Low Floor Articulated Bus, RTL, Longueuil (Montreal), Quebec

Placement

The major objective affecting door placement is the need to ensure even passenger loading and unloading across the length of the respective vehicles. Accordingly, doors should be positioned to divide BRT vehicles into sections of roughly equal capacity and circulation distances. A number of recent BRT applications (e.g., Las Vegas and various European and South American systems) have an even distribution of doors and entry/exit streams across vehicle length.

Both conventional buses and specialized vehicles are also available with doors on either the left side (e.g., the Volvo and Mercedes vehicles in Bogota, Colombia and Curitiba, Brazil) or both sides. For years, trolley buses using the tunnel to access Harvard Square Station on the MBTA Red Line had doors on both sides. This is done to allow vehicles to use center platforms exclusively, as for the South American systems, or both the center and side platforms, as planned for a number of U.S. systems such as Cleveland.

Floor Height

BRT vehicles can have one of three basic floor heights: (1) 100% low floor; (2) partial low floor (usually about 70%); and (3) high floor. Low floors (or the low floor portion of partially low-floor vehicles) are typically 11-13 inches from the pavement, while high-floor vehicles are typically from 25 inches to as much as 35 inches above the pavement.

High-floor vehicles have an advantage in BRT applications where absolute maximum carrying and/or seated capacity is necessary, because little or no interior space is consumed by wheel wells, under-floor mechanical equipment, fuel tanks, etc. However, they may have inordinately high boarding and alighting times, unless used in conjunction with some way of assuring no-gap, no-step boarding and alighting. Rapidly deployed door bridges or door flaps have been used for this purpose in high volume BRT applications in South America (Quito, Curitiba, and Bogota). The major disadvantage of high-floor vehicles is that they can usually be used only at stations with high platforms, thereby limiting operating/service flexibility. This issue could be overcome, as has been done on some light rail transit (LRT) systems, by having no-step high platform boarding on one side of the vehicle and stairs to permit boarding from low platform stations on the other side.

One hundred percent low-floor vehicles have the great advantage of low boarding and alighting times and the ability to place a door behind the rear axle. How-

ever, 100% low floor designs also typically lose between 4 and 8 seats to wheel well intrusion, even where relatively small wheel and tire sizes are used. Another disadvantage of 100% low floor designs is that mechanical and electrical equipment and fuel tanks must either be stored inside the vehicle, where they take up space, or put on the roof, where they are difficult to service.

Low profile tires and minimum wheel travel of low floor vehicles may also contribute to poor ride quality. A final disadvantage of 100% low floor vehicles is the difficulty of packaging conventional mechanical drive trains consisting of an engine, hydraulic-mechanical transmission, connecting drive shafts, a differential, and an axle. One hundred percent low floor designs with this type of drive train can also lose up to four seats or the equivalent standing area merely due to the engine and drive train's intrusion into the vehicle (see Van Hool's A300 series of vehicles). The reason that many low floor specialized BRT vehicles have electric drive trains utilizing hub-electric motors and a single wheel on each side bogies with special wide, high-load limit tires is to avoid propulsion and suspension system packaging difficulties. These features contribute to acquisition cost, weight, and maintenance complexity.

Propulsion Systems

Low air and noise emission vehicles are extremely desirable for BRT, especially in situations where frequent services converge, such as near or in central business districts (like Pittsburgh, Miami, Brisbane, and Ottawa). With busway volumes often exceeding 100 or more per hour in two directions, community acceptance may depend on use of low air and noise emission vehicles. Low on-board noise levels are also desirable from a customer perspective. Three basic types have been used in BRT applications in North America.

1. Internal combustion, hydraulic-mechanical transmission
2. Dual mode, diesel-electric
3. Internal combustion/electric hybrid

Internal Combustion Engines, Hydraulic-Mechanical Transmissions

The most common propulsion plant, and the one most likely to be used if a conventional bus is selected for a BRT application, is the internal combustion engine (i.e., clean diesel, CNG spark ignition) driving an automatic hydraulic-mechanical transmission. There have been significant improvements to this type of drive train over the last two decades in response to the need to reduce emissions.

Electronically controlled, drive-by-wire clean diesel engines will have significantly reduced particulate, hydrocarbon, and CO emissions from pre-emissions control level by orders of magnitude. Exhaust gas recirculation promises to do the same for NOx emissions.

Available electronically controlled, clean diesel engines and self-cleaning (regenerating) catalytic converters enabled by ultra low-sulphur fuel can have even lower particulate and hydrocarbon emissions than CNG spark ignition engines (but with slightly higher NOx emissions). The catalytic converter/ultra low sulphur fuel combination also contributes to reductions in noxious-smelling hydrogen sulphide gas emissions.

Contemporary spark ignition CNG engine systems have low particulate emissions and can be quieter than current diesel engines, but suffer from higher total system weight, have relatively high operating and maintenance costs, and higher initial capital costs of about $50,000 per vehicle. They also have additional fuelling infrastructure costs compared to clean diesel vehicles. Advances in CNG engine and fuel storage technologies may lower CNG vehicle weight and operating costs in the future.

Dual Mode/Dual Power

Dual mode vehicles essentially combine a full performance electric trolley bus with an internal combustion engine (e.g., diesel, CNG) that is also capable of providing full, stand-alone performance. Dual mode vehicles, therefore, have the advantages of both trolleys and conventional buses with internal combustion engines. Electricity is obtained from overhead contact wires for part of a given route's trajectory, typically in an environmentally sensitive city center or tunnel (like Seattle and Boston). Where an overhead contact system cannot be installed or used, (e.g., a freeway) or is not economical, these vehicles have full performance capabilities using an internal combustion engine.

Dual mode vehicles are attractive for BRT because they can combine the performance, environmental, and permanence advantages of trolleybuses, with the flexibility of conventional buses. The main disadvantages of dual mode vehicles are their greater weight and both initial and ongoing increased costs. Rather than needing to maintain a single internal combustion engine/hydraulic-mechanical transmission, dual mode vehicles usually require more maintenance effort and cost, because they have more components.

Hybrid Internal Combustion Engine/Electric

Hybrid vehicles combine an internal combustion engine (e.g., clean diesel, gasoline, CNG-fueled spark ignition, or gas turbine) with a drive system incorporating an electric motor/generator or motor/alternator and an on-board energy storage medium. Contemporary hybrid vehicles can perform significantly better than other vehicles in terms of noise, emissions, fuel consumption and acceleration. While hybrid vehicles are cruising, coasting, braking, or stopped at idle, the internal combustion engine can produce energy for storage, and using the electric motor as a generator/alternator during braking also reduces brake wear and tear. Peak noise levels are reduced, since high engine speeds are not required to provide power for acceleration or to climb hills. Peak requirements are met by stored energy being dumped into the system's motor/generator. The internal combustion engines used in hybrids are also smaller and lighter for the same reason. Air pollution and fuel consumption advantages stem from the more constant load on the internal combustion engine and the ability to tune the engine for peak fuel economy.

M.J. Bradley, Inc. and the University of West Virginia (2001) reported that hybrid vehicles using clean diesel engines with low sulphur fuel have better emissions characteristics than pure CNG engines. Revenue service experience in Seattle with a prototype of the hybrid diesel-electric vehicles they recently purchased also suggests significantly better fuel economy and better acceleration than standard diesel equipment.

Guidance

Guided vehicles, used in conjunction with stations having platforms at the same height as vehicle floors, can be expected to have boarding and alighting times similar to those on heavy rail and some LRT systems, or approximately 2-3 seconds per person per channel (25-35% savings), compared to passenger service times for conventional buses or streetcars with steps of 4 or more seconds per passenger per channel.

No-step, no-gap boarding and alighting can also significantly reduce the time it takes for customers carrying packages, having disabilities, and/or with children in strollers to board and alight from BRT vehicles. This, combined with wide aisles, can significantly reduce passenger service times for these customers, thereby improving schedule reliability. Guided vehicles also have advantages in terms of riding comfort and right-of-way width for dedicated transitways. (As previously noted,

another way to provide no-step boarding is through the use of vehicles with a ramp or bridge deployable at stations). The use of guided vehicles with narrow transitway lane widths has also been cited as a transit-only enforcement tool.

There are two basic types of vehicle guidance systems: mechanical and electronic. The first mechanical guidance system for buses was originally developed for the O-Bahn by Mercedes-Benz (now Evo-Bus). This guidance approach, similar to that utilized on the rubber-tired automated people mover systems found at airports, has been proven in service for many years in Essen, Germany and Adelaide, Australia and in newer, non-O-Bahn applications in a number of British cites, such as Leeds. These systems utilize a pre-cast, concrete track with low vertical side rails or curbs that are contacted by laterally mounted guide wheels that, in turn, are connected to the vehicle steering system's idler arm. More recent guidance systems (as seen in Bombardier's GST and the Translohr BRT vehicles) use a light-duty track embedded in the pavement to provide guidance and to serve as an electric return for the vehicle's electric power system.

O-Bahn type mechanical guidance systems add about $10-20,000 USD to the cost of each vehicle (depending on the numbers involved) along with some weight and complexity, while the incremental cost of the curbs necessary to guide the vehicles will depend on whether there are already curbs on the respective running ways. The mechanical systems using curbs provide positive guidance and are safe at relatively high operating speeds (in the case of the O-Bahn, over 60 mph [100 kph]).

One important new development in BRT vehicles is the use of advanced electronic technologies (ITS) to provide lateral and even longitudinal vehicle guidance. These systems, as distinct from mechanical guidance technologies, replace physical infrastructure with inexpensive-to-implement magnetic or optical markers on or in the running way. Because of their ease of driver-steered vehicle entry and extraction, the operator can take over at any time and they are compatible with operating plans that feature mixed local and express operations on a single guideway.

There are two types of electronic guidance systems currently in BRT operation: (1) optical, in which a video camera detects the position of a vehicle relative to painted lines on the pavement and steers via a servo motor in the steering mechanism, developed by Siemens and implemented on the Irisbus Civis vehicle; and (2) magnetic, that works essentially the same way as optical, but uses magnets buried in the pavement. The FROG system was implemented on the VL/APTS Phileus.

Figure 3. Boarding and Alighting Electronically Guided Irisbus Agora and Civis Vehicles, TEOR, Rouen, France

Figure 3 illustrates the docking accuracy possible with electronic guidance systems. Customers easily board and alight from electronically-guided Iris bus Agora and Civis vehicles used on the TEOR System in Rouen, France.

The current incremental costs of the electronics and steering servos necessary to make the ITS- driven guidance systems work are currently in the neighborhood of $75-100,000 USD per vehicle. This cost is expected to come down after manufacturers recover research and development costs. Infrastructure costs associated with the systems are modest, since no infrastructure beyond embedded magnets or painted stripes on running way pavements are necessary. A downside of these systems is that they lack the high-speed safety of positive, mechanical guidance.

Aesthetics, Identity and Branding

A unique vehicle identity for a particular BRT service, achieved through livery (paint schemes, colors, icons) and/or design, not only positions the system vis-à-

vis the rest of the transit system, emphasizing functional differences, but also tells the large number of infrequent customers (as high as 35-40% of overall ridership on many rail-based rapid transit systems) where they can board. System branding and identity convey important customer information such as routing and stations served. Vehicle design can complement maps, signs, and other information sources, further enhancing transit ridership.

Compare the exterior look of a specialized BRT Vehicle, the 60-foot articulated Irisbus Civis, to be used for Las Vegas' MAX line (Figure 4), with the conventional bus, an Orion 5, operated by Fairfax County, Virginia in the Dulles Corridor (Figure 5). Both vehicles are attractive and popular in their respective markets. The Fairfax County Connector bus, however, is essentially the same as vehicles serving other routes terminating at the same intermodal transfer facility (West Falls Church MetroRail Station).

Figure 4. Exterior Design, 60-Foot Irisbus Civis Specialized BRT Vehicle, Vegas Blvd. MAX, Las Vegas, Nevada

The uniquely styled Civis, on the other hand, is only used in places where it operates for specialized BRT services, sending a visual cue as to stopping locations and routes for the respective rapid services and advertising the BRT system as providing a distinct service.

Figure 5. Exterior Livery and Design, Orion 5 40-Foot Bus, Fairfax County Connector Dulles Corridor Express Services, Fairfax, County, Virginia

The low floor CNG Neoplan articulated vehicle used on MBTA's Silver Line in Boston (Figure 6) illustrates the creative use of color and livery on conventional equipment to provide a distinct image and identity, matching the color, route name, and map color. Contrast that with the livery of the 40-foot Nova Bus RTS used in regular MBTA local bus service (Figure 7). Such a branded appearance can distinguish a bus in BRT operation from one in local bus service. The vehicle livery and icon or flag should be different from other buses, but match that of BRT stops, stations, and terminals, information signs, graphics, and all printed matter. In this way, it emphasizes that BRT is an integrated system.

Figure 6. Exterior Livery, Neoplan 60-Foot CNG Articulated Bus, Silver Line, MBTA, Boston, Massachusetts

Figure 7. Exterior Livery, Nova Bus RTS 40-Foot Bus, Local Bus Service, MBTA, Boston, Massachusetts

As of 2003, at least five European bus manufacturers (Irisbus, Bombardier, Neoplan, APTS/VDL, and Translohr) have designed and built specialized BRT vehicles with an LRV-like exterior appearance, interior, and other features such as guidance. In Europe and South America, Volvo has BRT vehicle projects under way, while in North America, both New Flyer (Invero) and North American Bus Industries

(Compobus) have BRT vehicle projects either in production or close to the production of prototypes. Examples of their features include large sizes and distinct shapes (lengths from 45-83 feet [13.8 to 25 meters]), large, panoramic passenger windows, dramatically curved front windscreens, multiple doors, lateral guidance/precision docking, quiet, internal combustion-electric hybrid propulsion, and the option for the driver position to be in the center of the vehicle.

Conclusion

The importance of vehicles to the overall success of BRT systems cannot be overstated. Vehicle design affects every aspect of system performance and cost, and their appearance, both external and internal, is a key contributor to the system's image, identity and position in the transportation marketplace.

Based on documented experience to date, the following general guidelines should be considered in BRT system planning and project development:

- Vehicles should be planned and ultimately specified as a function of the type of services offered (e.g., local versus express, mixed) and the nature of the markets served (e.g., short non-work non-home related trips versus long home to work trips).
- Vehicles should provide sufficient passenger capacity at comfortable loading standards (i.e., 3 standees per square meter in North America) for anticipated ridership levels and planned service structure and frequencies. Lengths ranging from 40-45 feet (12.2 - 13.75 meters) for single unit vehicles through double articulated (82-foot [25 meters]) vehicles are in successful revenue service and can be considered.
- Vehicles should have high passenger appeal, be environmentally friendly, easy and convenient to use, and comfortable. Desirable features include air conditioning, bright lighting, panoramic windows, and real-time passenger information.
- Vehicles should be easy and rapid to board and alight. Low floor heights (i.e., less than 15 inches [38 cm]) above pavement level) are desirable unless technologies permitting safe and reliable level boarding and alighting (e.g., rapidly deployed ramps/bridges, some type of precision docking mechanism) can be used.
- A sufficient number of doors having sufficient width should be provided, especially where off-board fare collection is provided. Generally, one door

channel should be provided for each ten feet of vehicle length. Vehicles with doors on either or both sides are available and can enable use of side and/or center platform stations.

- The mix of space devoted to standees and seating will depend on the type of service and nature of the market served (e.g., express versus local). Because a seated passenger occupies more space than a standee, total capacity is higher where the number of seats is lower, all else being equal.
- Wide aisles and sufficient circulation space can lower dwell times and allow for better distribution of passengers, especially to the rear of articulated vehicles.
- Cost-effective bus propulsion systems are available for revenue service that, compared to conventional diesel engine/hydraulic mechanical systems:
    - virtually eliminate particulate emissions
    - are environmentally friendly in terms of CO, HC and NOx emissions
    - are relatively quiet
    - get improved fuel economy
    - accelerate faster.
- There are mechanical and electronic guidance systems in revenue service that can enable rail-like passenger boarding and alighting convenience and service times at stations, reduce right-of-way requirements, and provide a more comfortable ride than conventional buses.
- Vehicles should be well proven in revenue service before being introduced in large numbers for intense BRT operations. Controlling risk is extremely important in the operation of highly visible services.
- BRT operations with standard vehicles in use on other parts of the respective system are acceptable, as long as distinct livery (color schemes), graphics, icons, and other means are employed to provide a unique identity and image. No special features are required to provide acceptable capacity, levels of service, and passenger attractiveness.
- Even where standard buses are used for BRT operations, consideration should be given to internal layouts and door numbers and configurations consistent with the markets served and service provided.

- Use of specialized BRT vehicles is often desirable for high volume trunk routes where the operational benefits of the specialized vehicles will offset their incremental costs.
- Cost should be considered on a life cycle basis, as some of the features that add to initial acquisition costs (e.g., guidance, hybrid drives, stainless steel frames, and composite bodies) have the potential to reduce ongoing operating and maintenance costs, increase passenger revenue, and add to vehicle service life.
- It is critical that vehicle planning and design be fully integrated with planning and design for other BRT elements such as running ways, stations, fare collection, and service plans, if the overall system is to achieve its maximum effectiveness and efficiency.

References

Levinson, H., S. L. Zimmerman, J. Clinger, S. Rutherford, R. L. Smith, J. Cracknell, and R. Soberman. 2003. *TCRP Report 90: Bus Rapid Transit, Volume 1: Case Studies in Rapid Transit.* Washington, DC: Transportation Research Board of the National Academies.

Levinson, H., S. L. Zimmerman, J. Clinger, S. Rutherford, R. L. Smith, J. Cracknell, and R. Soberman. 2003a. *TCRP Report 90: Bus Rapid Transit, Volume 2: Implementation Guidelines.* Washington, DC: Transportation Research Board of the National Academies.

Danaher, A. and P. Ryus. 1999. TCRP Project A-15. *Transit Capacity and Quality of Service Manual, 1st Ed.* Washington, DC: Transportation Research Board.

Bradley and Associates. 2000. *Hybrid Electric Drive Heavy Duty Testing Project Final Emissions Report*. Boston, MA: Northeast Advanced Vehicle Coalition, West Virginia University.

About the Authors

Sam Zimmerman (*sam.zimmerman@dmjmharris.com*) is principal of the transportation planning section of DMJM+HARRIS. His 34-year career has focused on rapid transit planning, transit economics and finance, and travel demand forecasting. Mr. Zimmerman has consulted extensively on rapid transit and transportation planning projects, planning procedures, and policy throughout the U.S. and the world, and has written numerous papers. He was co-principal Investigator for the Transportation Research Board's Bus Rapid Transit (BRT) Implementation Guidelines project, and is project manager for the National Transit Institute's BRT Training course.

Herbert Levinson (*hslevinson@aol.com*) is one of the world's leading experts in traffic engineering, parking, transit planning, and the nexus between transit and traffic operations. He has authored over 150 papers, edited and otherwise contributed to numerous transportation planning and traffic engineering books, taught at major universities such as Yale and New York Polytechnic Institute, performed research for the USDOT, the American Association of Automobile Manufacturers, and the Transportation Research Board (TRB), and has consulted throughout the world. Mr. Levinson is a member of the National Academy of Engineering and serves on the Executive Committee of the Transportation Research Board. He was co-principal investigator for the Transportation Research Board's Bus Rapid Transit Implementation Guidelines project.