Real-Time Dispatching Control for Coordinated Operation in Transit Terminals

KURT KER-TSUNG LEE AND PAUL M. SCHONFELD

For real-time dispatching control in transit terminals, the holding time for each ready vehicle is optimized on the basis of predicted arrival delays of late vehicles and other factors such as expected transfer volumes and vehicle operating costs. Holding times for each ready vehicle are optimized with the proposed numerical approach, which evaluates the dispatching decision at frequent intervals for ready vehicles by evaluating a dispatching objective function. That function is computed by numerically integrating relevant probability distributions. The numerical results can provide general dispatching guidelines. However, the dispatching algorithms are efficient enough to be used in real time for each decision.

Schedule synchronization may greatly reduce transfer delays at transfer terminals where various routes interconnect. Moreover, probabilistic variations in traffic conditions and dwell times at stations may be accommodated to some extent by including safety factors, called slack times in schedules. However, at the scheduled departure time from a transfer terminal, some connecting vehicles may still be late. For any vehicle that is ready to be dispatched, the question is whether to dispatch it on schedule or to wait for late incoming vehicles with connecting passengers. There is a finite (typically a very small) number of such vehicles, and estimates of their late arrival times are presumed to be available. For example, among three vehicles from three connecting routes, suppose Vehicle 1 is ready on time, Vehicle 2 is 1 min late, and Vehicle 3 is 2 min late. In that case, the dispatching decision for Vehicle 1 has three choices: (a) dispatch immediately, (b) wait 1 min for Vehicle 2, or (c) wait 2 min for both Vehicles 2 and 3. For Vehicle 2, the decision will be to either dispatch immediately or wait another minute for Vehicle 3.

Such choices can be well formulated in objective functions that consider the operator cost of delaying a vehicle, the delay cost to users already on board or waiting downstream, and the missed connection cost to passengers transferring from late incoming vehicles (which depends on the wait time until the next suitable departure). The formulation can be extended to more complex cases involving dispatching with real-time information. In that case, the probability that a late vehicle arrives at the transfer terminal within the time interval of the dispatching decision may be obtained from the conditional probability distribution estimated from real-time information.

Several previous studies have concentrated on control strategies to maintain the reliability of service headways on transit routes. Turnquist analyzed wait time variances at stops along one transit line (1). The effects of frequency and reliability on the proportion of random and nonrandom arrivals were explored through a small empirical study. Turnquist also proposed four major classes of strategies to improve reliability: vehicle-holding strategies, reduction in the number of stops made by each bus, single preemption, and provision of exclusive right of way (2). This study showed that service frequency is the most important factor to affect the control strategies. Other findings are that for low-frequency service (fewer than 10 buses per hour), schedule-based holding strategies or zone scheduling are likely to work best. For midfrequency service (10 to 30 buses per hour), zone scheduling or single preemption is most effective, although headway-based holding can also work well if an appropriate control point can be found. In a high-frequency situation on the route (more than 30 buses per hour), an exclusive lane combined with single preemption should be considered.

The problem of determining the optimal dispatching decision for a system at a single service point of one or two vehicles is formulated by Osuna and Newell as a dynamic programming problem (3). They conclude that the optimal decision will hold a vehicle if it returns within less than about half the mean trip time for a one-vehicle route. However, the optimal decision will control the vehicles so as to retain nearly equally spaced dispatch time for a two-vehicle route.

A computer simulation was developed by Abbkowitz et al. (4) and used to comparatively evaluate four transfer strategies in a simple two-route case: (a) unscheduled transfers, (b) scheduled transfers without vehicle waiting, (c) scheduled transfers where the lower-frequency vehicle is held until the higher-frequency vehicle arrives, and (d) scheduled transfers where whichever vehicle arrives first waits for the later vehicle. This approach yielded interesting numerical results about the effects of various route characteristics on the preferred strategy. However, a simulation approach is computationally expensive, and the results are subject to the inherent variance of Monte Carlo methods. Henderson et al. used the ratio of how often passengers are late versus how often they are on time as the service reliability measure since it is more meaningful for passengers (5).

Strategies for controlling vehicle movements to improve service reliability along transit lines and dispatching decisions at stations have also been analyzed by simulation models. Araya and Sone examined the traffic dynamics of automated transit systems in which a fixed number of vehicles are operated according to a preestablished schedule along a single-loop track with on-line stations (6). They executed several simulations to demonstrate the usefulness of the proposed control algorithm. A detailed passenger
flow model is included in the simulator that gives the exact number of passengers, both in vehicles and at station. Van Breusegem et al. used a complete discrete-event traffic simulation model to determine the controlled speed along the route and the dispatch slack time at the station to guarantee system stability (7). They have developed a complete traffic analysis for open lines and loop lines in the class of sequential lines with or without reference to a nominal time schedule. Simulations have also shown the efficiency of the proposed traffic control algorithms and their robustness against disturbances occurring randomly on a loop line.

Lee and Schonfeld optimized the headways and slack times for the operation of multiple transit routes through a transfer terminal and suggested that real-time dispatching control can further improve such a timed transfer system (8).

After a review of these studies, it appears that all previous studies either optimized the preplanned scheduling or analyzed strategies for controlling vehicle movements along one route and for holding vehicles at stations or control points to improve "headway-based" reliability and reduce the waiting along one route rather than transfer delays among different routes. These deficiencies limit the applicability of timed transfer operation in transportation system. Therefore, this study focuses on dispatching control based on real-time computations in a transit timed-transfer system.

SYSTEM DEFINITION

Bus routes, rail transit routes, and other kinds of transit routes may be included in any combination in analyzed systems. It should be noted that a real-time holding or dispatching decision is considered only for coordinated transit operation.

It is assumed here that the probability distribution for travel times of late vehicles such as in Figure 1 and the current positions of late incoming vehicles are already known when a holding or dispatching decision is made. A reasonable number of monitoring points may be set up along each route. The travel time distribution from monitoring points to the transfer terminal can be obtained from the data collected at those points. To reduce the costs for data collection, the monitoring points could be set up at intersections with traffic control and bus stops. In normal traffic conditions (i.e., without incidents along the route), the mean and the standard deviation of travel times along the route should increase as the distances to the transfer terminal increase. When a dispatching or holding decision is made, the mean and standard deviation of the travel time from the current estimated position of late vehicles can be estimated. An example of the relations between the travel time and distance of late vehicles along the route is shown in Figure 2.

The holding times for vehicles ready to be dispatched are either continuous or discrete depending on the characteristics of the empirical distributions for late incoming vehicle arrivals. However, the holding or dispatching decision should be updated in every decision interval. If the optimized holding time of a vehicle is 0, that vehicle should be dispatched immediately.

TOTAL COST FUNCTION

From the system definition, a model for dispatching decisions with real-time computation is developed. The objective function is the

FIGURE 1 Predicted travel time distributions of late incoming vehicles.

FIGURE 2 Example of interpolation of late vehicle arrival times along route.
Lee and Schonfeld

The total cost associated with holding or dispatching decisions. This cost includes the delay cost to vehicles that are ready to be dispatched and to passengers already on board or waiting down­stream along the route and the missed connection cost of late incoming transfer passengers. There will be no further interrelation among the ready routes at the decision time since the connections among them have already been made. Thus, the cost due to holding or dispatching decisions is separable for each ready vehicle. That means that the total cost due to the decisions of all ready routes should be the simple summation of the cost for each ready route. Therefore, the holding or dispatching decisions can be made independently for each ready vehicle.

The total cost for ready Route \( i \) due to the decision includes the delay cost for ready vehicles and passengers and the missed connection cost for late incoming transfer passengers.

\[
C_i = Y_i + U_i
\]

where

\[
C_i = \text{cost due to holding or dispatching decision on Route } i \text{ ($)},
\]

\[
Y_i = \text{delay cost of holding vehicles that are ready to be dispatched and passengers already on board on Route } i \text{ ($)},
\]

\[
U_i = \text{missed connection cost of late incoming passengers transferring to Route } i \text{ ($)}.
\]

In each holding or dispatching decision for each vehicle, the delay cost to ready vehicles, to passengers already on board, and to passengers waiting downstream on Route \( i \) \( Y_i \) can be formulated as

\[
Y_i = (Q_w + B) T_i
\]

where

\[
i = \text{route index of ready routes},
\]

\[
Q_w = \text{number of passengers already on board on Route } i \text{ and waiting passengers downstream along Route } i,
\]

\[
T_i = \text{holding time on Route } i \text{ (min)},
\]

\[
u_w = \text{time value of passengers already on board ($/passenger-min)},
\]

\[
B_i = \text{vehicle operating cost on Route } i \text{ ($/vehicle-min)}.
\]

An hourly operating cost function of the type used by Jansson is used here if the vehicle size \( S_i \) is used on each Route \( i \) (9).

\[
B_i = a_i + b_i S_i
\]

where

\[
a_i = \text{fixed coefficient in vehicle operating cost function on Route } i \text{ ($/vehicle-min)},
\]

\[
b_i = \text{variable coefficient in vehicle operating cost function on Route } i \text{ ($/vehicle-min)},
\]

\[
S_i = \text{vehicle size on Route } i \text{ (seats/vehicle)}.
\]

Therefore, Equation 2 can be formulated as

\[
Y_i = (Q_w + a_i + b_i S_i) T_i
\]

For each dispatching decision, the missed connection cost of late incoming passengers transferring to each ready vehicle \( i \) when it was held for time interval \( T_i \) is determined from the conditional probability that late incoming vehicles arrive after the holding time \( T_i \). Figure 3 shows this conditional probability in which \( A \) indicates the probability that late Vehicle \( k \) arrives after \( T_i \) and \( D \) is the probability that Vehicle \( k \) arrives late. Therefore, \( A/D \) is the conditional probability that late Vehicle \( k \) arrives after \( T_i \) when \( k \) is already late. The total missed connection cost of late incoming passengers transferring to Route \( i \) \( U_i \) is then formulated as

\[
U_i = \sum_{k=1}^{n_k} q_{kw} u_w H_i \left[ \frac{\int_{t_i}^{t_i + 2H_i} f(t_i) dt}{\int_{t_i}^{t_i + 2H_i} f(t_i) dt} \right]
\]

where

\[
k = \text{route index of late routes},
\]

\[
n_k = \text{number of routes with late arrivals},
\]

\[
q_{kw} = \text{transfer passengers from Route } k \text{ to Route } i \text{ (passengers)},
\]

\[
w_k = \text{preplanned slack time of Route } k \text{ at the transfer terminal (min)},
\]

\[
H_i = \text{headway for Route } i \text{ (min)},
\]

\[
H_k = \text{headway for Route } k \text{ (min)},
\]

\[
t_i = \text{vehicle arrival time on Route } k \text{ (min)},
\]

\[
f(t_i) = \text{probability density function for vehicle arrival time on Route } k,
\]

\[
u_w = \text{time value of late incoming transfer passengers ($/passenger-min$)}.
\]

The passenger volumes on board or waiting downstream can be more accurately estimated with advanced fare collection system.
(e.g., smart cards and electronic fare boxes) in which the origins and destinations of passengers and the boarding time can be obtained automatically and transmitted to the control center. However, without advanced ticketing system, the passenger volumes already on board and the transfer volume of late incoming passengers can be estimated from the vehicle loads, (i.e., the volume multiplied by the headway and the volume of passengers waiting downstream along routes can be still be estimated from the volume multiplied by the headway and the holding time).

\[ Q_i = \sum_{j=1}^{n_i} q_i H_j + g_i (H_i + T) \]  \hspace{1cm} (6)

\[ q_u = r_u H_k \]  \hspace{1cm} (7)

where

- \( n_i \) = number of routes ready to be dispatched,
- \( q_u \) = volume of transfer passengers from Route \( j \) to Route \( i \) (passengers/min),
- \( r_u \) = volume of transfer passengers from Route \( k \) to Route \( i \) (passengers/min),
- \( j \) = route index of ready routes,
- \( g_i \) = volume of passengers waiting downstream on Route \( i \) (passengers/min).

Equations 4 and 5 can be substituted into Equation 1 to determine the total cost due to dispatching decision \( C_j \):

\[ C_j = (Q \mu_u + a_i + b_i S_i) T_i + \sum_{k=1}^{n_i} q_k u_k H_k \left[ \int_{t_k}^{t_{k+1}} f(t) dt \right] \]  \hspace{1cm} (8)

Since the total probability that a vehicle arrives late should be 1 when that vehicle is already late, Equations 5 and 7 can be modified as

\[ U_i = \sum_{k=1}^{n_i} q_k u_k H_k \left[ \int_{t_k}^{t_{k+1}} f(t) dt \right] \]  \hspace{1cm} (9)

\[ C_i = (Q \mu_u + a_i + b_i S_i) T_i + \sum_{k=1}^{n_i} q_k u_k H_k \left[ \int_{t_k}^{t_{k+1}} f(t) dt \right] \]  \hspace{1cm} (10)

When the arrival time of late vehicles can be predicted precisely with the advanced real-time information (i.e., deterministic arrival of late vehicles), the only possible candidate holding times for ready vehicles will be the predicted arrival time of each late vehicle \( Z_i \) (i.e., possible holding times are discrete rather than continuous). Therefore, the total cost due to dispatching decision \( C \) can be simplified as

\[ U_i = \sum_{k=1}^{n_i} q_k u_k H_k \left[ \sum_{t_k}^{t_{k+1}} f(t) \right] \]  \hspace{1cm} (11)

where

- \( L \) = route index of late routes in which \( Z_i \) is greater than \( Z_{ub} \),
- \( n_b \) = number of late routes where \( Z_i \) is greater than \( Z_{ub} \),
- \( Z_i \) = predicted late arrival time of Vehicle \( k \), and
- \( Z_{ub} \) = predicted late arrival time of Vehicle \( L \).

If the arrival times of late vehicles are distributed according to a general discrete distribution, Equations 5 and 7 may be expressed

\[ U_i = \sum_{k=1}^{n} q_k u_k H_k \left[ \sum_{a=1}^{n_a} f(a) \right] \]  \hspace{1cm} (12)

\[ C_i = (Q \mu_u + a_i + b_i S_i) T_i + \sum_{k=1}^{n} q_k u_k H_k \left[ \sum_{a=1}^{n_a} f(a) \right] \]  \hspace{1cm} (13)

It should be noted that if it is possible to identify unusual delays of late vehicles due to incidents or vehicle breakdowns from the differences of speed or travel time between the previous vehicles and current vehicle along the route, then very useful real-time information can be provided and delay costs to ready Vehicle 4 avoided.

**OPTIMAL HOLDING TIME AND DISPATCHING DECISION**

After formulating the two components of total cost \( C_i \) (Equation 1) as functions of holding time \( T_i \), the optimal value of \( T_i \) (i.e., \( T^*_i \), where * indicates optimal value) can be sought numerically since the probability distributions are too complex for analytic integration. Thus, numerical integration of conditional probabilistic vehicle arrival distributions is used to compute the missed connection cost. Such numerical integration is much faster and more precise than simulation. Afterward, the following algorithm is used to make each dispatching decision for ready routes.

The algorithm starts from the ready route with the highest passenger volume already on board, with an initial holding time of 0 to determine the total cost due to that decision using Equation 4. Then for each ready route, holding time is increased until total costs do not decrease further. That determines the optimal holding time. The decision for each ready route should be updated in each decision interval or when new events such as arrivals of late incoming vehicles occur. In normal traffic conditions, the optimal holding time of ready vehicles in the current decision should be smaller than the optimal holding time in the last decision. If that does not happen, one may suspect that incidents are delaying the late vehicles. The steps in this algorithm can be stated as follows:

1. Collect empirical data on travel time distributions along each route in each demand period.
2. Estimate relations for means and standard deviations of those travel time distributions and distances along each route.
3. Estimate the passenger volume already on board, the volume of passengers waiting downstream, and the transfer passenger volumes from late incoming vehicles to ready vehicles at the decision-making time.
4. Estimate the mean and standard deviation of travel time distributions for each late vehicle using the relations developed in Step 2 and the current estimated positions of late vehicles.
5. Start from the ready route with the highest passenger volume already on board with an initial holding time \( T_i \) of 0 to determine the total cost due to that decision by Equation 8.
6. Repeat Step 5 by increasing \( T_i \) until no lower cost is obtained to determine the optimal \( T_i^* \).
7. Repeat Steps 5 and 6 for every ready route in the order of increasing ready passenger volume.
8. Update the information of Steps 3 and 4 in the next decision interval \( z \).
9. Repeat Steps 5 through 8 for every ready vehicle in each decision interval \( z \).
10. If \( T_i^* = 0 \), dispatch Vehicle \( i \) immediately. Otherwise, hold Vehicle \( i \) for another decision interval.

At the conclusion of Step 10, the results include the optimal holding time for each ready vehicle and the dispatch decision. These decisions should be reevaluated in each decision interval.

**NUMERICAL RESULTS**

Numerical results were computed mainly for the purpose of investigating the sensitivity of optimal holding times for ready vehicles to various factors such as the ratio of time value of passengers already on board to time value of passengers of late incoming transfer passengers, the ratio of passengers volumes already on board to passenger volumes of late incoming transfer passengers, the vehicle operation costs, and the mean and standard deviations of vehicle late arrival times.

Normal distributions are used here for the numerical analysis. However, the proposed optimization models can work with any late arrival distributions identified from real-time information. It should be noted that since travel times must necessarily be positive, arrival time distributions with infinite left tails, and hence with some negative arrival times, cannot strictly represent reality. However, even when such distributions are used to approximate the true late arrival distributions, the probabilities of early arrivals are small enough to be negligible. Of course, if real-time information is used for late arrivals of incoming vehicles, negative arrivals will never appear.

A three-route example is considered in the numerical analysis. The purpose of this example is to explore the relations among variables and particular parameters through sensitivity analysis. However, the holding or dispatching decisions based on real-time computation can be made for relatively large systems. (For practical purposes the number of routes turns out to be unlimited.)

The baseline parameter values were selected for the numerical analysis because they appeared reasonable and typical; they are as follows:

\[
\begin{align*}
i & = 1.0 \\
\beta_i & = 0.667 \\
\beta_i & = 0.0042 \\
\alpha & = 0.2 \\
\omega & = 0.25
\end{align*}
\]

The demand generated randomly and other input data for each route in the numerical example are shown in Table 1. The transfer passenger volumes for each pair of routes are the

\[
q_j = Q_j (1 - p_c) \left( \sum_{k=1}^{j-1} Q_k - Q_j \right)
\]

where \( p_c \) is the percentage of passenger volumes from each route to transfer terminal and \( Q_j \) is the total passenger demand on Route \( j \) in passengers per hour. With these baseline values, the optimal holding time for Route 1 in the first decision is 1.375 min. The relations between the holding time and the cost components of the cost function due to the first holding or dispatching decision are shown in Figure 4. This figure clearly shows that the optimal holding time represents a trade-off between the dispatching delay cost of ready vehicles and passengers, the vehicle operating cost, and the missed connection cost of late incoming transfer passengers. Figure 6 shows the effects of standard deviations of late arrival times on optimal holding times. The slopes of the \( T_i^* \) curves are determined by the slopes of the normal distributions as standard deviations change. Thus, in Figure 6, \( T_i^* \) first increases as the standard deviation increases. At high values of \( T_i^* \), \( T_i^* \) approaches 0 while \( T_i^* \) still increases. That limits to a finite value of the magnitude of the optimal value of \( T_i^* \).

The effect of the vehicle operating cost on the optimal holding time \( T_i^* \) is shown in Figure 5. It is reasonable that for a given passenger volume, \( T_i^* \) should decrease at a decreasing rate as the cost of delaying vehicles increase. Figure 6 shows the effects of standard deviations of late arrival times on optimal holding times. The slopes of the \( T_i^* \) curves are determined by the slopes of the normal distributions as standard deviations change. Thus, in Figure 6, \( T_i^* \) first increases as the standard deviation increases. At first, the additional uncertainty provides economic justification for a larger safety factor (i.e., holding time). However, as the standard deviation approaches a significant fraction of the headway, it becomes preferable to reduce holding time and allow a higher probability of missed connections in the “tail” of the late vehicle arrival distribution.

Beyond a certain critical standard deviation,
the optimal holding time $T^*$ should be 0, implying that as vehicle arrivals become more uncertain and headway magnitudes do not produce excessive missed connection costs, it becomes uneconomical to leave any safety factors in the holding or dispatching decision. Conversely, holding times is most feasible and desirable when arrival uncertainties are low. In Figure 7, the optimal holding times are 0 when the common headway is too small to be worth coordinating and increase at a decreasing rate beyond certain critical headways. The reason that the optimal holding time remains constant even when common headway increases significantly is that the probability of missing a connection beyond a certain holding time becomes negligible.

To identify the time series of holding or dispatching decisions for the three-route example, the means and the standard deviations of travel time distributions along each route in each decision interval and the numerical results are given in Table 2.

### Table 2: Optimal Results for Preplanned and Real-Time Optimization

<table>
<thead>
<tr>
<th>Cost</th>
<th>Zero Slack</th>
<th>Zero Slack</th>
<th>Optimal Slack</th>
<th>Optimal Slack</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_v$</td>
<td>36.027</td>
<td>36.027</td>
<td>36.027</td>
<td>36.027</td>
</tr>
<tr>
<td>$C_w$</td>
<td>13.128</td>
<td>13.128</td>
<td>13.128</td>
<td>13.128</td>
</tr>
<tr>
<td>$C_m$</td>
<td>38.189</td>
<td>38.189</td>
<td>38.189</td>
<td>38.189</td>
</tr>
<tr>
<td>$C_N$</td>
<td>89.344</td>
<td>89.344</td>
<td>89.344</td>
<td>89.344</td>
</tr>
<tr>
<td>$C_s$</td>
<td>0</td>
<td>0</td>
<td>3.091</td>
<td>3.091</td>
</tr>
<tr>
<td>$C_m$</td>
<td>12.578</td>
<td>12.578</td>
<td>8.256</td>
<td>8.256</td>
</tr>
<tr>
<td>$C_F$</td>
<td>22.454</td>
<td>22.454</td>
<td>18.026</td>
<td>18.026</td>
</tr>
<tr>
<td>$C_y$</td>
<td>109.798</td>
<td>109.798</td>
<td>105.370</td>
<td>105.370</td>
</tr>
<tr>
<td>$C_p$</td>
<td>0</td>
<td>0</td>
<td>2.198</td>
<td>2.198</td>
</tr>
<tr>
<td>$C_D$</td>
<td>10.672</td>
<td>10.672</td>
<td>8.067</td>
<td>8.067</td>
</tr>
<tr>
<td>$C_F$</td>
<td>120.470</td>
<td>114.576</td>
<td>111.549</td>
<td>111.549</td>
</tr>
</tbody>
</table>

- $C_v =$ Vehicle Running Cost ($/min.)
- $C_w =$ User Waiting Cost ($/min.)
- $C_m =$ User In-Vehicle Cost ($/min.)
- $C_N =$ Total Non-Transfer Cost ($/min.)
- $C_s =$ Slack Delay Cost ($/min.)
- $C_m =$ Missed Connection Cost ($/min.)
- $C_D =$ Connection Cost ($/min.)
- $C_F =$ Total Transfer Cost ($/min.)
- $C =$ Total Cost of Preplanned Optimization ($/min.)
- $C_p =$ Holding Cost for Ready Vehicles and Passengers ($/min.)
- $C_M =$ Missed Connection Cost due to Dispatching ($/min.)
- $C_D =$ Real-Time Control Cost ($/min.)
- $TC =$ Total System Cost ($/min.)
- $AC =$ Average Cost ($/strip)$
holding times for Routes 1 and 2 are 1.14 and 0.66 min, respectively. The total delay cost on Route 2 is $0.759, and the missed connection cost saved on Route 3 due to holding on Route 2 is $1.089.

CONCLUSIONS AND EXTENSIONS

Real-time optimization of the holding time for each ready vehicle at a transfer terminal, on the basis of the predicted arrival delays of late incoming vehicles, is proposed to make holding or dispatching decisions. An algorithm reevaluates the dispatching decision at frequent intervals for vehicles that are held. The real-time computation of the objective function is achieved by integrating probability distributions numerically.

Qualitatively, the conclusions from the numerical results may be summarized as follows:

1. Ready vehicles with higher passenger volumes on board should be dispatched immediately when the connecting passenger volume on late vehicles is relatively low.
2. Ready vehicles should be dispatched immediately when the uncertainty about late vehicle arrivals is relatively large.

Possible extensions of the analyses and mathematical models developed in this study may include

1. Applying real data in the numerical analysis to reflect the real arrival patterns,
2. Adapting these models for other types of transportation terminals such as airline hubs, and
3. Applying advanced real-time information from intelligent vehicle-highway systems to improve system operation.

REFERENCES


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