# Service Reliability and Optimal Running Time Schedules

Peter G. Furth and Theo H. J. Muller

**To improve reliability, transit routes have time points at which early vehicles are held. Holding reduces waiting time and the amount of time passengers have to budget for a trip. However, it also reduces operating speed and thus increases passenger riding time and, potentially, operating cost. A new approach is presented for quantifying the user costs associated with unreliability. User cost has three components: excess waiting time, potential travel time or buffer time (related to budgeted travel time), and mean riding time, of which the first two are reliability impacts. For long headway service, these costs can be determined from 2-percentile departure times, 95-percentile arrival times, and mean arrival and departure times at stops. With a simple route operations model on a hypothetical route, impacts of scheduling with different numbers of time points and with different levels of running time and cycle time supplements are explored, and optimal running time schedules are determined. For a typical case, the optimal time point schedule offers net benefits equivalent to 4.5 min of riding time per passenger compared with operation without time point control. Optimal route running times are roughly mean plus one standard deviation of uncontrolled running time, and optimal cycle time is roughly mean plus two to three standard deviations of uncontrolled route running time. Surprisingly, it was found that in an optimal schedule, inserting slack at time points does not increase cycle time, because slack time inserted en route simply substitutes for slack time needed in layover.**

To improve reliability, transit routes have scheduled departure times at terminals and at time points located along the route, and early vehicles are held until the scheduled departure time. Holding at time points truncates the early part of the departure time distribution, converting what would be early departures into on-time departures. The more slack time is inserted into the schedule, the greater the reliability, because slack time raises the probability of an early arrival and therefore (with holding) an on-time departure. However, holding lowers operating speed, affecting both riding time and, potentially, operating cost. Weighing the competing concerns of speed and reliability, there is an optimal amount of slack to insert into the schedule at time points and at the terminal, where it takes the form of layover or recovery time.

The question that is addressed in this paper is, how does one determine how much slack to insert into a route's schedule, both en route and at terminals? The analysis presented here is restricted to longer headway routes, on which passengers do not arrive uniformly over time, but target a particular scheduled departure. Other researchers who have addressed this problem include Wirasinghe (*1*) and Lesley (*2*).

Two problems have hindered past efforts to optimize running time schedules with due consideration for reliability. First, reliability is a matter of extremes and is therefore impossible to analyze without a large sample size. The availability of archived automatic vehicle location (AVL) data at a growing number of transit agencies makes reliability-based scheduling more within reach. Second, there is no method for expressing unreliability as a user cost. On long headway routes, the most common measure of reliability is on-time performance, which is a good indicator operational quality, but is not in the form of a user cost. While everybody will agree that raising on-time performance from 80% to 90% is a good thing, is this improvement worth adding 2 min to passengers' riding time? This kind of tradeoff, which is essential to designing running time schedules as well as other methods of operational control, requires that unreliability be expressed as a user cost.

The approach that this paper proposes extends the authors' earlier work on waiting time (*3*) by accounting for unreliability in riding time as well as waiting time. Unreliability is reflected in two kinds of user costs: excess waiting time, because passengers have to arrive early to make sure they do not miss their bus, and potential or buffer travel time, which is related to budgeted travel time, because passengers have to budget additional time for their trip when their arrival time at the destination is uncertain.

Under the rubric of intelligent transportation systems (ITS), many people have conceived of real-time control systems that employ various operational control tactics, including short-turning, expressing, and deadheading as well as holding, and that inform passengers about vehicle arrival times. The authors believe that while real-time, centrally controlled systems are essential with highly chaotic, short headway service and for responding to large schedule disturbances, holding at time points based on a carefully tuned schedule is an appropriate strategy for dealing with the small disturbances that routinely afflict most transit lines. Time point holding can help prevent small disturbance from becoming major disturbances, and is part of fulfilling the promise to customers inherent in a published schedule. Although holding to a schedule is not usually considered an ITS strategy, in fact, AVL systems are vital to an effective system of time point scheduling and holding. AVL systems can help operators keep to schedule by displaying schedule deviation in real time, by triggering conditional priority at traffic signals, by providing management with a record of operators who habitually depart early, and by providing the archive of operations data needed to create well-tuned schedules (*4*).

P. G. Furth, Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115. T. H. J. Muller, Faculty of Civil Engineering and Geosciences, Transportation and Planning Section, Delft University of Technology, Postbus 5048, 2600 GA Delft, Netherlands. Corresponding author: P. G. Furth, pfurth@coe.neu.edu.

*Transportation Research Record: Journal of the Transportation Research Board, No. 2034,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 55–61.
DOI: 10.3141/2034-07

## REVIEW OF PRACTICE

In designing running time and recovery time schedules, some transit schedulers are forced to work with almost no historical running time data, making adjustments by feel based on a single day's observation or in reaction to complaints. Where enough running time data are available to measure reliability, schedulers typically apply rules of thumb to find a balance between speed and reliability. One common rule of thumb for bus routes, supported by software supplied by scheduling vendors (*4*) and cited in the *Transit Capacity and Quality of Service Manual* (TCQSM) (*5*), is to set running time between time points equal to the mean observed running time. A rule of thumb commonly practiced in the Netherlands is Muller's passing moments method (*6*), designed to achieve a high level of reliability and to create an incentive for operators to comply with time points by setting route running time at 85-percentile uncontrolled running time (essentially, mean plus one standard deviation), and using 85-percentile completion times as a basis for determining segment-level running time schedules.

For determining recovery time or layover at the end of a bus line, a common rule is to use a fixed percentage (typically 15% to 20%) of scheduled running time. Of course, schedulers know that a route's need for recovery time actually depends on its running time variability. A statistically formulated rule of thumb, cited in the TCQSM and in studies of operational control (*7, 8*), is to set scheduled cycle time equal to an arbitrarily chosen percentile, such as the 95-percentile, route running time.

On passenger railroads, holding to scheduled departure times is common practice. Running times between stations are scheduled to be the minimum possible running time plus a supplement to provide a buffer against delays caused by interference with other trains. On the Dutch railroads, for example, supplements are typically 5% to 7%.

Many metros, including some with automated control, use time point control. Running time between stations is scheduled for less than maximum acceleration, effectively putting slack into the schedule. With automated control, acceleration rates and dwell times are adjusted in real time in order to maintain the schedule.

On bus routes, time point discipline is strong in some cities and weak in others. For example, in many European cities, time points are strictly respected, helping to contribute to transit's reputation there for punctuality. And while some U.S. transit systems take pride in complying with schedules, in others it is not uncommon to see on-time performance reports in which 10% to 20% of departures were early. When time point discipline is not enforced, running time schedules are essentially meaningless.

One reason sometimes given for poor holding discipline is the difficulty of enforcement. Another reason is that running time schedules are often unrealistic. Who could expect drivers hold at a time point if doing so will make them arrive late downstream because the schedule is too tight on later segments? This valid concern emphasizes the need for well-tuned, data-driven running time schedules.

As mentioned earlier, optimal running time scheduling requires that unreliability be expressed as a user cost. Common measures of service reliability are reviewed in Furth and Muller (*3, 6*); the most commonly used measure for long headway routes—percentage of departures or arrivals that are early, on time, or late—is not a user cost measure. Wirasinghe's model answers that need by using penalties for buses arriving later than scheduled (*1*). This paper's approach includes some newly proposed user costs related to reliability and explicitly considers the propagation of unreliability from one cycle to another.

## HOLDING, RELIABILITY, AND USER COST

### Impact of Holding on Departure and Arrival Distributions

On bus routes, randomness in dispatching, running time between stops, and dwell time cause variability in buses' arrival and departure time at stops. Holding at time points helps to counteract that variability by truncating the early tail of the departure time distribution, thereby tightening the departure time distribution at the time point, and having a similar but muted tightening effect on arrival and departure time distributions at downstream stops.

Figure 1*a* shows typical distributions of departure time from a stop *i* and arrival time at a downstream stop *j* on a route without holding. Four summary measures of the departure and arrival time distributions are shown: mean and 2-percentile departure time, and mean and 95-percentile arrival time. Figure 1*b* shows the very differently shaped departure and arrival time distributions that occur when time point holding is applied at stop *i* and at other time points between *i* and *j*. At *i*, the early tail of the departure time distribution is truncated, with the truncated probability appearing as a spike around the scheduled departure time. At *j*, the arrival time distribution is tighter than in the uncontrolled case because of holding at upstream time points. Holding results in nonstandard shapes of the departure and arrival time distributions, requiring that analysis be done using numerical methods.

### User Cost Components

Three components of user cost for a passenger traveling from *i* to *j* are also shown in Figure 1. The first is excess waiting time at the origin stop *i*. As argued in Wirasinghe (*1*), passengers on long headway routes want to limit the probability that they miss their targeted departure, and therefore can be expected to arrive before the bus's 2-percentile departure time. The time between a passenger's arrival and the 2-percentile departure time is unknown and unavoidable, having mostly to do with uncertainty in access time, and because it is independent of both the schedule and the service reliability, it may be omitted from the cost function. Waiting that occurs beyond the 2-percentile departure time, however, is avoidable, because if service were perfectly on time nobody would wait beyond this point. Because waiting time ends when the bus departs, on average,

$$\text{excesswait}(i) = E[D_i] - D_i^{0.02} \tag{1}$$

where

$D_i$ = departure time at stop *i*,
$E[\,]$ = expected value, and
$D_i^{0.02}$ = 2-percentile departure time at *i*.

Roughly speaking, excess waiting time is the early tail of the departure time distribution. Comparing the top and bottom diagrams in Figure 1, one can see how holding dramatically reduces excess waiting time at *i* by truncating that early tail.

Some may wonder why this proposed measure of waiting time does not include a term for the occasional long wait that occurs when trips depart early and passengers miss their targeted trip. First, regular passengers will arrange their arrival times to make sure this unhappy event happens with a low but essentially fixed probability *p*, causing an expected additional waiting time of *ph,* where *h* is service headway. Because both *p* and *h* are independent of both the running time

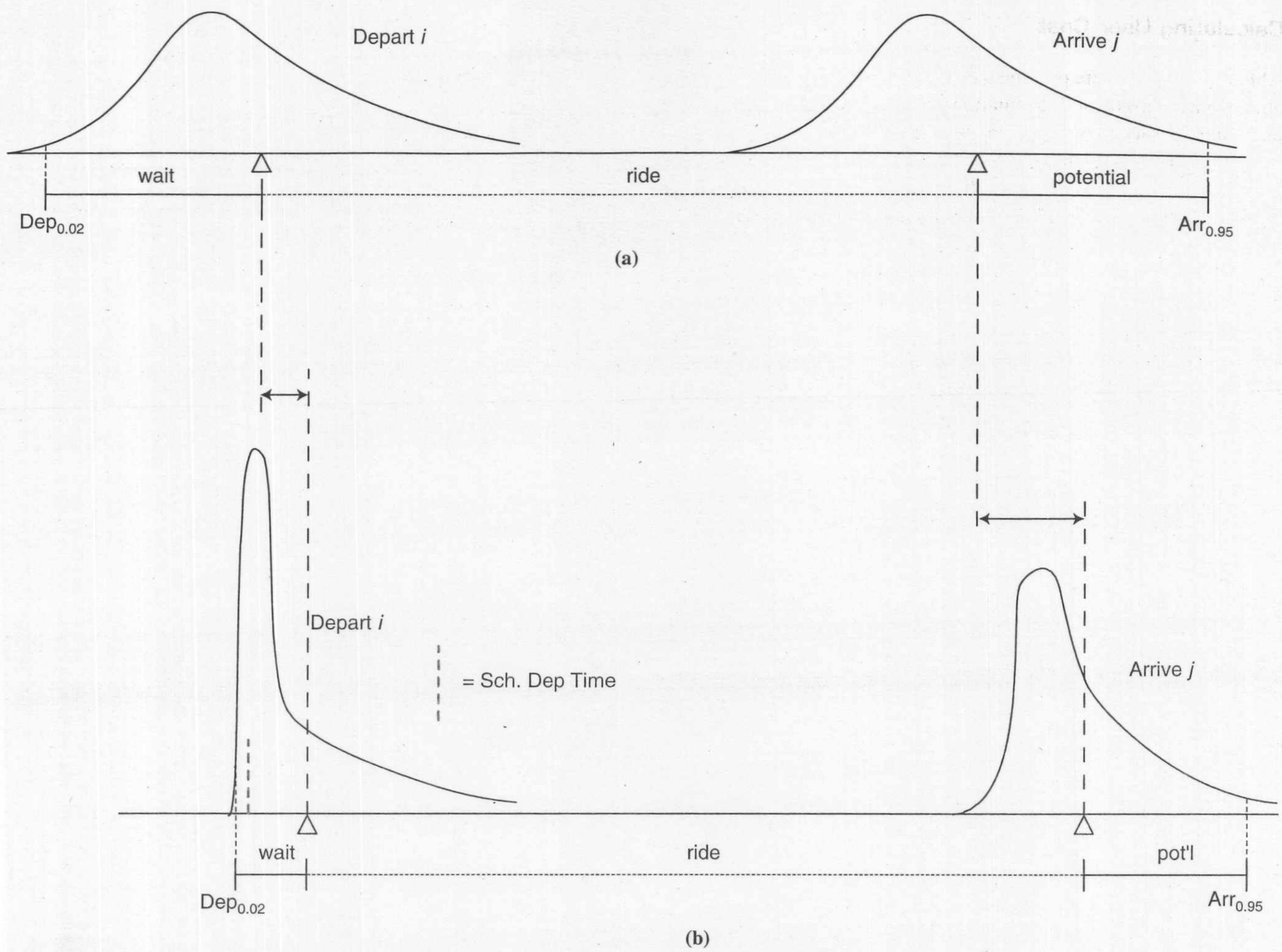


**FIGURE 1 Impact of holding on departure and arrival time distributions and on user cost components.**

schedule and the service reliability, this waiting time increment is a constant that can be omitted from an optimization framework. More importantly, though, this formulation captures the effect of the long wait experience through its arrival time assumption. Consider a passenger on a bus route with a 20-min headway. Suppose he arrives at a stop 1 min before its scheduled departure, but misses the bus because it departed early. On that day, he will suffer an extra 20 min of waiting. But what will he do on subsequent days (if he is still a transit user)? He will arrive at least 3 min early, thereby internalizing the long wait penalty. This is precisely the effect that this paper's measure of waiting time captures by pegging excess waiting time to the 2-percentile departure time.

The second reliability-related user cost component is potential travel time, also called the travel time buffer. Similar to the previous argument proposed by Wirasinghe (*1*), passengers cannot plan a trip based on expected arrival time; rather, they have to budget for a high percentile arrival time—say, the 95-percentile value—to limit their probability of arriving late. Time that is budgeted for travel but not actually consumed waiting or riding, called buffer or potential travel time (pottime), still represents a cost to passengers, because it cannot be relied on or used as freely if it were not so encumbered. It applies at a passenger's destination stop, given by

$$\text{pottime}(j) = A_j^{0.95} - E\left[A_j\right] \tag{2}$$

where $A_j$ is bus arrival time at $j$ and $A_j^{0.95}$ is 95-percentile bus arrival time at $j$. One can see in Figure 1 how holding decreases potential travel time for passengers arriving at $j$: it increases mean arrival time at $j$ by a significant amount, while having little effect on 95-percentile arrival time, because most trips arriving "very late" at $j$ were also late at upstream time points and therefore unaffected by time point holding.

The third user cost component is mean riding time, given by

$$\text{ridetime}(i, j) = E\left[A_j\right] - E\left[D_i\right] \tag{3}$$

Holding increases average riding time, as seen in Figure 1, where holding increases mean arrival time at $j$ by more than it increases mean departure time at $i$.

Excess waiting time and potential travel time are both measures that, like riding time, are user costs that lend themselves to economic evaluation by applying a value of time. As is argued in Furth and Muller (*3*), a plausible set of unit costs for excess waiting time and potential travel time, consistent with demand and value of time studies, are 1.5 and 0.75 times the unit cost of riding time, respectively.

For service quality monitoring, excess waiting time and potential travel time are customer-oriented indicators of service quality, suitable for service standards, improvement targets, and performance incentives.

### Calculating User Cost

Although user costs are presented above in terms of a passenger traveling from an origin $i$ to a destination $j$, origin–destination detail is in fact not necessary for determining user costs; on and off volumes by stop ($\text{on}_i$, $\text{off}_i$) will suffice. Waiting time occurs only at one's origin stop, and potential travel time applies only at one's destination stop, so their evaluation can be done using on and off volumes. While riding time is most naturally formulated as in terms of $T_{ij}$ = passenger trips from $i$ to $j$, it reduces to a function of ons and offs by stop when summed over a route. Using the fact that for a given $i$, $\sum_j T_{ij} = \text{ons}_i$, and for a given $j$, $\sum_i T_{ij} = \text{offs}_j$, total riding time on a route is given by

$$\begin{aligned}\sum_i \sum_j T_{ij}\,\text{ridetime}_{ij} &= \sum_i \sum_j T_{ij}\left\{E[A_j] - E[D_i]\right\} \\ &= \sum_j E[A_j]\left(\sum_i T_{ij}\right) - \sum_i E[D_i]\left(\sum_j T_{ij}\right) \\ &= \sum_j E[A_j]\,\text{offs}_j - \sum_i E[D_i]\,\text{ons}_j \end{aligned} \quad (4)$$

Although user costs have been formulated in terms of arrival and departure times, it is mathematically equivalent and convenient to express them instead as deviations from scheduled arrival and departure times. This permits one to aggregate over all the trips in a route, direction, or period, providing a larger sample size.

Sample calculations of user impacts are shown in Table 1. In the table, *A*dev and *D*dev indicate deviations from scheduled arrival and departure times, and $A_{\text{sched}}$ and $D_{\text{sched}}$ indicate scheduled arrival and departure times, measured as cumulative running times from the start of the route. The two parts of the riding time impact correspond to the two terms in Equation 4. While measuring service reliability on a route requires a large sample of AVL data, the required demand measures (ons and offs by stop) can be estimated from a small sample of manual or automated passenger counts, because only mean values are needed.

## SIMPLE OPERATIONS MODEL

For optimizing time point schedules, an operations model is needed that will predict arrival and departure time distributions when holding is applied. This section demonstrates a simple model applied to a hypothetical route with 16 segments and 17 stops followed by layover.

Three random processes are modeled. The first is segment running time, which includes dwell time at a segment's end stop. On this hypothetical route, each segment has the same running time distribution, following the shifted lognormal distribution with a 1.2-min minimum. In the base case, mean uncontrolled running time for the route is 40 min, and $\sigma_{\text{route}}$, the standard deviation of uncontrolled route running time, is 5 min.

The second random process is departure time. Due to human factors, the departure time of a held bus will deviate slightly from scheduled departure time. Figure 2 shows the distribution used for departure time deviation after holding at a time point (Figure 2*a*) and at a terminal (Figure 2*b*). Negative deviations represent (slightly) early departures.

The third random process is necessary layover. To be distinguished from scheduled layover, it is the time that operators need at a terminal regardless of how late they may arrive, and is modeled with the distribution shown in Figure 2*c*. This distribution has considerable variation to reflect operators' varying needs and behavior; it also includes a minimum time that may be either a layover guarantee or time needed to turn a bus around and pick up waiting passengers.

With the use of these random processes, recursive calculations are done to track the distributions of arrival and departure at each stop, tracking a bus's progress along the route. By convention, scheduled departure time at Stop 1 is Time 0. Calculations begin by assuming a departure time distribution at Stop 1. Because arrival time is given by

$$A_{i+1} = D_i + RT_i \quad (5)$$

where $RT_i$ is the running time on the segment starting at $i$; the distribution of $A_{i+1}$ is simply the convolution of the distributions of $D_i$ and $RT_i$. Convolutions were calculated numerically using a discretized grid with 0.1-min intervals.

The distribution of departure time $D_i$ is determined by modifying the distribution of arrival time $A_i$ to account for holding. If stop $i$ is not a time point, $D_i = A_i$, and so $D_i$ has the same distribution as $A_i$. If $i$ is a time point, trips arriving at $i$ later than $\text{schdep}_i$, the scheduled departure time at $i$, depart with the same distribution as they arrived, while trips arriving early are held, and depart with the distribution shown in Figure 2*a*, shifted by $\text{schdep}_i$. Because this departure time model permits small negative deviations, additional logic was needed to ensure that buses do not depart before they arrive.

Layover is treated as if it were a running time segment following stop $n$, and so the distribution of the time at which a bus is ready to start the next cycle is the convolution of $A_n$ and layover time. The distribution of dispatch time for the next cycle, whose scheduled departure time is the scheduled cycle time *c*sch, is determined by applying

**TABLE 1 Example of User Cost Calculations**

| | | | Summary of Schedule Deviations | | | | Cumulative Scheduled Running Time | | Excess Wait (pax-min/trip), Ons* | Potential Time (pax-min/trip), Offs* | Ride Time (pax-min/trip) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ons (pax/trip) | Offs (pax/trip) | $D\text{dev}^{0.02}$ (min) | $D\text{dev}_{\text{mean}}$ (min) | $A\text{dev}_{\text{mean}}$ (min) | $A\text{dev}^{0.95}$ (min) | $A_{\text{sched}}$ (min) | $D_{\text{sched}}$ (min) | ($D\text{dev}_{\text{mean}} - D\text{dev}_{0.02}$) | ($A\text{dev}_{0.95} - A\text{dev}_{\text{mean}}$) | Part 1 −Ons* ($D_{\text{sched}} + D\text{dev}_{\text{mean}}$) | Part 2 Offs* ($A_{\text{sched}} + A\text{dev}_{\text{mean}}$) |
| Stop 1 | 20 | 0 | −2.1 | 0.9 | | | | 0.0 | 60.0 | | −18.0 | |
| Stop 2 | 10 | 0 | −3.0 | 2.5 | 1.8 | 4.6 | 5.0 | 5.0 | 55.0 | 0.0 | −25.0 | 0.0 |
| Stop 3 | 5 | 10 | −3.4 | 4.6 | 3.2 | 8.2 | 10.0 | 10.0 | 40.0 | 50.0 | −23.0 | 132.0 |
| Stop 4 | 0 | 25 | | | 5.0 | 13.0 | 15.0 | | | 200.0 | | 500.0 |
| Total | 35 | 35 | | | | | | | 155.0 | 250.0 | 566.0 | |

pax = passengers.

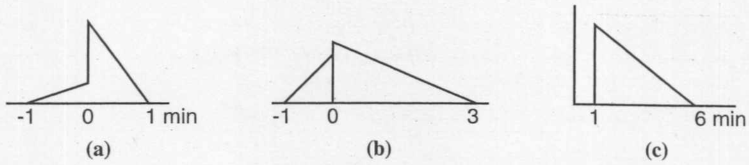


FIGURE 2 Distribution for (*a*) departure time deviation for buses held at time point, (*b*) departure time deviation for buses held at Stop 1, and (*c*) necessary layover at terminals.

holding just as at time points, except using the departure deviation distribution in Figure 2*b* for buses that are held. To prevent arrival and departure time distribution lengths from growing without bound in successive convolutions, upper tails were truncated once the cumulative distribution reached $1 - 10^{-5}$.

### Achieving Steady State

For an operation to be sustainable, it should be in or nearly in steady state; that is, the distribution of Stop 1 departure time (also called dispatch time) should be the same from cycle to cycle, shifted by *c*sch. Otherwise, successive trips will start later and later, imposing a cost on future passengers that is not captured in the model.

### Optimization and Computational Framework

Conceivably, scheduled departure time could be optimized for every time point. However, in order to obtain results that would be generalizable, schedules were parameterized in terms of a running time supplement and a cycle time supplement. Unlike in railroad scheduling, supplements are added to "mean" uncontrolled running time, not to "minimum" running time. Scheduled running time is mean uncontrolled running time (from departure at Stop 1 to arrival at stop *n*) plus the running time supplement, which is divided proportionally over all the route segments. Scheduled cycle time is mean uncontrolled running time plus the cycle time supplement. Scheduled layover is the difference between scheduled cycle time and scheduled running time.

The assumed passenger demand profile was symmetric, with 74 ons and offs per trip and a maximum load of 36 passengers. Stops in the first half of the route had 5 ons and 2 offs, and those in the latter half of the route, 5 offs and 2 ons, with two exceptions: first, there were no offs at the first two stops and no ons at the last two stops; and second, Stops 1, 5, 9, 13, and 17, which are the route's end-, mid-, and quarter-points, had twice their neighboring stops' demand. Modeling stops with demand concentration adds some realism because such stops are natural locations for time points.

Unit costs were \$12 per passenger hour, \$8 per passenger hour, and \$6 per passenger hour for excess waiting time, riding time, and potential travel time. Operating cost, proportional to the cycle length, was set at \$80 per vehicle hour.

Analysis and optimization procedures were programmed in MatLab. The analysis procedure is not a Monte Carlo simulation; rather, the arrival and departure time distributions for each stop are calculated numerically. Useful features of the MatLab platform are the availability of a convolution function, the ability to work with arrays of unspecified length, and the ability to call an optimizer with some schedule parameters held fixed and others free.

Four warm-up cycles were sufficient to achieve near steady state (less than 1-s discrepancy in mean dispatch deviation) for every set of schedule parameters tested. Optimizations with respect to two parameters took about 5 min to run.

## EXPERIMENTAL RESULTS

A series of experiments was run to explore the impacts of scheduling with varying running time and cycle time supplements, and to get insight into the structure of optimal schedules. The first experiment shows how user and operating costs vary with the running time supplement, with every stop with nonzero boardings treated as a time point and the scheduled cycle time optimized for every case. Results are in Figure 3. Negative running time supplements indicate that a running time schedule is tighter than mean running time. The most negative running time supplement, about −13 min, is the same as having no time point control, because buses always arrive late and are therefore never held.

Four cost components and their total can be seen in the figure. All costs are expressed in \$/trip and are shown as differences from when there is no time point control.

Waiting cost falls dramatically with the running time supplement because of the beneficial effect of holding. Potential travel time likewise falls because holding reduces arrival time variability. Riding time increases with scheduled running time, as expected.

An unexpected result was that the operating cost, and therefore the scheduled cycle time, stays essentially unchanged as the running time supplement increases, until the running time supplement becomes quite large. Recall that cycle time was not fixed in these experiments, but was optimized for each level of running time supplement. This result indicates that in an optimal schedule, each minute consumed by

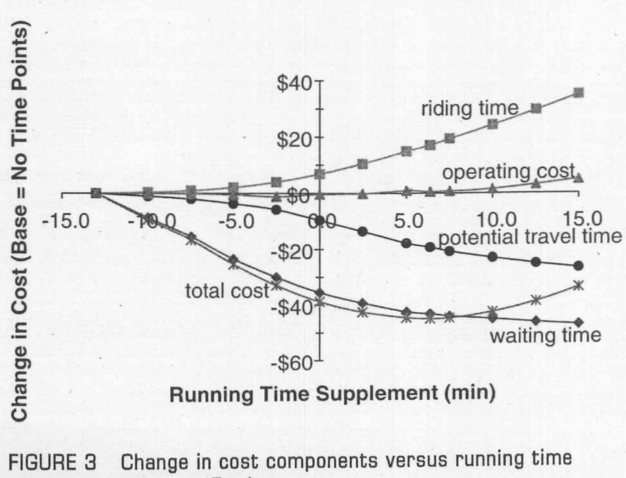


FIGURE 3 Change in cost components versus running time supplement, $\sigma_{route} = 5$ min.

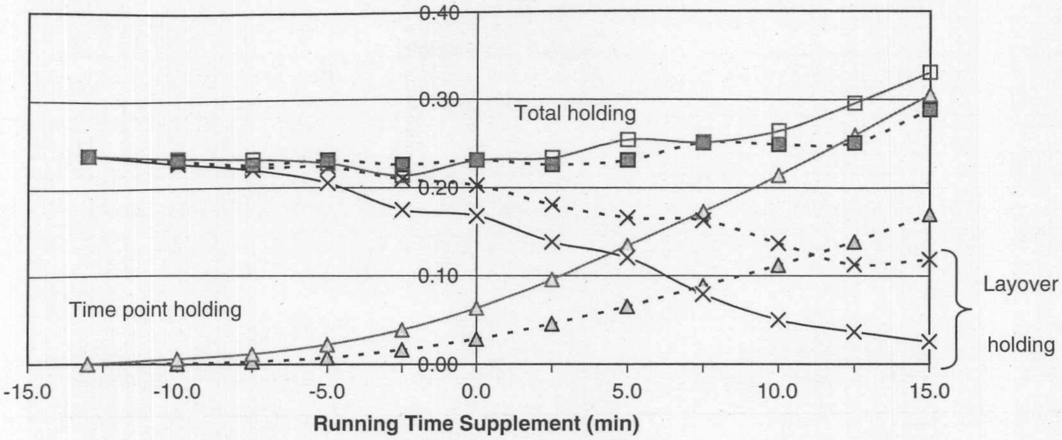


**FIGURE 4 Distribution of holding time (as fraction of uncontrolled running time) versus running time supplement. Solid lines = every stop as a time point; dashed lines = one midroute time point only.**

holding at time points reduces by 1 min the holding time needed at layover. This is a significant result which runs counter to the conventional wisdom that holding at time points is expensive. This analysis suggests that holding at time points may not cost anything at all; slack time en route simply substitutes for layover slack time.

Summing the four cost components yields a total cost curve with a U-shape, with a clear minimum when the running time supplement is about 6.25 min, or 1.25 $\sigma_{route}$. Because operating cost is essentially unchanged, the solution that minimizes user cost is also the solution that minimizes societal cost. With the optimal schedule, the shift of slack time from the terminal to time points reduces user cost by \$45 per trip. On a per-passenger basis, excess waiting time falls by 2.9 min, potential travel time falls by 2.6 min, and average riding time increases by 1.7 min, for a net benefit equivalent to a 4.5-min reduction in riding time per passenger.

Average holding time is not at all the same as the running time supplement, because only early buses are held. Figure 4 offers some insight into how long buses are held at time points and at the dispatch terminal when operating under a schedule with varying running time supplements. Again, cycle length is optimized, which means that it is essentially constant. Results are shown for two cases, with the solid lines for the case in which every stop is a time point (as in Figure 3), and the dashed lines for when the route has only one midroute time point (Stop 9). One can see how the total time spent holding stays roughly constant, and that with increasing the running time supplements, buses are held less at the terminal and more at time points. The substitution effect is weaker when there is only one time point, with a greater share of the holding being retained at the terminal.

Intuitively, one might expect the optimal running time and cycle time supplements to be roughly proportional to $\sigma_{route}$ because the need for slack time and holding grows with running time variability. To test this hypothesis, optimizations were done for varying levels of $\sigma_{route}$, optimizing both the running time and cycle time supplement. Results, shown in Figure 5, show some irregularity at low values of $\sigma_{route}$, but then stabilize for larger values of $\sigma_{route}$. Again, results from two cases (every stop a time point, only one time point) are shown. On the whole, it was found that the optimal running time supplement is approximately 1 $\sigma_{route}$, which is consistent with the passing moments method (*7*).

This result contrasts with the scheduling rule of thumb commonly used in America that scheduled running time should follow mean running time (unless mean running time already includes its appropriate share of holding). The welfare loss associated with using mean uncontrolled running time rather than mean plus one standard deviation is \$6 per trip, the equivalent of 0.6 min of riding time per passenger. This difference is not very large, fortunately, due to the flatness of the total cost curve near the optimum. Still, the results suggest that typical American scheduling practice provides too little slack in the running time schedule, resulting in schedule reliability being below its optimal level.

Based on Figure 5, the optimal cycle length supplement is around 2.5 $\sigma_{route}$. For the route analyzed, this corresponds to roughly the 98.5-percentile running time. However, in this assumed model, randomness in necessary layover makes a large contribution to the variance of needed cycle length. If necessary layover is treated as part of a route's running time, then the optimal cycle length supplement would be at the 95-percentile running time, consistent with the rule of thumb cited earlier.

The final experiment explores the difference between having many or few time points on a route. Results given in Figure 6 are based on optimal schedules for cases of 0, 1, 3, 7, and 14 time points. Because holding control is assumed at the first stop, these alternatives correspond to having 1, 2, 4, 8, and 15 control points. In the last case, every stop with nonzero boardings is a time point.

Results are presented for three cases of running time variability, $\sigma_{route}$ = 3, 5, or 7 min. The first set of results, shown as vertical bars, is the change in societal cost compared with no control. Naturally,

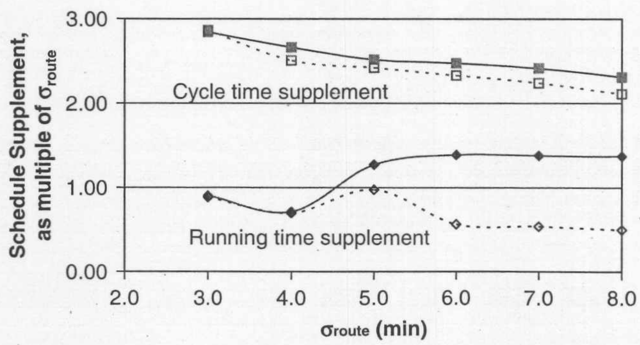


**FIGURE 5 Optimal running time and cycle time supplements as a multiple of $\sigma_{route}$. Solid lines = every stop as a time point; dashed lines = one midroute time point only.**

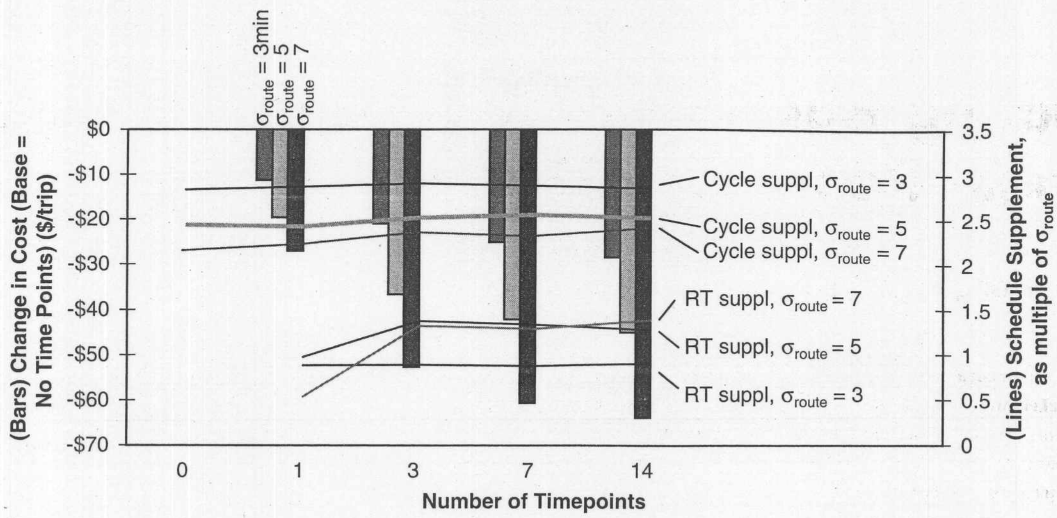


FIGURE 6 Optimal running time and cycle time supplements and cost impact versus number of time points.

impacts are greater with greater values of $\sigma_{route}$. The main result is that increasing the number of time points increases benefits, but with diminishing returns. For the middle variability case ($\sigma_{route}$ = 5 min), about 77% of the benefit of making every stop a time point can be obtained by placing time points at three key stops. Still, the incremental benefit of making every stop a time point is not insignificant; compared to having three time points, the additional benefit is the equivalent of 1 min of riding time per passenger.

Also shown in Figure 6 are the optimal running time and cycle time supplements. The optimal cycle time supplement is not affected by the number of time points, confirming the earlier finding. The optimal running time supplement grows as a route goes from 1 to 3 time points; then as more time points are added, the optimal running time supplement holds constant, meaning the same total amount of extra running time is distributed, but in smaller pieces.

Overall, this final experiment tells us that it is better to spread holding over many points than to hold a lot at a few time points. From the point of view of customer perception, it is also preferred to hold in many small increments rather than to race along and then hold a long time.

## CONCLUSION

For long headway transit routes, excess waiting time and potential travel time are measures of service reliability that can be determined from extreme values of departure and arrival time distributions. Because they can be expressed as user costs, they permit economic trade-offs of reliability against riding time and operating cost, which is essential for designing running time schedules.

With a simple route operations model on a hypothetical route, operating and user impacts were evaluated for different running time schedules, and schedules were optimized with respect to overall societal cost. Experimental results showed substantial benefits from time point holding, estimated at the equivalent of 4.5 min of riding time per passenger for one typical case. Benefits increase as more stops become time points. Route running times should be set at roughly mean plus one standard deviation of uncontrolled running time, and cycle time at roughly mean plus two to three standard deviations of uncontrolled route running time. Significantly, these results suggest that adding slack to running time schedules does not necessarily increase operating costs, because time spent holding en route can simply substitute for holding time at the layover point.

## REFERENCES


1. Wirasinghe, S. C. Initial Planning for Urban Transit Systems. In *Advanced Modeling for Transit Operations and Service Planning* (W. H. K. Lam and M. G. Bell, eds.), Pergamon, Elmsford, N.Y., 2003, pp. 1–30.
2. Lesley, L. J. The Role of the Timetable in Maintaining Bus Service Reliability. *Proc., Symposium on Operating Public Transport,* University of Newcastle, Newcastle-upon-Tyne, United Kingdom, 1975, pp. 36–53.
3. Furth, P. G., and T. H. J. Muller. Service Reliability and Hidden Waiting Time: Insights from Automatic Vehicle Location Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1955,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 79–87.
4. Furth, P. G. *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management Reviewe and Potential.* Transportation Research Board of the National Academies, Washington, D.C., 2006.
5. Kittelson & Associates, Inc., KFH Group, Inc., Parsons Brinckerhoff Quade & Douglass, Inc., and K. Hunter-Zaworski. *TCRP Report 100: Transit Capacity and Quality of Service Manual,* 2nd ed. Transportation Research Board of the National Academies, Washington, D.C., 2003.
6. Muller, T. H. J., and P. G. Furth. Integrating Bus Service Planning with Analysis, Operational Control, and Performance Monitoring. *Proc., Intelligent Transportation Society of America Annual Meeting,* Boston, Mass., 2000.
7. Strathman, J. G., K. J. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, S. Callas, D. Griffin, and J. Hopper. Automated Bus Dispatching, Operations Control, and Service Reliability: Baseline Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1666,* TRB, National Research Council, Washington, D.C., 1999, pp. 28–36.
8. Abkowitz, M., and I. Engelstein. Methods for Maintaining Transit Service Regularity. In *Transportation Research Record 961,* TRB, National Research Council, Washington, D.C., 1984, pp. 1–8.


*The Bus Transit Systems Committee sponsored publication of this paper.*