Optimal Design of Transit Short-Turn Trips

Avishai Ceder

A set of procedures is presented for efficiently designing transit timetables with trips that are initiated beyond the route departure point, or terminated before the route arrival point, or both ("short-turn trips"). In practice, transit frequency is determined at the route segment with heaviest load, whereas at other segments the operation may be inefficient because of partial loads (empty seats). Transit schedulers attempt to overcome this problem by manually constructing short-turn trips to reduce the number of vehicles required to carry out the transit timetable. The study presented herein was meant to improve and automate this task by identifying feasible short-turn points, deriving the minimum fleet size required by a given schedule, and adjusting the number of departures at each short-turn point to that required by the load data (provided that the maximum headway associated with passenger wait time is minimized). Other objectives included minimizing the number of short-turn trips while ensuring that the minimum fleet size is preserved and creating vehicle schedules (blocks). A simple example is used throughout to illustrate the procedures developed.

The first phase of this research, which has been completed and documented (1, 2), provides procedures for using passenger load data to derive alternative timetables along an entire transit route, without short-turn trips. A short-turn trip begins beyond the route departure terminal or is terminated before the route arrival terminal, or both. The possibility of generating short lines permits further saving of vehicles while ensuring that the passenger load in each route segment will not exceed the desired occupancy (load factor).

Schedulers at most transit properties usually include the short-turn operating strategy in their efforts to reduce the cost of service. The procedures commonly used are based only on visual observation of the load profile, that is, the distribution of the loads along the entire route. A potential turn point is determined at the adequate time point (major stop) nearest to the stop at which a sharp decrease or increase in the passenger load is observed. Although this procedure is intuitively correct, the schedulers do not know if all the short-turn trips are actually needed to reduce the fleet size. Unfortunately, each short-turn trip limits service and hence tends to reduce the passenger level of service.

Furth et al. (3) presented an overview of operating strategies on major downtown-oriented bus routes. Among the strategies discussed were short-turn trips, in which the service trip begins farther along the route, but the arrival point of all the trips is the same. The present work designates all the possible categories of short-turn trips for any type of transit lines (crosstown routes, downtown-oriented routes, feeder routes, etc.).

The major objectives set forth herein are as follows:

- To identify feasible short-turn points based on passenger load profile data;
- To derive the minimum fleet size required to carry on a given timetable (including the consideration of deadheading, i.e., nonrevenue trips);
- To adjust the number of departures at each short-turn point to that required by the load data, provided that the maximum headway to be obtained is minimized (this objective results in the maximum possible short-turn trips and the minimum required fleet size);
- To minimize the number of short-turn trips, provided that the minimum fleet size is maintained (for a given timetable, this objective results in increasing the level of service seen by the passengers); and
- To create vehicle blocks for the final derived timetable (a block is a sequence of revenue and nonrevenue activities for an individual vehicle).

To satisfy these objectives, several methods were developed. These methods are based on procedures and algorithms that use data commonly inventoried or collected by most transit properties. Furth (4) uses origin-destination (O-D) data to assess short-turn strategies for route 16 in Los Angeles (SCRTD) between West Hollywood and downtown. Although use of O-D data can improve the scheduling of short-turn trips, this information is not commonly available at transit agencies. The current work is not based on O-D data, but its methods can be extended to include such data whenever they are available.

APPROACH AND BACKGROUND

Framework

The initial information required for constructing the short-turn trips includes:

- A complete timetable for each route timepoint;
- Passenger loads for each time period across all timepoints;
- Minimum frequency or policy headway; and
- A set of candidate short-turn points.

This information is given for both route directions (each direction requires its own data). The complete timetable can be
Initialization:
- Timetable (set of departure times at all timepoints)
- Passenger Loads (for each time period across all time points)
- Minimum Frequency (policy headway)
- Set of candidate short-turn points

Creation of $R_j$ = set of feasible short-turn points, for each time period $j$.

Deficit Function theory and algorithms

Result: A procedure to insert deadheading trips to minimize the route fleet size.

Determination of the required number of departures (max. load/desired occ.) for each timepoint $\epsilon R V_j$.

Result: A new timetable with the maximum possible short-turn trips.

A procedure to exclude departure times from each timepoint $\epsilon R V_j$ except the max. load points. It is carried out to minimize the maximum obtained headway while reducing the number of departures to that required.

Theory and algorithms to maximize the extensions of short turns.

A procedure to insert deadheading trips to minimize the sum of all the maximal deficit function values.

Apply the procedure to extend the short turns toward their original trips without increasing $R_{\text{min}}$.

Result: final optimal timetable with short turns.

FIGURE 1 Flow chart describing the design of transit timetables and vehicle schedules with short turns.

provided by the scheduler or may be derived from the passenger load information ($I$, $2$). Candidate short-turn points are usually all the major route stops (timepoints) at which a public timetable is posted. In some cases, the scheduler may limit the candidate points to only those timepoints at which the vehicles can actually turn back.

The overall program to accomplish the objectives of this work is presented in flowchart form in Figure 1. It starts with a procedure to determine the set of feasible short-turn points, $R_j$, among the candidate points. Then the deficit function theory, as explained below, is used to derive the minimum number of vehicles required to carry out all the trips in the complete two-direction timetable, $N_{\text{min}}$. The required number of departures is determined at each feasible short-turn point, and then the so-called minimum $H$ algorithm is applied. The basis of the algorithm is the elimination of some departures from the complete timetable to obtain the number of departures required. In that procedure, the algorithm minimizes the maximum difference between two adjacent departure times (headway). At this stage, as shown in Figure 1, the deficit function method derives the minimum required fleet size with short turns, $N_{\text{min}}$. If this minimum is less than the size required
without short turns, then another procedure is applied. This second procedure inserts the maximum possible departures back among those previously eliminated, provided that the minimum fleet size, \( N_{\text{min}} \), is maintained. The final step of the overall program is to create vehicle blocks to cover all the trips that appear in the last version of the two-direction timetable.

**Deficit Function: Background**

The deficit function approach for assigning the minimum number of vehicles to carry out a given timetable can be described as follows. A deficit function is simply a step function that increases by one at the time of each trip departure and decreases by one at the time of each trip arrival. Such a function may be constructed for each terminal in a multi-terminal transit system. The only information needed to construct a set of deficit functions is the transit timetable.

The main advantage of the deficit function is its visual nature. Let \( d(k, t) \) denote the deficit for point \( k \) at time \( t \). This point \( k \) can be either a terminal or a timepoint, provided that some trips are initiated or terminated (or both) at this point. The value of \( d(k, t) \) represents the total number of departures less the total number of trip arrivals up to and including time \( t \). The maximal value of \( d(k, t) \) over the schedule horizon is designated \( D(k) \).

It is possible to partition the schedule horizon of \( d(k, t) \) into a sequence of alternating hollow and maximal intervals. The maximal intervals define the interval of time over which \( d(k, t) \) takes on its maximum value. A hollow interval is defined as the interval between two maximal intervals. Hollows may consist of only one point, and if this case is not on the schedule horizon boundaries, the graphical representation of \( d(k, t) \) is emphasized by a clear dot.

If the set of all the route and points (terminals or timepoints) is \( E \), the sum of \( D(k) \) for all \( k \in E \) is equal to the minimum number of vehicles required to service the set \( E \). This is known as the fleet size formula, independently derived by Bartlett (5), Gertsbach and Gurevich (6), and Salzborn (7, 8). Mathematically, for a given fixed schedule:

\[
N = \sum_{k \in E} D(k) = \sum_{k \in E} \max d(k, t)
\]  

(1)

where \( N \) is the minimum number of vehicles required to service the set \( E \).

When deadheading (DH) trips are allowed, the fleet size may be reduced below the level described in Equation 1. Ceder and Stern (9) describe this procedure in the construction of a unit reduction deadheading chain (URDHC). Such a chain is a set of nonoverlapping DH trips that, when inserted into the schedule, reduces the fleet size by one. The procedure continues to insert URDHCs until no more can be inserted or until a lower bound on the minimum fleet size is reached. (determination of the lower bound is detailed in the work of Stern and Ceder (10)). The deficit function theory for transit scheduling is extended by Ceder and Stern (11, 12) to include possible shifting in departure times within bounded tolerances.

**INITIAL PROCEDURES**

**Feasible Short-Turn Points**

The short-turn points are usually route timepoints at which the vehicle can turn back without interfering with the traffic flow. It is therefore anticipated that for each route the initial set of candidate short-turn points is given by the scheduler.

Let the set of candidate short-turn points be designated as set \( R_1 \) for one direction and \( R_2 \) for the opposite route direction. Note that \( R_1 \) does not necessarily coincide with \( R_2 \). More specifically,

\[
R_1 = \{r_{11}, r_{12}, \ldots, r_{1n}\}
\]

(2)

\[
R_2 = \{r_{21}, r_{22}, \ldots, r_{2m}\}
\]

(3)

where \( r_j \) is the \( j \)th candidate short-turn point in the direction.
For a given time period, the fluctuation of a passenger load along the entire route (load profile) may reveal that some short-turn points are actually redundant. For example, consider a load profile that consists of 20 stops and 9 candidate short-turn points, as shown in Figure 2. Theoretically, each segment between two adjacent short-turn points can be treated independently with respect to its required frequency. This frequency is determined by the maximum observed load in the segment, which is marked by a hatched area in Figure 2. In the short-turn strategy, however, all the trips must serve the heaviest load segment of the route (in the example, all trips must cross the \( r_{15} - r_{16} \) segment). Hence fewer trips are required between \( r_{14} \) and \( r_{15} \) than between \( r_{15} \) and \( r_{16} \), while the latter group of trips must cross the \( r_{15} - r_{16} \) segment. Consequently, the point is redundant. The same argument also holds for \( r_{16} \), which is located after the max load segment.

The exclusion of the redundant points at each time period \( j \) results in a set of feasible short-turn points, \( R_j \), and this analysis is important from the computational time viewpoint. In the formal description of the algorithm, there is an additional analysis of the difference between the required frequency at the short-turn point associated with the max load segment and a considered short-turn point. If this difference is small, the short-turn point under consideration can be deleted. The limit on this difference can be determined by the scheduler; otherwise, it is automatically set to 1.0. Note that the difference in the frequencies is equivalent to the difference in the load, and there are always stochastic variations in that load. Hence, if this difference is small, it is not reasonable to consider short-turn trips from the associated short-turn point. This procedure is similar to the manual procedure performed in current practice, where the scheduler selects the short-turn points on the basis of observed sharp increase or decrease in the load on the load profile.

Finally, for subsequent analyses, the union of all \( R_j \) for all time periods \( j \) is denoted set \( R \), or mathematically, \( \cup R_j = R \ \forall j \).

**Minimum Fleet Size for Complete Timetable:**

**Example**

The deficit function theory described in the previous section is used to determine the minimum number of vehicles required to cover the complete timetable without short-turn trips. This minimum size is designated \( N_{\text{min}} \), as shown in Figure 1.

A simple example is used to illustrate the deficit function approach and the procedures developed. This example, which is given in Figure 3, is also used in the work of Ceder (13). It is based on a timetable that covers a schedule of about 2 hr (these hours refer to the departure times at the maximum load points). The route (set \( R \)) includes three breakpoints \( (A, B, C) \), and the average travel times for service and deadheading trips are given in Figure 3.

By using the deficit function approach as a basis, it is possible to construct \( d(A, t) \) and \( d(C, t) \). The minimum number of vehicles required, without deadheading trips, is \( D(A) + D(C) = 11 \). However, a DH trip can be inserted from \( A \) to \( C \), departing just before the start of the first maximal interval of \( d(C, t) \). Both \( d(A, t) \) and \( d(C, t) \) are then changed, according to the dashed line in Figure 3. \( D(C) \) is reduced from 6 to 5, and the overall fleet size is reduced from 11 to 10. After that, it is impossible to reduce the fleet size any farther through DH trip insertions; hence, \( N_{\text{min}} = 10 \). This condition can also be detected automatically by the lower bound test. The simple lower bound, 10, is equal to the maximum value of the combined function (with respect to the time): \( d(A, t) + d(C, t) \). From the DH trip insertion procedure, the maximum of the combined functions is 10, and therefore \( N_{\text{min}} \) reaches its lower bound. An improved lower bound method appears elsewhere (11).

**PROCEDURE TO EXCLUDE DEPARTURE TIMES: MINIMAX \( H \) ALGORITHM**

**Establishing Level-of-Service Criterion**

The basic information required for considering short turns is the load profile along the entire route. These data are available at most bus properties worldwide and are called ride check information (loads and running times along the entire route). On the basis of this load profile information, each route segment between two adjacent short-turn points can be treated separately. That is, the required number of trips between the \((k-1)\)th and \(k\)th short-turn points for a given direction and time period is:

\[
F_m = \max \left( \frac{P_k}{d}, F_{\text{min}} \right)
\]

where \( P_k \) is the maximum load observed between the two adjacent short-turn points, \( d \) is the desired occupancy (load standard), and \( F_{\text{min}} \) is the minimum required frequency (the reciprocal of what is known as the policy headway).

In current practice the complete timetable is based on the maximum load, \( P_m \), observed along the entire route in a given time period. If the frequency determined from this max load is not based on the policy headway, then its formulation is

\[
F_m = \frac{P_m}{d} \quad P_m = \max_k P_k
\]

The manual procedure performed by the scheduler to create short-turn trips is simply exclusion of departure times to set the frequency at each short-turn point \( k \) to \( F_k \) instead of \( F_m \). The exclusion of departure times is usually performed without any systematic instructions, in the belief that by doing so, it is possible to reduce the number of vehicles required to carry out the timetable.

The result of excluding certain departure times is that some passengers will have to extend their wait at the short-turn points. To minimize this adverse effect, it is possible to set the following (minimum \( H \)) criterion: "Delete \( F_m - F_k \) departure times at \( k \) with the objective of minimizing the maximum wait obtained."

It is known that in a deterministic passenger arrival pattern, the wait time is half the headway. Therefore the minimum \( H \) criterion attempts to achieve the minimization of maximum wait. This criterion is called "minimax \( H \)," and it can represent an adequate passenger level of service whenever the scheduler's strategy allows elimination of some departure times.
Minimax $H$ Algorithm

To solve the optimization problem with the minimax $H$ criterion, a theory was developed by Ceder (14) on the basis of the representation of the problem on a directed network with a special pattern, application of a modified shortest-path algorithm on the network to determine the minimax headway, and application of an algorithm to ensure that the exact number of required departures will be included in the optimal solution. The minimax $H$ algorithm will now be outlined and applied to the example problem presented in Figure 3.

Let $G_m = \{N_m, A_m\}$ be the special network, consisting of a finite node set $N_m$ and a finite set $A_m$ of directed arcs. A general illustration of the special network, accompanied by an example, is presented in Figure 4. In general, $n$ departures are given from the complete timetable, and it is required that only $m < n$ will remain while satisfying the minimax $H$ criterion. The construction of $G_m$ is based on $m - 2$ equally spaced departure times between the first and last given departures, $t_1$ and $t_n$, respectively. These equally spaced departure times are denoted by $t_j$, $j = 2, 3, \ldots, m - 1$ and have the equal headway of $t_j = (t_n - t_1)/(m - 1)$. The $G_m$ network has the following six characteristics:

- $G_m$ consists of $m$ rows. The first and last rows are only nodes $t_1$ and $t_n$, respectively, and there is a row for each $t_j$, $j = 2, 3, \ldots, m - 1$.
- Each node in $N_m$ represents a departure time in the given set of departures; however, it is not necessary that all the given departures be included in $N_m$ (see 7:10 and 8:50 in the
example in Figure 4). Also, the same departures may be represented by several nodes (see 7:45, 8:00, and 8:20 in Figure 4).

- The nodes in each row are organized in increasing time order from left to right with respect to their associated \( t_i \).
- That is, all given nodes \( t_i \) such that \( t_i \leq t_i' < t_i + 1 \) are positioned twice, to the right of \( t_i' \) and to the left of \( t_i + 1 \), where \( t_i \), \( t_i + 1 \), are two adjacent, equally spaced departure times. An exception is that in the second and the \((m - 1)\)th rows, only one node is positioned to the left of \( t_i \) and to the right of \( t_i - 1 \), respectively. These single nodes, \( t_1 \) and \( t_{m-2} \) in Figure 4, are selected such that \( t_1 \) is the node closest to \( t_i' \), provided that \( t_i' < t_1 \), and \( t_{m-2} \) is the node closest to \( t_{m-1} \), provided that \( t_{m-1} < t_{m-2} \).
- The directed arcs in \( A_m \) connect only nodes from the \( k \)th row to the \((k + 1)\)th row, where \( k = 1, 2, \ldots, m - 1 \).
- A directed arc from \( t_i \) to \( t_j \) is included in \( A_m \) if \( t_j > t_i \). An arc from \( t_i \) to \( t_j \) is included if and only if without this arc, \( G_m \) is disconnected.
- The length of an arc from \( t_i \) to \( t_j \) is exactly \( t_j - t_i \).

After constructing \( G_m \), a modified shortest-path algorithm is applied. This is a modified version of the efficient algorithm initially proposed by Dijkstra (15). The Dijkstra method is based on assigning temporary labels to nodes. The label on the node is an upper bound on the path length from the origin node to that node. These labels are then updated (reduced) by an iterative procedure. At each iteration, exactly one of

\begin{figure}
\centering
\includegraphics[width=\textwidth]{network_representation.png}
\caption{General network representation of a given timetable at one location and an example of this network construction approach.}
\end{figure}
the temporary labels becomes permanent, implying that it is no longer the upper bound but rather the exact length of the shortest path from the origin to the considered node.

The modification of the Dijkstra method is in the computation step in which the labels are updated. It is modified from

$$\pi(t_i) = \min \{ \pi(t_i), \pi^*(t_j) + (t_j - t_i) \}$$

(6)

to

$$\pi(t_i) = \min \{ \pi(t_i), \max \{ \pi^*(t_j), (t_j - t_i) \} \}$$

(7)

where $\pi(t_i), \pi^*(t_j)$ are temporary and permanent labels of nodes $t_i$ and $t_j$, respectively. This algorithm is applied to $G_m$, where the origin node is $t_1$, and the algorithm terminates when the temporary label on node $t_n$ becomes permanent.

The third part of the minimax $H$ algorithm ensures that the optimal result will include exactly the required $m$ departures.

Although the modified shortest-path algorithm on $G_m$ determines the value of the minimax headway, it does not ensure that the result will include all $m$ required departures. The detailed description of this third part appears in the work of Ceder (14), along with a procedure to treat also multiple departures (same and more than one departure time in the given timetable).

The minimax $H$ algorithm is now applied to the example problem presented in Figure 3. The given and required number of departures for each hour and direction of travel are indicated in Figure 5. Four $G_m$ networks are constructed to derive the minimax headway. This derivation appears in the figure with a dashed line indicating the optimal path on $G_m$ and the labels of the shortest-path algorithm according to Equation 7. A dashed line with an arrowhead indicates the direction of another optimal solution. Also note that $t_n$ between
7:00 and 8:00 becomes \( t_i \) for 8:00–9:00 to preserve the continuity of the analysis. The other parts of Figure 5 are self-explanatory. In all four cases there is no need to proceed to the third part of the minimax \( H \) algorithm because all required departures are determined by the modified shortest-path algorithm.

**Practical Consideration**

The next step, according to the flowchart (Figure 1), is to use the deficit function approach to determine the minimum fleet size required for the new timetable. Before proceeding to that step, however, the new timetable may need adjustment to comply with two operational transit issues.

The first issue is avoiding a skip-stop operation. The new timetable, following the minimax \( H \) algorithm, may have a trip with missing passage (departure) times at some feasible timepoints located after the trip's initial departure point and before its arrival point. Such a situation can occur because of the independent treatment of each feasible timepoint. The interpretation of such a missed passage time is that the trip may skip the considered stop. Such a strategy, at the planning stage, certainly does not lead to saving vehicles (assuming that the trip’s travel time is not adjusted and remains as in the original timetable), and it has an adverse effect on the passenger level of service. Therefore these missed passage times in the new timetable are inserted back. This act may further reduce the minimax headway at the timepoints at which the missed times were inserted.

The second issue concerns the beginning and end-of-day operational characteristics. Observation of public timetables worldwide reveals that in the early morning and late at night, the short-turn trips are not inserted in an alternating fashion. That is, the first trip to cover the whole route often starts late compared to other initial (short-turn) trips in the day. Similarly, the last trip in the day to cover the whole route starts early in comparison to the other end-of-day trips. Therefore, in the procedures developed, it is optional for the scheduler to decide whether or not the following strategy should be used: in the first and last time periods in the day, the excluded departure times are all extracted from the beginning and end of the times that appear in the complete timetable for that day.

**OPTIMAL EXTENSION OF DETERMINED SHORT-TURN TRIPS**

In the example problem in Figure 3, after the deletion of departures at timepoints \( A \) and \( B \) in directions \( A \rightarrow C \) and \( C \rightarrow A \) (see Figure 5), it is possible to construct the new timetable with the deficit functions. This time, however, all three timepoints \((A, B, C)\) are involved. That is, in the modified timetable, some trips begin at \( B \) and some terminate at \( B \) in directions \( A \rightarrow C \) and \( C \rightarrow A \), respectively. Hence point \( B \) becomes also an end/start point, and the deficit function description can be applied to it. The new timetable and deficit functions are presented in Figure 6. By using the deficit function approach, it is possible to insert a single DH trip from \( C \) to \( B \) to arrive before or at 8:35 (the beginning of the \( d(B, t) \) maximal interval). This results in a minimum fleet size of \( N_{\text{min}} = 9 \) vehicles, which is a saving of one vehicle in comparison with the fleet required for the timetable in Figure 3.

The timetable in Figure 6 is characterized by the maximum determined short-turn trips for minimizing the fleet size. A method to reduce (minimize) the number of short-turn trips, provided that \( N_{\text{min}} \) is maintained, is presented next.

**Extensions of Deadhead Trips**

Denote the modified timetable with maximum short turns by \( T' \); the route and points by \( r_i, i = 1, 2 \), and the intermediate short-turn points (belonging to the set \( R \)) by \( U_j, U_j \in R, j = 1, 2, \ldots, V \), where there are \( V \) short-turn points. To attain \( N_{\text{min}} \), the overall schedule to carry out \( T' \) might also include DH trips. This overall schedule is designated \( S \). In this section, the deficit function properties are exploited to check whether a DH trip can be interpreted as an extension of a short-turn trip in \( T' \).

By using the deficit function theory, a DH trip can be inserted in a certain time window to reduce the fleet size by one. To simplify this possibility, a DH trip is inserted from one terminal to terminal \( k \) so that its arrival time always coincides with the first time that \( d(k, t) \) attains its maximum. The following steps attempt to describe the procedure used to convert DH trips in \( S \) into service trips used in the original timetable:

1. Select a DH trip in \( S \) and call it DH; if there is no such trip in \( S \), stop.
2. If the DH trip is from \( U_i \) to \( r_1 = r_i \) \((i = 1 \text{ or } 2, U_i \in R)\), go to Step 3. If the DH trip is from \( r_1 = r_i \) to \( U_j \) \((i = 1 \text{ or } 2, U_j \in R)\), go to Step 4. If the DH trip is from \( U_i \) to \( U_j \) \((U_i, U_j \in R)\), go to Step 5.
3. Examine an arrival in \( d(U_i, t) \) left of the departure time of DH (start with the one closest to that departure time and proceed to the left) to see whether it can be extended to \( r \) (by replacing DH). If the considered arrival is associated with trip \( P_i \), the extension can be executed if and only if the following three conditions are met: the arrival time of \( P_i \) at \( U_i \) is within the hollow that contains the DH departure time, \( P_i \) was originally planned to continue toward \( r \) (as DH), and the originally planned arrival time of \( P_i \) at \( r \) is equal or less than the arrival time of DH. If all the three conditions are fulfilled, delete DH from \( S \), update \( T' \), \( d(U_i, t) \), and \( d(r, t) \), and go to Step 1. Otherwise, DH remains in \( S \); go to Step 1.
4. Examine a departure in \( d(U_i, t) \) that is right of the arrival time of DH (start with the one closest to that arrival time and proceed to the right) to see whether it can be extended to \( r \) (by replacing DH). If the considered departure is associated with trip \( P_i \), the extension can be executed if and only if the following three conditions are met: the departure time of \( P_i \) at \( U_i \) is less or equal to the arrival time of DH, \( P_i \) was originally planned to start at \( r \) (as DH), and the originally planned departure time of \( P_i \) at \( r \) is within the hollow that contains the DH departure time. If all these three conditions are fulfilled, delete DH from \( S \), update \( T' \), \( d(r, t) \), and \( d(U_i, t) \), and go to Step 1. Otherwise, DH remains in \( S \); go to Step 1.
5. Set \( U_k = r \) and use the procedure in Step 3: if it is terminated successfully (DH is converted to a service trip), execute Adjustment A as follows and go to Step 1. Otherwise, set \( U_k = r \) and use the procedure in Step 4. If it is terminated
Directions: A-C, C-A

<table>
<thead>
<tr>
<th>Time-point</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>C</th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00</td>
<td></td>
<td></td>
<td></td>
<td>7:00</td>
<td></td>
<td>7:35</td>
</tr>
<tr>
<td>7:10</td>
<td>7:15</td>
<td>7:40</td>
<td>7:15</td>
<td>7:15</td>
<td>7:35</td>
<td>-</td>
</tr>
<tr>
<td>7:20</td>
<td>7:20</td>
<td>7:20</td>
<td>7:20</td>
<td>7:20</td>
<td>7:40</td>
<td>-</td>
</tr>
<tr>
<td>Time-table</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7:45</td>
<td>8:05</td>
<td>8:30</td>
<td>7:40</td>
<td>8:00</td>
<td>8:15</td>
<td></td>
</tr>
<tr>
<td>8:10*</td>
<td>8:35</td>
<td>8:35</td>
<td>8:10</td>
<td>8:15</td>
<td>8:25</td>
<td></td>
</tr>
<tr>
<td>8:35*</td>
<td>9:00</td>
<td>9:10</td>
<td>8:20</td>
<td>8:45</td>
<td>8:55</td>
<td></td>
</tr>
<tr>
<td>8:45*</td>
<td>9:25</td>
<td>8:27</td>
<td>8:52</td>
<td>9:07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Departures at B (direction A - C), and arrivals at B (directions C - A)
** with DH trip from C to B (8:30 - 8:35)

FIGURE 6 New timetable with maximum excluded departure times and associated deficit functions.

successfully, execute Adjustment A and go to Step 1. Otherwise, DH remains in S; go to Step 1. Adjustment A: Delete DH from S; update $T'$, $d(U_i, t)$, and $d(U_k, t)$.

**Extensions at Intermediate Short-Turn Points**

$T_i$ is used to denote updated timetable $T'$, including the extensions of DH trips. This $T_i$ is now subjected to further extensions at each $U_i \in R$. An extension of a short-turn trip can be viewed as stretching the trip toward the route end points, $r_i (i = 1, 2)$. An extension does not necessarily mean that the short-turn trip is converted to a full trip along the entire route because it can be only partially extended. That is, an extension can be performed from $U_i$ to $U_k$ ($U_i, U_k \in R$). The three stages at which the extensions at $U \in R$ can be analyzed and executed are as follows: zeroing the maximum deficit function, stretching the maximal interval, and treating the deficit function hollows.

**Zeroing the Maximum Deficit Function**

On the basis of the deficit function properties, it is possible to prove that while $N_{min}$ is preserved, the number of extensions in each $U_i \in R$ from $U_i$ to $r_i$ ($i = 1$ or 2) is greater than or equal to $D(U_i)$. This rule is based on the observation that in each $U_i \in R$, exactly $D(U_i)$ departures can be extended to their original departure point without increasing the required fleet size $N_{min}$. This procedure will eventually lead to
$D(U) = 0$ for all $U_i \in R$. These extensions are obtained through the following basic steps, where $t_i$ and $t_e$ denote the beginning and end of the deficit function maximal interval.

1. Initialization: set $R = \overline{R}$.
2. Select an intermediate timepoint $U_i \in \overline{R}$; if $\overline{R} = \emptyset$ (empty), stop.
3. Check to see whether $D(U_i) = 0$; if so, delete $U_i$ from $\overline{R}$ and go to Step 2; otherwise, continue.
4. Identify a trip (there might be more than one) whose departure time is at $t_i$ of $d(U_i, t)$ and extend this departure to its original time at $t_i$ ($i = 1$ or 2). Update $T_i$, $d(U_i, t)$, and $d(r_i, t)$ and go to Step 3.

Figure 7 illustrates an example of five extensions on the deficit function $d(k, t)$. The first two (numbered 1 and 2) induce $D(k)$ to decrease from 2 to 0. Each extension in Figure 7 refers to a different case, and $d(k, t)$ is updated in sequence. The maximal interval of $d(k, t)$ is indicated by its boundaries $t_i$ and $t_e$.

**Stretching the Maximal Interval**

After reducing all $D(U_i)$ to zero, it is possible to prove the following rule: while preserving $\overline{N}_{\text{min}}$, further extensions can be performed from $U_i \in R$ to $r_i$ ($i = 1$ or 2), up to the point at which the maximal interval is stretched over the whole span of the schedule horizon. This rule is based on the observation that certain arrivals can be extended without increasing $D(U_i)$.
above 0. The span of the schedule horizon is determined by the earliest departure and latest arrival of the original timetable.

These additional extensions to the route end points are executed using the following steps, where $T_3$ denotes the updated timetable after the stage described previously (zeroing the function):

1. Initialization: set $R = R'$.
2. Select $U_i \in R$; if $R = \emptyset$, stop.
3. Check whether the $U_i$'s maximal interval (from $t_i$ to $t_i$) coincides with the span of the schedule horizon; if so, delete $U_i$ from $R'$ and go to Step 2. Otherwise, continue.
4. Identify a trip (there might be more than one) such that its arrival is at $t_i$, of $d(U_i, t)$ and extend this arrival to its original time at $r_i (i = 1 \text{ or } 2)$. Update $T_3$, $d(U_i, t)$, and $d(r_i, t)$, and go to Step 3.

The above procedure is demonstrated by cases c, d, and e in Figure 7. In each case, $r_i$ is updated, and in case f the procedure stops when the $d(k, t)$ maximal interval coincides with the span of the schedule horizon.

**Treating the Deficit Function Hollows**

At this third stage, a search is made to determine more extensions at $U_i \in R$ for departures and arrivals in hollows. Each hollow in $d(U_i, t)$ contains the same number of arrivals as it does departures. The procedure developed does not treat hollows, which consist of only one point. In Figure 7, Case f, for example, there are two hollows. The first consists of two arrivals followed by two departures, and the second is a single arrival and departure. The deficit function theory (8) permits construction of the following extension search procedure in which $T_3$ denotes the updated timetable after the stages just discussed:

1. Initialization: set $R = \overline{R'}$.
2. Select $U_j \in R'$; if $R' = \emptyset$, stop.
3. Check the next trip (with respect to time) in $d(U_j, t)$. If it is the last departure, go to Step 2. If it is an arrival, go to Step 5. Otherwise, continue.
4. Examine this departure. Extend it to its original time at $r_i (i = 1 \text{ or } 2)$. Execute this extension if $D(r_i)$ is unchanged or if $D(r_i)$ is increased, but if it can be reduced (back) through the Unit Reduction DH Chain (URDHC) procedure (9), update $T_3$ and all the involved deficit functions. Then, if $t_i$ of $d(U_j, t)$ does not coincide with the right boundary of the schedule horizon, go to the extension procedure described in the previous stage (Stretching the Maximal Interval). Otherwise, go to Step 3. If the extension cannot be made, repeat this extension examination toward a different intermediate short-turn point $U_k$ (instead of $r_i$) each time, selecting the points backward from $r_i$ to $U_j$.
5. Examine this arrival. Extend it to its original time at $r_i (i = 1 \text{ or } 2)$ and use the URDHC procedure to check whether $D(U_i)$ can remain the same. If so, execute the extension, update $T_3$ and all the involved deficit functions. Otherwise, repeat this extension examination the same way as in Step 4.

Finally, if a new DH trip is introduced according to this procedure, the procedure for extensions of deadhead trips needs to be repeated.

**Extensions on the Example Problem**

The minimax $H$ method was applied in Figure 5 to the example problem described in Figure 3. The resultant timetable $T_i$ (with maximum short-turns) appears in the upper part of Figure 8. The deficit functions of this timetable show that 10 vehicles are required to carry out the timetable without DH trips and that 9 vehicles are required with a single DH trip from $d(C, t)$ to arrive at $d(B, t)$ at 8:35. This is shown explicitly by Figure 6, with $N_{\min} = 9$.

The procedure for extensions of deadhead trips is used to determine whether the DH trip can be converted into a service trip. This examination reveals that the conversion cannot be performed, and hence the DH trip remains in the schedule.

After this first attempt, the procedures described for intermediate short-turn point extensions are applied. Because $D(B) = 0$, the algorithm in the first stage cannot be utilized. Because of the algorithm in the second stage, Extension 1 can be performed (see Figures 8 and 9). Then the algorithm in the third stage is used. It can be observed that Extension 2 alone affects $D(B)$, increasing it by one at 8:10. The URDHC procedure therefore searches for a DH trip that can arrive at $B$ at 8:10. Such a DH trip is inserted from $A$ while ensuring that $D(A)$ remains 3.

The final step is to check the newly inserted DH trip with the procedure for extension of deadhead trips. This permits Extension 3 to be performed. Consequently, among the eight short-turn trips in timetable $T_i$ of Figure 8, three were extended to their original schedule, but $N_{\min}$ remains 9. In other words, the procedures developed identify the minimum (crucial) allowed short-turn trips that are required to reduce fleet size. Figure 9 illustrates the updated deficit functions after the three extensions. It can be observed that no more extensions can be made. In addition, timetable $T_j$ in Figure 9 is the final recommended timetable.

**VEHICLE BLOCKS**

In this section, a procedure is described to assign each of the $N_{\min}$ vehicles to a group of trips in the final schedule. A single group of trips, called a block, exhibits a sequence of service and deadhead trips for an individual vehicle.

The task of scheduling vehicles to chains of trips can be carried out by the first in—first out (FIFO) rule or by a chain extraction procedure described by Gertsbach and Gurevich (6). Here, the FIFO rule is used to construct the $N_{\min}$ vehicle blocks. This rule can be stated as follows: Arrange the list of all trips in the final schedule (the trips in the final timetable and the required DH trips for achieving $N_{\min}$) by their departure time order (disregard locations and direction of travel). Select the first trip on the list and join it to its first feasible successor, which is the first trip down the list that has the same departure location as the arrival location of the last trip selected and has a departure time greater than or equal to the arrival time of the last trip selected. If no such trip is found, then remove all the selected trips from the list, maintaining their order, and assign them to a single block. This
forms the first vehicle block. Repeat the process with the reduced list to create the next vehicle block. Stop when the list is empty.

This FIFO rule is applied to the example problem to construct the nine blocks given in Tables 1 and 2. The basic 23-trip list for this method is presented in the upper part of Table 1, where no DH trip is required to obtain $N_{\text{min}} = 9$.

In the software developed in this work, each block is also accompanied by certain performance measures. These computer-generated measures, along with the generated final timetable, are required for the next scheduler's task (assigning drivers to the blocks). Each block has six performance measures:

- Total idle time (minutes),
- Total service kilometers (or miles),
- Total deadheading time (minutes),
- Total block time (minutes).

Idle time is the vehicle's waiting time in the same location between two adjacent trips in the block. In addition, the software provides the sum, over all blocks, of each measure.

**CONCLUSIONS AND FUTURE WORK**

The final product of this work is a set of programs that execute all the components and tasks described in Figure 1. The outcome of this work can generally be presented in light of the five objectives set forth in the first section. The procedures
developed provide the transit scheduler with a set of feasible short-turn points extracted from the description of passenger load profile in each time period of the day. They also provide the approach and methods for determining the minimum fleet size required to carry out a given schedule. These procedures fulfill the first and second objectives.

The third objective of this study is to reduce the number of departures at each short-turn point to that required from a passenger load standpoint while attempting to minimize the adverse effects of the reduction on the passenger level of service. This objective is fulfilled by adopting the minimax headway criterion, or in other words, by minimization of the maximum passenger wait time. The minimax $H$ algorithm described in this paper provides the mathematical tool to handle this criterion. Moreover, the procedures described thereafter allow for an additional improvement of the passenger level of service while preserving the minimum fleet size obtained through the elimination of some departure times. These procedures fulfill the fourth objective of this study.

The need to construct vehicle blocks (schedules) with short turns for the final timetable is expressed in the last objective of this work. The FIFO rule presented in the last section ensures that exactly $N_{\text{min}}$ blocks are created, where $N_{\text{min}}$ is the optimal (minimum) fleet size required to provide adequate transit service.

Future work should be concentrated along the following lines. The methods should be extended to handle origin-destination data whenever it is available [part of this work is described elsewhere (13)]. More than two end route points should be accommodated. A transit route may consist of branches, and the procedures developed can easily be extended to consider such cases. Finally, the procedures should be mod-

![Figure 9: Optimal timetable of the example problem, with updated deficit functions and the three extensions.](image-url)
### TABLE 1 TRIPS BY DEPARTURE TIME ORDER

<table>
<thead>
<tr>
<th>Trip No.</th>
<th>Departure Time</th>
<th>Departure Location</th>
<th>Arrival Time</th>
<th>Arrival Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7:00</td>
<td>A</td>
<td>7:40</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>7:00</td>
<td>C</td>
<td>7:35</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>7:10</td>
<td>A</td>
<td>7:50</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>7:15</td>
<td>C</td>
<td>7:50</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>7:20</td>
<td>C</td>
<td>7:55</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>7:25</td>
<td>A</td>
<td>8:05</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>7:25</td>
<td>C</td>
<td>7:45</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>7:30</td>
<td>C</td>
<td>7:50</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>7:40</td>
<td>C</td>
<td>8:15</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>7:45</td>
<td>A</td>
<td>8:30</td>
<td>C</td>
</tr>
<tr>
<td>11</td>
<td>7:50</td>
<td>C</td>
<td>8:25</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>7:50</td>
<td>B</td>
<td>8:15</td>
<td>C</td>
</tr>
<tr>
<td>13</td>
<td>7:50</td>
<td>A</td>
<td>8:35</td>
<td>C</td>
</tr>
<tr>
<td>14</td>
<td>7:55</td>
<td>B</td>
<td>8:20</td>
<td>C</td>
</tr>
<tr>
<td>15</td>
<td>8:00</td>
<td>A</td>
<td>8:45</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>8:05</td>
<td>C</td>
<td>8:45</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>8:15</td>
<td>C</td>
<td>8:55</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>8:20</td>
<td>C</td>
<td>9:00</td>
<td>A</td>
</tr>
<tr>
<td>19</td>
<td>8:25</td>
<td>C</td>
<td>9:05</td>
<td>A</td>
</tr>
<tr>
<td>20</td>
<td>8:25</td>
<td>A</td>
<td>9:10</td>
<td>C</td>
</tr>
<tr>
<td>21*</td>
<td>8:30</td>
<td>C</td>
<td>8:35</td>
<td>B</td>
</tr>
<tr>
<td>22</td>
<td>8:35</td>
<td>B</td>
<td>9:00</td>
<td>C</td>
</tr>
<tr>
<td>23</td>
<td>8:40</td>
<td>A</td>
<td>9:25</td>
<td>C</td>
</tr>
</tbody>
</table>

*Deadhead.

### TABLE 2 DESCRIPTION OF BLOCKS DERIVED BY THE FIFO METHOD

<table>
<thead>
<tr>
<th>Block (Vehicle) No.</th>
<th>Departure Location</th>
<th>Departure Time</th>
<th>Arrival Location</th>
<th>Arrival Time</th>
<th>Trip No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>7:00</td>
<td>C</td>
<td>8:40</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>7:40</td>
<td>A</td>
<td>8:15</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>8:25</td>
<td>C</td>
<td>9:10</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>7:00</td>
<td>A</td>
<td>7:35</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>7:45</td>
<td>C</td>
<td>8:30</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>C*</td>
<td>8:30</td>
<td>B</td>
<td>8:35</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>8:35</td>
<td>C</td>
<td>9:00</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>7:10</td>
<td>C</td>
<td>7:50</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>7:50</td>
<td>A</td>
<td>8:25</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>8:40</td>
<td>C</td>
<td>9:25</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>7:15</td>
<td>A</td>
<td>7:50</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>7:50</td>
<td>C</td>
<td>8:35</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>7:20</td>
<td>A</td>
<td>7:55</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>8:00</td>
<td>C</td>
<td>8:45</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>7:25</td>
<td>C</td>
<td>8:05</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>8:05</td>
<td>A</td>
<td>8:45</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>7:25</td>
<td>B</td>
<td>7:45</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>7:50</td>
<td>C</td>
<td>8:15</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>8:15</td>
<td>A</td>
<td>8:55</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>7:30</td>
<td>B</td>
<td>7:50</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>7:55</td>
<td>C</td>
<td>8:20</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>8:20</td>
<td>A</td>
<td>9:00</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>8:25</td>
<td>A</td>
<td>9:05</td>
<td>19</td>
</tr>
</tbody>
</table>

*Deadhead.
ified to handle a network of interlinking routes, in which a vehicle can transverse from one route to another in its block. When interlinking routes are allowed, the minimum fleet size can be reduced further (in comparison with the operation of independent routes).

ACKNOWLEDGMENT

This study was supported by the Urban Mass Transportation Administration and is part of the UMTA/TSC Project Evaluation Series under the Service and Management Demonstration Program.

REFERENCES


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