Avishai (Avi) Ceder
e-mail: **a.ceder@auckland.ac.nz**
Department of Civil and Environmental Engineering
Faculty of Engineering, University of Auckland,
Auckland, New Zealand

Tutorial Handout: Scheduled Service Management

The 20th ISTTT at Noordwijk, The Netherlands

July 16, 2013

1. Introduction

This introduction and other parts of this Tutorial notes are based on the book:

> **Ceder, A. "Public Transit Planning and Operation: Theory, Modeling and Practice", Elsevier, Butterworth-Heinemann, Oxford, UK, 640 p. March 2007.** This book was translated to Chinese by the Tsinghua publishing house, Beijing, China, **June 2010**, and its **2nd Edition** will appear **early in 2014** by Spun Press – Taylor and Francis.

Further reading and more references appear at the end of these notes.

The term public-transport or public-transportation will be also referred as 'transit' or 'public-transit' in these notes. The transit-operation planning process commonly includes four basic activities, usually performed in sequence: (1) network route design, (2) timetable development, (3) vehicle scheduling, and (4) crew scheduling. Figure 1 shows the systematic decision sequence of these four planning activities. The output of each activity positioned higher in the sequence becomes an important input for lower-level decisions. Clearly the independence and orderliness of the separate activities exist only in the diagram; i.e., decisions made further down the sequence will have some effect on higher-level decisions. It is desirable, therefore, that all four activities be planned simultaneously in order to exploit the system's capability to the greatest extent and maximize the system's productivity and efficiency. Occasionally the sequence in Figure 1 is repeated; the required feedback is incorporated over time. However since this planning process, especially for medium to large fleet sizes, is extremely cumbersome and complex, it requires separate treatment for each activity, with the outcome of one fed as an input to the next.

The quantitative treatment of the transit planning process is reflected in the welter of professional papers on these topics and in the development of numerous computer programs to automate (at least partially) these activities. In the last twenty five years, a considerable amount of effort has been invested in the computerization of the four planning activities outlined in Figure 1 in order to provide more efficient, controllable, and responsive schedules. The best summary of this effort, as well as of the knowledge accumulated, was presented at the second through twelve international workshops on Computer-Aided Scheduling of Public Transport, which changed its name to Conference

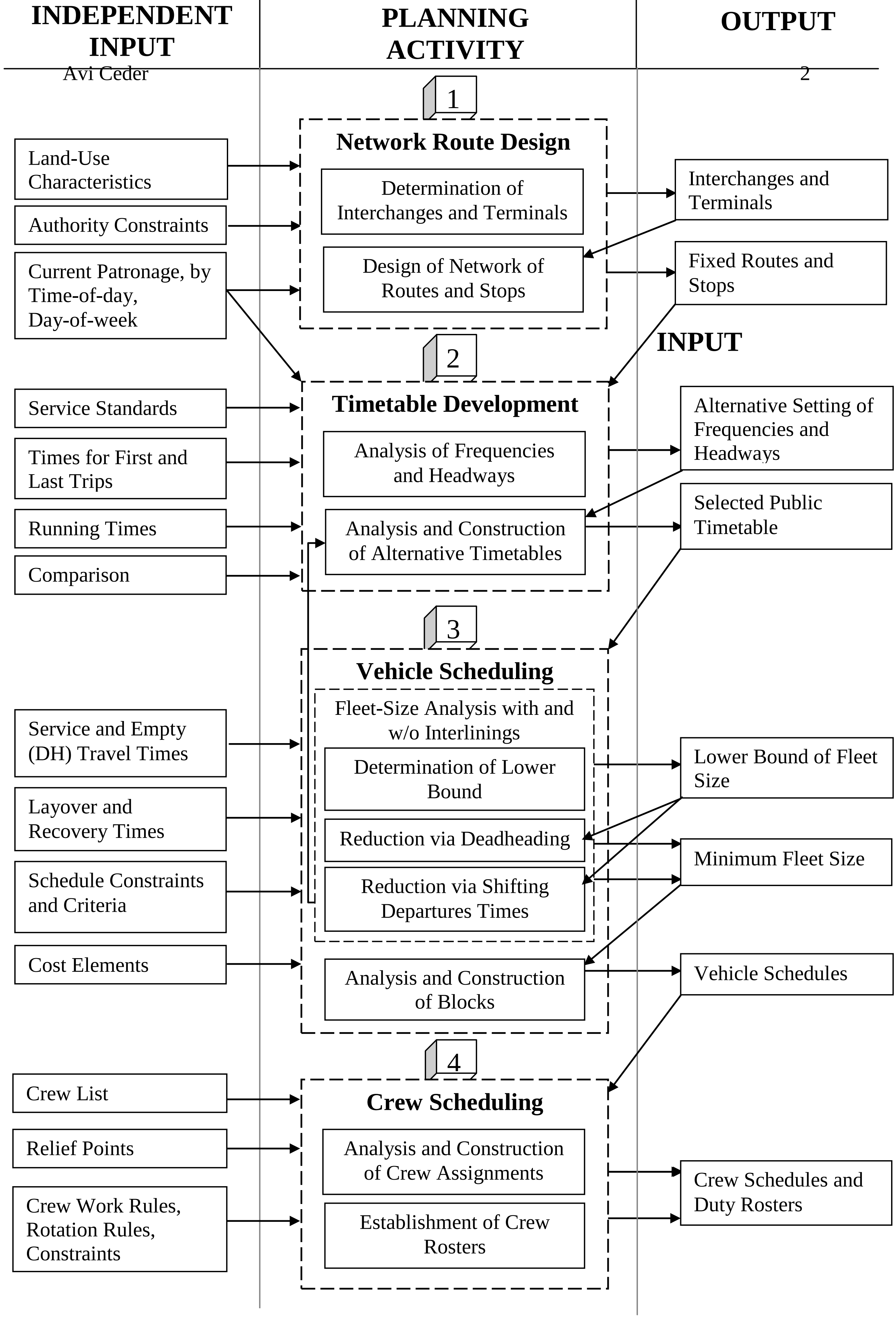

Figure 1 Functional diagram (System Architecture) of a common transit-operation planning

of Advanced Systems of Public Transport (CASPT), in the books edited by Wren (1981), Rousseau (1985), Daduna and Wren (1988), Desrochers and Rousseau (1992), Daduna, Branco, and Paixao (1995), Wilson (1999), Voss and Daduna (2001), Hickman et al. (2008), and in the Proceedings of the 10th CASPT (2006), 11th CASPT (2009), and 12th CASPT (2012). There are also some commercially available software in the area of transit scheduling, such as (in alphabetical order): **AUSTRICS** (www.austrics.com.au), **HASTUS** (www.giro.ca), **ILOG** (www.ilog.com), **MERAKAS Ltd** (http://www.merakas.lt/pikas), **PTV** (www.ptv.de), **ROUTELOGIC** (www.routelogic.com), **ROUTEMATCH** (www.routematch.com), **ROUTEMATE** (www.nemsys.it), **SYSTRA** (www.systra.com), and **TRAPEZE** (www.trapezesoftware.com). These software packages concentrate primarily on the activities of vehicle and crew scheduling (activities 3, 4 in Figure 1) because, from the agencies' perspective, the largest single cost of providing service is generated by drivers' wages and fringe benefits. Focusing on activities 3 and 4 would seem to be the best way to reduce this cost. However, because some of the scheduling problems in these software packages are over- simplified and decomposed into sub-problems, a completely satisfactory or optimal solution is not assured, thus making room for decisions by experienced schedulers. After all, experience is what we gain when expecting something else; said another way, the exam is given first and the lesson after.

An argument in favor of automating activities 3 and 4 is that this scheduling process is extremely cumbersome and time consuming to undertake manually. In addition to the potential for more efficient schedules, the automated process enables services to be more controllable and more responsive. The cost and complexity of manual scheduling have served to discourage adjusting activities 1 and 2. Only with automated scheduling methods, which are becoming more widely accepted, is it feasible to focus on higher levels in the planning process. Nonetheless, a case can be made that these higher levels have received short shift by both researchers and practitioners.

The network route-design activity in Figure 1 focuses almost entirely on individual routes that, for one reason or another, have been identified as candidates for change. Occasionally sets of interacting (e.g., overlapping or connecting) routes are subject to redesign, usually after a series of incremental changes to individual routes has resulted in a confusing, inefficient local system. Although it is difficult to predict the benefits that will result from redesigning any transit network without conducting a detailed assessment, it is reasonable to believe that they will be large compared with the benefits of additional efforts aimed just at problematic scheduling activities (2, 3, and 4 in Figure 1). The approach described in Ceder (2007) generates all feasible routes and transfers connecting each place (node) in the network to all others. From this vast pool of possible routes and transfers, smaller subsets are generated that maintain network connectivity. For each subset thus generated, transportation demand is met by calculating the appropriate frequency for each route. Next, pre-specified optimization parameters are calculated for each subset. Based on the specific optimization parameter desired by the user, it is then possible to select the most suitable subset. This method has been designed as a tool for the planning of future transit networks, as well as for the maintenance of existing networks.

The method presented ensures flexibility by allowing the user either to input own data or to run the analysis automatically.

The timetable-development activity aim is to meet general public transportation demand. This demand varies during the hours of the day, the days of the week, from one season to another, and even from one year to another. It reflects the business, industrial, cultural, educational, social, and recreational transportation needs of the community. The purpose of this activity, then, is to set alternative timetables for each transit route in order to meet variations in public demand. Alternative timetables are determined on the basis of passenger counts, and they must comply with service-frequency constraints. Below alternative timetables are constructed with either even headways, but not necessarily even loads on board individual vehicles at the peak-load section, or even average passenger loads on board individual vehicles, but not even headways. Average even loads on individual vehicles can be approached by relaxing the evenly spaced headways pattern (through a rearrangement of departure times). This dynamic behavior can be detected through passenger-load counts and information provided by road supervisors. The key word in the even-load cases is the ability to control the loading instead of being repeatedly exposed to an unreliable service resulting from an imbalance in loading situations.

The vehicle-scheduling activity in Figure 1 is aimed at creating chains of trips; each is referred to as a vehicle schedule according to given timetables. This chaining process is often called vehicle blocking (a block is a sequence of revenue and non-revenue activities for an individual vehicle). A transit trip can be planned either to transport passengers along its route or to make a deadheading trip in order to connect two service trips efficiently. The scheduler’s task is to list all daily chains of trips (some deadheading) for each vehicle so as to ensure the fulfillment of both timetable and operator requirements (refueling, maintenance, etc.). The major objective of this activity is to minimize the number of vehicles required. Ceder (2007) describes a highly informative graphical technique for the problem of finding the least number of vehicles. The technique used is a step function, which is introduced as far back as 20 years ago as an optimization tool for minimizing the number of vehicles in a fixed-trip schedule. The step function is termed deficit function, as it represents the deficit number of vehicles required at a particular terminal in a multi-terminal transit system. Below the fixed-schedule case is extended to include variable trip schedules, in which given shifting tolerances allow for possible shifts in departure times. This opens up an opportunity to reduce fleet size further. The deficit function, because of its graphical characteristics, has been programmed and is available on a web site. In these notes, the deficit function is applied and linked to the following activities: vehicle scheduling with different vehicle types, the design of operational transit parking spaces, network route design, and short-turn design of individual and groups of routes. The value of embarking on such a technique is to achieve the greatest saving in number of vehicles while complying with passenger demand. This saving is attained through a procedure incorporating a man/computer interface allowing the inclusion of practical considerations that experienced transit schedulers may wish to introduce into the schedule.

The crew scheduling activity goal is to assign drivers according to the outcome of vehicle scheduling. This activity (not covered in these notes) is often called driver-run cutting (splitting and recombining vehicle blocks into legal driver shifts or runs). This crew-assignment process must comply with some constraints, which are usually dependent on a labor contract. A brief summary is given of the conceptual analytical tools used in the modeling and software of this complex, combinatorial problem. The crew-rostering component of this activity normally refers to priority and rotation rules, rest periods, and drivers' preferences. Any transit agency wishing to utilize its resources more efficiently has to deal with problems encountered by the presence of various pay scales (regular, overtime, weekends, etc.) and with human-oriented dissatisfaction. The crew-scheduling activity is very sensitive to both internal and external factors, a factor that could easily lead to an inefficient solution.

2. Max Load (Point Check) Methods

One of the basic objectives in the provision of transit service is to ensure adequate space to accommodate the maximum number of on-board passengers along the entire route over a given time period. Let us denote the time period (usually an hour) as j. Based on the peak-load factor concept, the number of vehicles required for period j is:

$$F_j = \frac{\overline{P}_{mj}}{\gamma_j \cdot c} \qquad (1)$$

where $\overline{P}_{mj}$ is the average maximum number of passengers (Max load) observed on-board in period j, c represents the capacity of a vehicle (number of seats plus the maximum allowable standees), and γ_j is the load factor during period j, $0 < \gamma_j \leq 1.0$. For convenience, let us refer to the product $\gamma_j \cdot c$ as d_{oj}, the desired occupancy on the vehicle at period j. The standard γ_j can be set so that d_{oj} is equal to a desired fraction of the capacity (e.g., d_{oj} = number of seats). It should be noted here that if $\overline{P}_{mj}$ is based on a series of measurements, one can take its variability into account. This can be done by replacing the average value in Equation (1) with $\overline{P}_{mj} + b \cdot S_{pj}$, where b is a predetermined constant and S_{pj} is the standard deviation associated with $\overline{P}_{mj}$.

The Max load data is usually collected by a trained checker, who stands and counts at the transit stop located at the beginning of the Max load section(s). This stop is usually determined from old ride-check data or from information given by a mobile supervisor. Often, the checkers are told to count at only one stop for the entire day instead of switching among different Max load points, depending on period j. Certainly it is less costly to position a checker at one stop than to have several checkers switching among stops. Given that a checker is assigned to one stop, that which apparently is the heaviest *daily* load point along the route, we can establish the so-called *Method 1* for determining the frequency associated with this single stop:

$$F_{1j} = \max\left(\frac{P_{mdj}}{d_{oj}}, F_{mj}\right), \quad j = 1,2...,q \qquad (2)$$

$$P_{md} = \max_{i \in S} \sum_{j=1}^{q} P_{ij} = \sum_{j=1}^{q} P_{i^*j}$$

$$P_{mdj} = P_{i^*j}$$

where F_{mj} is the minimum required frequency (reciprocal of policy headway) for period j, there are q time periods; S represents the set of all route stops i excluding the last stop, i* is the *daily Max load point*, and P_{ij} is a defined statistical measure (simple average or average plus standard deviation) of the total number of passengers on-board all the vehicles departing stop i during period j. The terms P_{mdj} and P_{md} are used for the (average) observed load at the daily Max load point at time j and the total load observed at this point, respectively.

Figure 2 exhibits an example of passenger counts along a 10-km route with six stops between 6:00 a.m. and 11:00 a.m. The second column in the table in this figure presents the distances, in km, between each stop. The desired occupancy and minimum frequency are same for all hours, and hence their time period subscript is dropped: d_o=50 passengers and F_m=3 vehicles, respectively. The set of stops S includes 5 i's, j=1, 2,..., 5, each period of one hour being associated with a given column. The last column in the table represents $\sum_{j=1}^{5} P_{ij}$ in which each entry in the table is P_{ij} (an average value across several checks). Thus, i* is the 3rd stop with P_{md} = 1,740, and P_{mdj} in Equation (3.2) refers only to those entries in the 3rd row.

The second point-check method, or *Method 2*, is based on the Max load observed in each time period. That is,

$$F_{2j} = \max\left(\frac{P_{mj}}{d_{oj}}, F_{mj}\right), \qquad j = 1,2..., q \qquad (3)$$

where $P_{mj} = \max_{i \in S} P_{ij}$, which stands for the maximum observed load (across all stops) in each period j.

In the table in Figure 2, the values of P_{mj} are circled, and a rectangle is placed around P_{md}. Figure 2 also illustrates passenger counts in three dimensions (load, distance, and period), from which the hourly Max load is observed for the first three hours. The results of Equations (1) and (2) applied to the example of Figure 2 appear in Table 1 for both frequency (F_{kj}) and headways (H_{kj}) rounded to the nearest integer, where k=1, 2. The only non-rounded headway is H_{kj}=7.5 minutes, since it fits the so-called *clock headways*: these have the feature of creating timetables that repeat themselves every hour, starting on the hour. Practically speaking, H_{kj}=7.5 can be implemented in an even-headway timetable by alternating between H_{kj}=7 and H_{kj}=8.

Stop #	Distance (km) to next stop	Average Observed Load (passengers), by hour					Total Load (passengers)
		6:00 - 7:00	7:00 - 8:00	8:00 - 9:00	9:00 - 10:00	10:00 - 11:00	
1	2	50	136	245	250	95	776
2	1	100	510	310	208	122	1250
3	1.5	400	420	400	320	200	1740
4	3	135	335	350	166	220	1206
5	2.5	32	210	300	78	105	725

Notes: (1) Route length is 10 km, and stop #6 is the last stop

(2) For all hours, d_o=50, c = 90 passenger, F_m =3 veh./hr

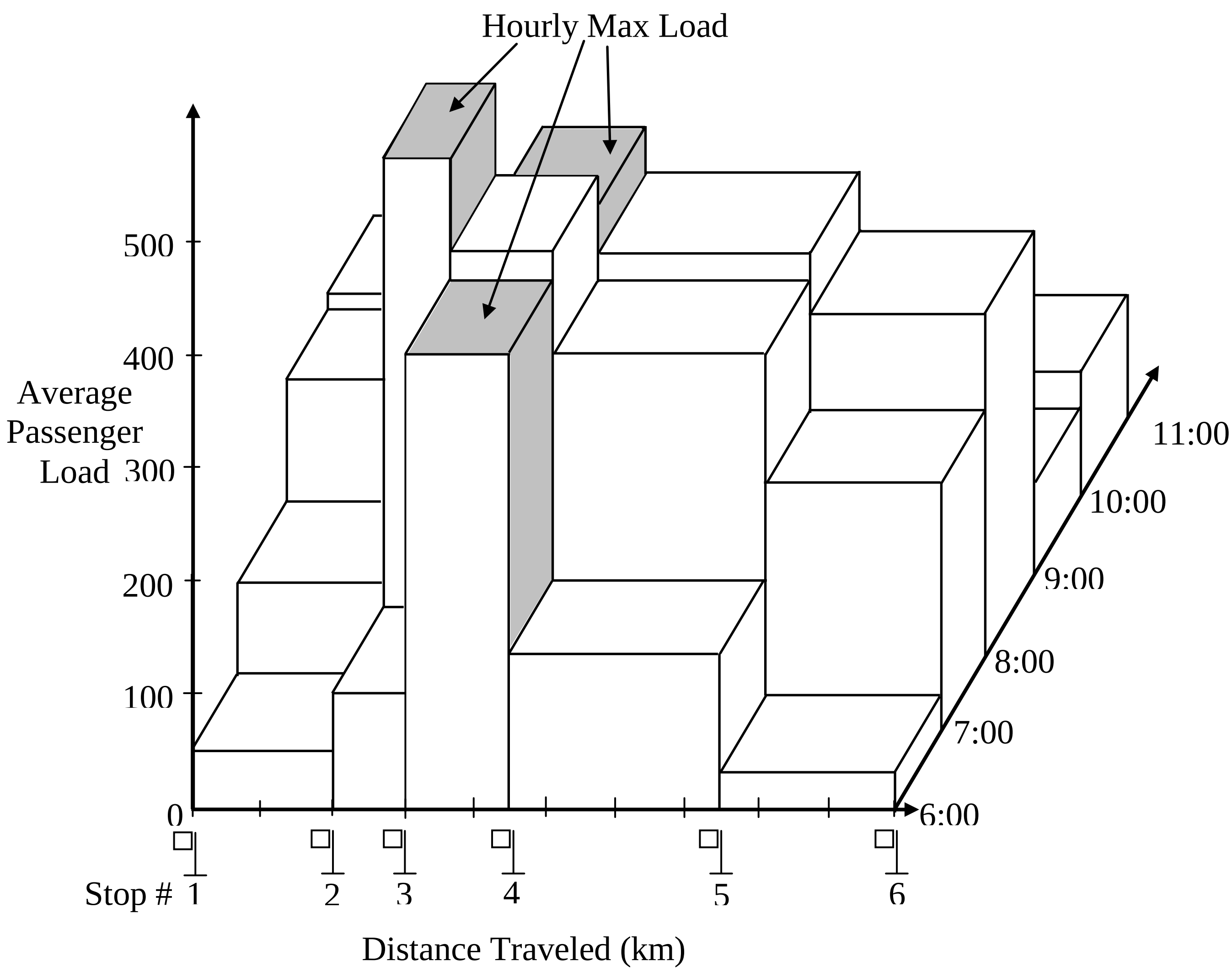

Figure 2 Five-hour load profiles, with indications of hourly and daily Max load points

We will also retain non-rounded F_{kj}'s and show in the next chapter how to use these determined values for constructing timetables with and without even headways.

Table 1 Frequency and headway results for the example in Figure 2, according to Methods 1 and 2

Period j	**Method 1** (Daily Max Load Point)		**Method 2** (Hourly Max Load Point)	
	$\mathbf{F_{1j}}$ (veh/hr)	$\mathbf{H_{1j}}$ (minutes)	$\mathbf{F_{2j}}$ (veh/hr)	$\mathbf{H_{2j}}$ (minutes)
6:00 – 7:00	8.0	7.5	8.0	7.5
7:00 – 8:00	8.4	7	10.2	6
8:00 – 9:00	8.0	7.5	8.0	7.5
9:00 – 10:00	6.4	9	6.4	9
10:00 – 11:00	4.0	15	4.4	14

3. Load Profile (Ride-Check) Methods

The data collected by ride check enables the planner to observe the load variability among the transit stops, or what is termed the *load profile*. Usually a recurrent, unsatisfactory distribution of loads will suggest the need for possible improvements in route design. The most common operational strategy resulting from observing the various loads is short turning (shortlining). A start-ahead and/or turn-back point(s) after the start and/or before the end of the route may be chosen, creating a new route that overlaps the existing route. This short-turn design problem is covered in Ceder (2007). Other route-design-related actions using load data are route splitting and route shortening, both of which are dealt with in Ceder (2007). This section will use the ride-check data for creating more alternatives to derive adequate frequencies, while assuming that the route remains same. Nevertheless, we know that in practice the redesign of an existing route is not an activity often undertaken by transit agencies.

Two examples of load profiles are illustrated in Figure 3. These profiles are extracted from the example in Figure 2 for the first and third hours. It may be noted that in most available transit-scheduling software (see the Introduction section above), these load profiles are plotted with respect to each stop without relating the x-axis to any scale. A more appropriate way to plot the loads is to establish a passenger-load profile with

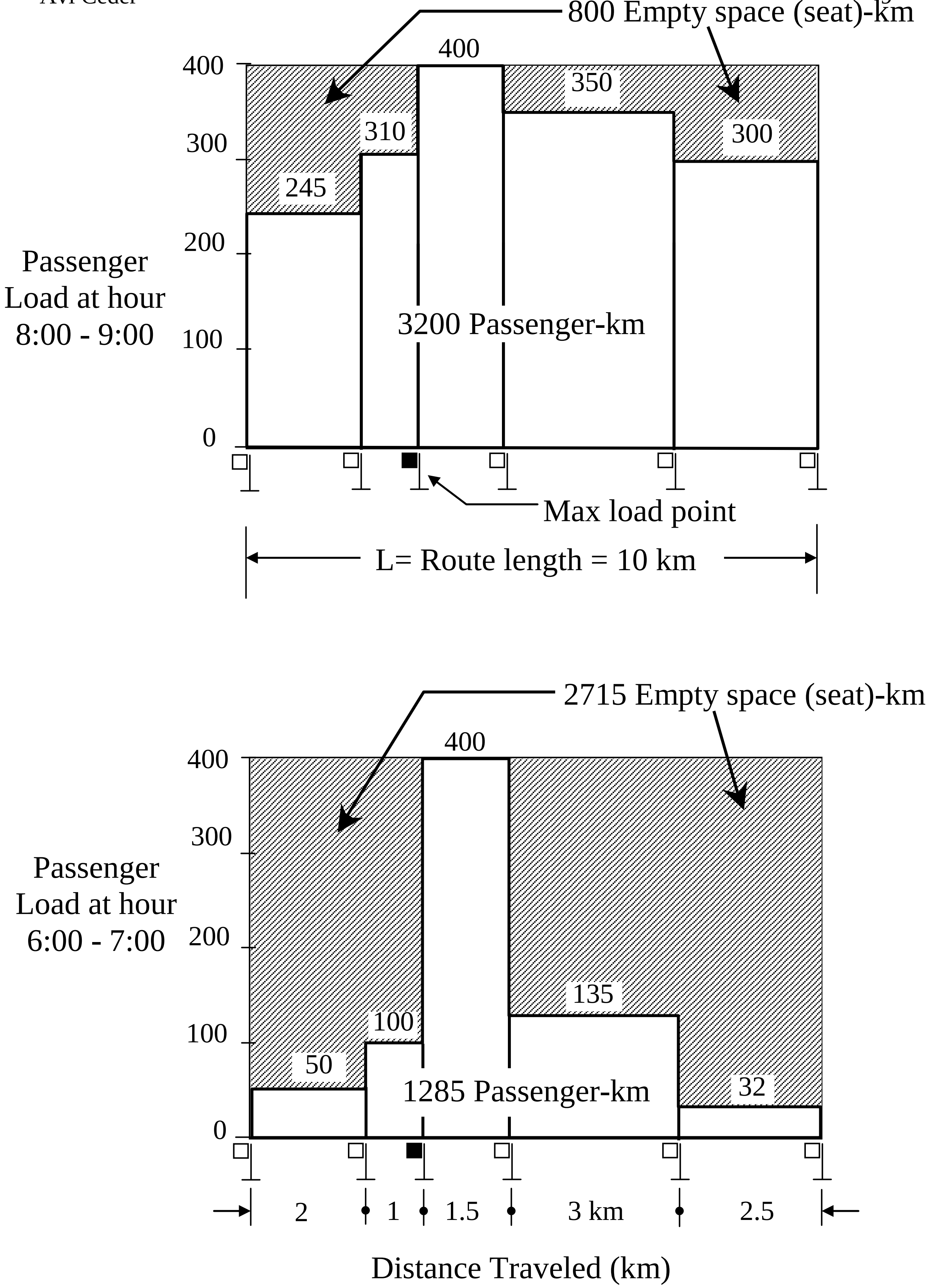

Figure 3 Two load profiles from the example in Figure 2 with the same 8-vehicle frequency, but with different passenger- and empty space-km

respect to the distance traveled from the departure stop to the end of the route. It is also possible to replace the (deterministic) distance by the average running time; in the latter case, however, it is desirable for the running time to be characterized by low and persistent variations. These plots furnish the important evaluation measures of passenger-km and passenger-hour as is also shown in Figure 3.

Let us observe the area marked by dashed lines in Figure 3. If a straight line is drawn across the load profile where the number of passengers is equal to the observed average hourly Max load, then the area below this line but above the load profile is a measure of superfluous productivity. When Method 2 is used to derive the headways, this area represents empty space-kilometers. Furthermore, if d_{oj} in Method 2 is equal to the number of seats--often this is the desired occupancy or load factor used--then this measure is empty seat-kilometers. In light of this measure of unproductive service, we can see in Figure 3 that the 8:00-9:00 load profile is more than twice as productive as the 6:00-7:00 profile, though both have the same (Max load point-based) frequency. We can now use the additional information supplied by the load profile to overcome the problem exhibited in Figure 3 when using Method 2. This can be done by introducing frequency-determination methods based on passenger-km rather than on a Max load measure. The first load-profile method considers a lower-bound level on the frequency or an upper bound on the headway, given the same vehicle-capacity constraint. We call this *Method 3, and it is* expressed as follows:

$$F_{3j} = \max\left[\frac{A_j}{d_{oj} \cdot L}, \frac{P_{mj}}{c}, F_{mj}\right] \qquad (4)$$

$$A_j = \sum_{i \in S} P_{ij} \cdot \ell_i, \qquad L = \sum_{i \in S} \ell_i$$

Where ℓ_i is the distance between stop i and the next stop (i+1), A_j is the area in passenger-km under the load profile during time period j, and L is the route length. The other notations were previously defined in Equations (1), (2), and (3).

One way to look at Method 3 is to view the ratio A_j/L as an average representative of the load P_{ij} (regardless of its statistical definition), as opposed to the Max load (P_{mj}) in Method 2. Method 3 guarantees, on the average basis of P_{ij}, that the on-board passengers at the Max load route segment will not experience crowding above the given vehicle capacity c. This method is appropriate for cases in which the planner wishes to know the number of vehicle runs (frequency) expected, while relaxing the desired occupancy standard constraint and, at the same time, avoiding situations in which passengers are unable to board the vehicle in an average sense. Using the results of Method 3 allows planners to handle: (i) demand changes without increasing the available number of vehicles; (ii) situations in which some vehicles are needed elsewhere (e.g., breakdown and maintenance problems or emergencies); (iii) occasions when there are fewer drivers than usual (e.g., owing to budget cuts or problems with the drivers' union). On the other hand, Method 3 can result in unpleasant travel for an extended distance in which the load (occupancy) is above d_{oj}.

To eliminate or control the possibility of such an undesirable phenomenon, we introduce another method, called *Method 4*. This_Method 4 establishes a level-of-service consideration by restricting the total portion of the route length having loads greater than the desired occupancy. Method 4 takes the explicit form:

$$F_{4j} = \max\left[\frac{A_j}{d_{oj} \cdot L}, \frac{P_{mj}}{c}, F_{mj}\right] \qquad (5)$$

$$\text{subject to (s.t.)} \qquad \sum_{i \in I_j} \ell_i \leq \beta_j \cdot L,$$

where mathematically $I_j = \{ i : \frac{P_{ij}}{F_j} > d_{oj} \}$; in other words, I_j is the set of all stops i in time period j, such that the load P_{ij} exceeds the product of d_{oj} times the frequency F_{4j}, and β_j is the allowable portion of the route length at period j in which P_{ij} can exceed the product $F_{4j} \cdot d_{oj}$. The other notations in Equation (5) were previously defined. By controlling parameter β_j, it is possible to establish a level-of-service criterion. We should note that for $\beta_j = 0$ and $\beta_j = 1.0$, Method 4 converges to Method 2 and Method 3, respectively.

The load profile of Example 1 (see Figure 2) is presented in Figure 4 pertaining to the hour 9:00-10:00. The *considered load level* associated with Method 3 is simply the area under the load profile, divided by L=10 km, or 188.1 passengers in this case. This is an average load profile. However, all loads between stops 1 and 4 (4.5 km), which stretch out for 45% of the route length, exceed this average. To avoid the load exceeding the desired profile for more than a predetermined % of the route length, Method 4 can be introduced. If this percentage is set at 40% of the route length, the considered load will

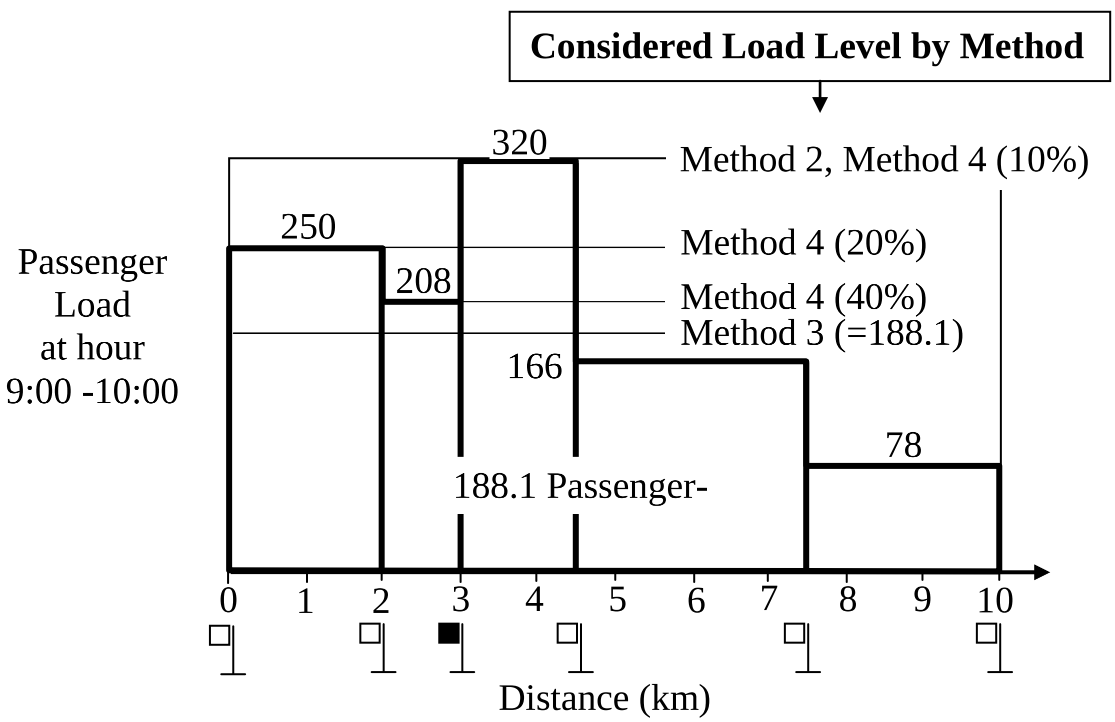

Figure 4 Load profile from the example in Figure 2 between 9:00 and10:00 with considered load levels for three methods and Method 4 standing for 10%, 20%, and 40% of the route length

be 208 passengers, thus allowing only the stretch between stops 1 and 2 (2 km) and that between stops 3 and 4 (1.5 km) to have this excess load. We term this situation Method 4 (40%). Setting the percentage to 20% results in an average of 250 passengers; in the case of 10% (1 km), the considered load level converges to the Method 2 average of 320 passengers.

Figure 5 illustrates the fundamental trade-off between the load profile and Max load concepts. We will show it for the case of Method 4 (20%) and the 9:00-10:00 hour. Based on Figure 2 data and Equation (5), with $\beta_j = 0.2$, we attain F_4 = max (250/50, 320/90, 3) = 5 veh/hr from raising the considered load level from 188.1 to 250 passengers. The 5 assigned departures will, in an average sense, carry 320/5 = 64 passengers between stops 3 and 4 (14 more than the desired load of 50). Nonetheless, this excess load for 1.5 is traded off with (320-250)·8.5 = 595 empty space-km as is shown in Figure 5. This trade-off can be interpreted economically and perhaps affect the ticket tariff.

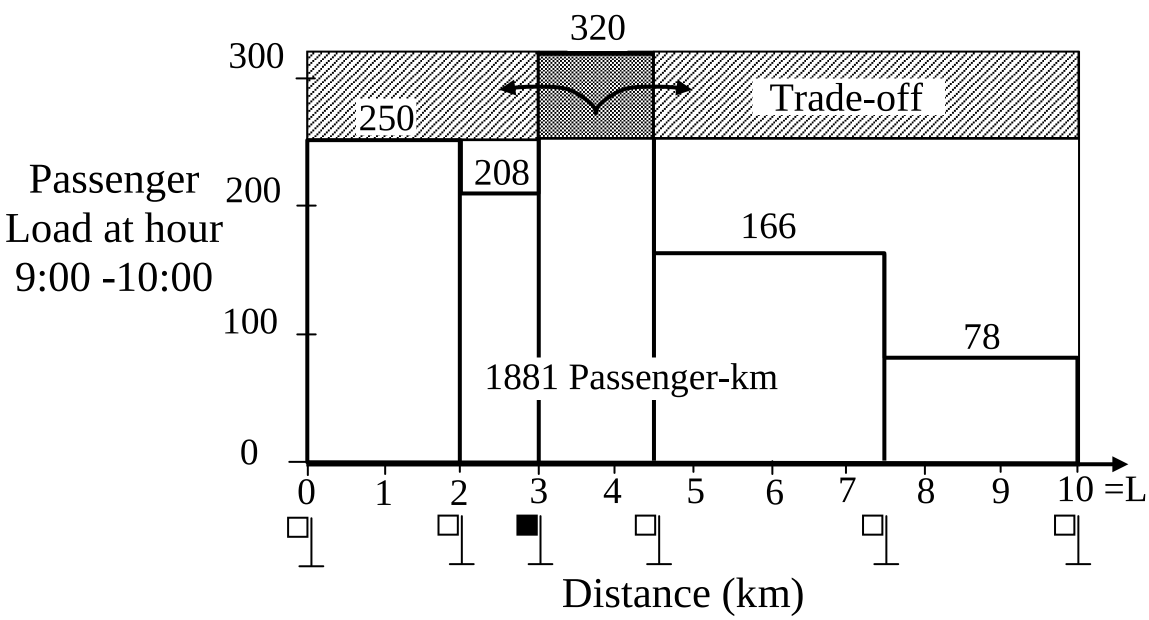

Figure 5 Load profile from the example in Figure 2 (9:00 -10:00) using Method 4 (20%), with an indication of trade-off between this method and Method 2 (more crowding in return for less empty space-km)

The considered load will be determined by the product $\frac{P_{mj}}{c} \cdot d_{oj}$. Table 2 shows the results for Methods 3 and 4, in which the percentage of route length allowed to have an excess load in Method 4 is set at 10%, 20%, and 30%.

We may observe that in the Method 3 results in Table 2, the first hour relies on $\frac{P_{mj}}{c}$ or, specifically, 400/90=4.44 veh/hr. When we turn to Method 4 for this first hour, the results of Method 2 for 10% are attained, since the Max load stretches along more than 10% of the route length. For 20% and 30%, the vehicle-capacity constraint still governs.

Table 2 Frequency and headway results for the example in Figure 2 for Methods 3 and 4

Period j	**Method 3**		**Method 4**					
			10%		**20%**		**30%**	
	F_{3j} (veh/hr)	**H_{3j}** (min)	**F_{4j}** (veh/hr)	**H_{4j}** (min)	**F_{4j}** (veh/hr)	**H_{4j}** (min)	**F_{4j}** (veh/hr)	**H_{4j}** (min)
6:00-7:00	4.44	14	8.00	7.5	4.44	14	4.44	14
7:00-8:00	5.88	10	8.40	7	8.40	7	6.70	9
8:00-9:00	6.40	9	8.00	7.5	7.00	9	7.00	9
9:00-10:00	3.72	16	6.40	9	5.00	12	5.00	12
10:00-11:00	3.07	20	4.40	14	4.40	14	4.00	15

In the second hour, 7:00-8:00, the following is obtained for Method 3: F_3=max(2942/50·10, 510/90, 3) = 5.88 veh/hr. Continuing in the second hour for Method 4 (10%), the average Max load of 294.2 passengers rises to 420, resulting in F_4=8.40 veh/hr. Table 2 continues to be fulfilled in the same manner. It should be noted that, as in Table 1, all the headways are rounded to their nearest integer.

Although we aim at a resource saving using Methods 3 and 4, there is a question as to whether this saving justifies the additional expense involved in using ride check as opposed to point check. The next section attempts to answer this question by constructing a criterion suggesting when to use the point check or, otherwise, the ride-check data-collection technique.

4. Alternative Timetables

4.1 Optional Timetables

Three categories of options may be identified: (i) type of timetable, (ii) method or combination of methods for setting frequencies, and (iii) special requests. These three groups of options are illustrated in Figure 6. A selected path in this figure provides a single timetable. Hence, there are a variety of timetable options.

The first category in Figure 6 concerns alternative types of timetables. The even-headway type simply means constant time intervals between adjacent departures in each time period, or the case of evenly spaced headways. Even average load refers to unevenly spaced headways in each time period, but the observed passenger loads at the hourly Max load point are similar on all vehicles. A second type of timetables entails situations in which even headways will result in significantly uneven loads. Such uneven-load circumstances occur, for example, around work and school dismissal times, but they may in fact, occur on many other occasions. Figure 6 shows that the average even load can be managed either at the hourly Max load point (even loads on all vehicles at that point) or at each individual vehicle's Max load point. The average even load at the hourly Max load point type is dealt with in this section, and the other type in Ceder (2007).

In the second category of options, it is possible to select different frequency or headway-setting methods. This category allows for the selection of one method or for combinations of methods for different time periods. The methods considered, and indicated in Figure 6, are the two-point check and the two-ride check, both described in the previous sections. In addition, there might be procedures used by the planner/scheduler that are not based on data, but on observations made by road supervisors and inspectors or other sources of information.

In the third category of selections, we allow for special scheduling requests. One characteristic of existing timetables is the repetition of departure times, usually every hour. These easy-to-memorize departure times are based on so-called *clock headways*: 6, 7.5, 10, 12, 15, 20, 30, 40, 45, and 60 minutes. Ostensibly, headways less than or equal to 5 minutes are not thought to influence the timing of passenger arrivals at a transit stop. The clock headway is obtained by rounding the derived headway down to the nearest of these *clock* values. Consequently, and similar to the "rounding off upward" of frequencies, clock headways require a higher number of departures than what is actually necessary to meet the demand.

A second possible special request is to allow the scheduler to predetermine the total number of vehicle departures during any time period. This request is most useful in crises, when the agency needs to supply a working timetable for an operation based on tightly limited resources (vehicles and/or crews). By controlling the total number of departures while complying with other requests, the scheduler achieves better results than by simply dropping departures without any systematic procedure. Furthermore, there might be cases in which the agency would like to increase the level of service by allowing more departures in the belief that passenger demand can be increased by

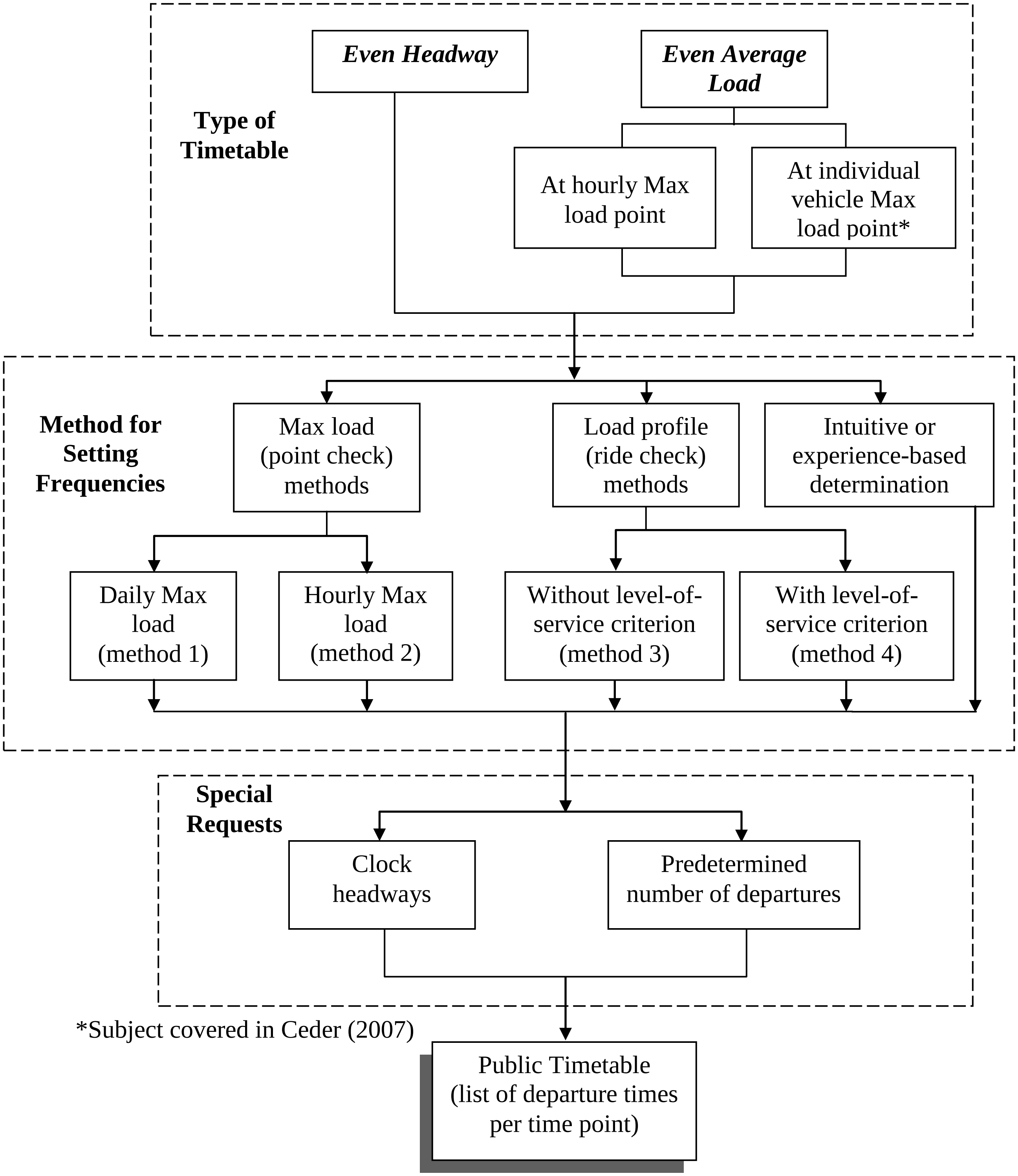

Figure 6 Optional public timetables

providing improved (more frequent) service. Certainly, this special request can also be approached through varying the desired occupancy (load factor) standard; however, this option can be a compulsory standard.

Finally, it should be emphasized that not all the paths concerning clock headways in Figure 6 are meaningful. The selection of the even average load type of timetable cannot be performed if there is a clock-headway constraint. Moreover, the number of departures cannot be predetermined for clock headways because of the specific time restrictions on those headways.

4.2 Comparison Measures

With computerized timetable construction, the transit agency can assess a variety of optional timetables rather than being limited to examining one or a few. Two interrelated measures may be useful for the agency to compare optional timetables: (i) number of required runs (departures); and (ii) required single-route fleet size.

The first comparison measure, total number of departures, can serve as an indicator of the number of vehicles required and also whether or not it is possible to save vehicle runs.

The second comparison measure refers to each route separately and provides an estimate of the required fleet size at the route level. In a large transit agency, an efficient arrangement of vehicle blocks includes interlining (switching a vehicle from one route to another) and deadheading trips. Hence, fleet size is not determined at the route level, but at the network level. The second comparison measure, however, is based on a simple formula derived by Salzborn (1972) for a continuous time function and explicitly shown by Ceder (2007) for discrete time points. This formula states that if T is the round-trip time, including layover and turn- around time, then the minimum fleet size is the largest number of vehicles departing at any time interval during T. This value is adequate for a single route with a coinciding departure and arrival location. Consequently, the second comparison measure can be used for each direction separately, as well as for both directions when selecting the maximum of two derived values. This single-route fleet-size formula is elaborated in Ceder (2007).

5. Even Headways with Smooth Transitions

One characteristic of existing timetables is the repetition of the same headway in each time period. However, a problem facing the scheduler in creating these timetables is how to set departure times in the transition segments between adjacent time periods. This section addresses the issue.

5.1 Underlying Principle

A common headway smoothing rule in the transition between time periods is to use an average headway. Many transit agencies employ this simple rule, but it may be shown that it can result in either undesirable overcrowding or underutilization. For example, consider two time periods, 06:00-07:00 and 07:00-08:00, in which the first vehicle is predetermined to depart at 06:00. In the first time period, the desired occupancy is 50 passengers, and in the second 70 passengers. The observed maximum demand to be considered in these periods is 120 and 840 passengers, respectively. These observed loads at a single point are based on the uniform passenger-arrival-rate assumption. The determined frequencies are 120/50 = 2.4 vehicles and 840/70 = 12 vehicles for the two respective periods, and their associated headways are 25 and 5 minutes, respectively. If one uses the common average headway rule, the transition headway is (25 + 5)/2 = 15 minutes; hence, the timetable is set to 06:00, 06:25, 06:50, 07:05, 07:10, 07:15,…, 07:55, 08:00. By assuming a uniform passenger arrival rate, the first period contributes to the vehicle departing at 07:05 the average amount of (10/25)·50 = 20 passengers at the Max load point; the second period contributes (5/5)·70 = 70 passengers. Consequently, the expected load at the Max load point is 20 + 70 = 90, a figure representing average overcrowding over the desired 70 passengers after 7:00. Certainly, the uniform arrival-rate assumption does not hold in reality. However, in some real-life situation (e.g., after work and school dismissals), the observed demand in 5 minutes can be more than three times the observed demand during the previous 10 minutes, as is the case in this example. In order to overcome this undesirable situation, the following principle, suggested by Ceder (2007), may be employed.

Principle 1: Establish a curve representing the cumulative (non-integer) frequency determined versus time. Move horizontally for each departure until intersecting the cumulative curve, and then vertically; this will result in the required departure time.

Proposition 1: Principle 1 provides the required evenly spaced headways with a transition load approaching the *average desired occupancies* of d_{oj} and $d_{o(j+1)}$ for two consecutive time periods, j and j+1.

***Proof*:** Figure 7 illustrates Principle 1. Since the slopes of the lines are 2.68 and 3.60 for j = 1 and j = 2, respectively, the resultant headways are those required. The transition load is the load associated with the 7:05 departure, which consists of arriving passengers during 16 minutes for j = 1, and of arriving passengers during 5 minutes for j = 2. Therefore, (16/22)·50 + (5/17)·60 = 54 approximately. This transition load is not the exact average between d_{o1} = 50 and d_{o2} = 60, since departures are made in integer minutes. That is, the exact determined departure after 7:00 is (3-2.68)·60/3.60 = 5.33 minutes. Inserting this value, instead of the 5 minutes mentioned above, yields a value that is closer to the exact average. Basically, the proportions considered satisfy the proof-by-construction of Proposition 1.

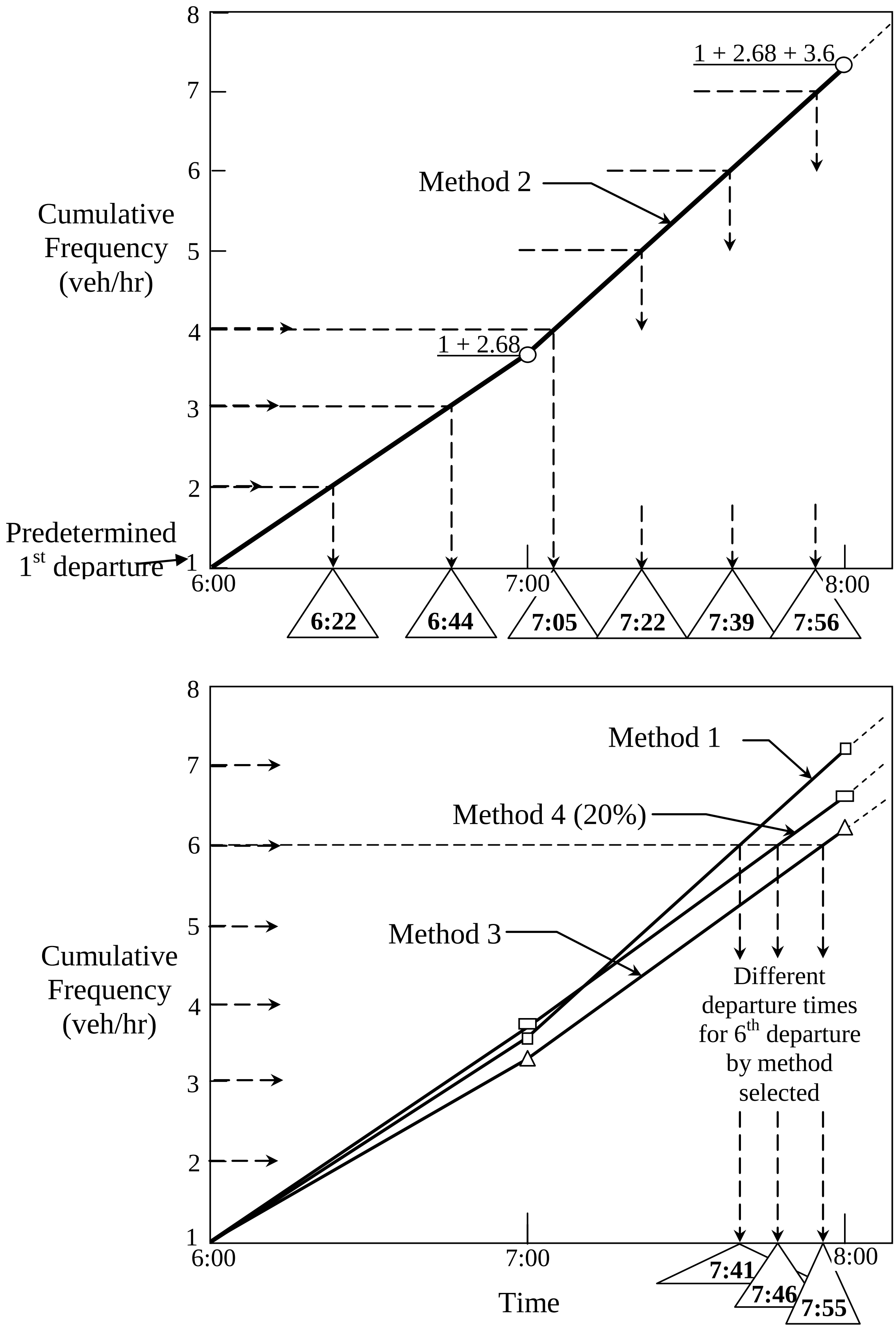

Figure 7 Determination of departure times for evenly spaced headways

6. Headways with Even Average Loads

This section opens with the following premise: transit managers/planners/schedulers who believe that problems related to attracting more transit users and reliability problems are drowned in the "ocean" of even-headway timetables should be told that these problems know… how to swim. In other words, even-headway timetables do not necessarily deliver the merchandise (satisfactory transit service) to the customer (passengers).

We have already noted that passenger demand varies even within a single time period, hence resulting for even headways in an imbalanced load on individual vehicles at the hourly Max load point. On heavy-load routes and short headways, the even-headway timetable suffices. However, in the course of reducing reliability problems, we may occasionally prefer to use the even-load instead of the even-headway procedure. Moreover, the availability of APCs (automatic passenger counters) provides a framework in which to investigate systematically the variation in passenger demand. With the anticipated vast amount of passenger load data, we can then better match vehicle departure times with variable demand. Two procedures carry out this endeavor. The first, addressed in this section, deals with average even load on individual vehicles at the hourly (or daily) Max load point. The second procedure, addressed in Ceder (2007), ensures an average even load at each individual vehicle Max load point.

6.1 Underlying Principle

A simple example is presented here to illustrate the underlying load-balancing problem. Consider an evenly spaced headway timetable in which vehicles depart every 20 minutes between 07:00 and 08:00; i.e., at 07:20, 07:40, and 08:00. The observed load data consistently show that the second vehicle, which departs at 07:40, has significantly more passengers than the third vehicle. The observed (average) Max load during this 60-minute period is 150 passengers, and the desired occupancy is 50 passengers. Hence, using Method 2, three vehicles are required to serve the demand as in the case of the evenly spaced headways timetable. The average observed loads at the hourly Max load point on the three vehicles are 50, 70, and 30 passengers, respectively. Given that these average loads are consistent, then the transit agency can adjust the departure times so that each vehicle has a balanced load of 50 passengers on the average at the hourly Max load point. The assumption of a uniform passenger- arrival rate results in 70/20 = 3.5 passengers/minute between 07:20 and 07:40, and 30/20 = 1.5 passengers/minute between 07:40 and 08:00. If the departure time of the second vehicle is shifted by X minutes backward (i.e., an early departure), then the equation 3.5X = 70-50 yields the balanced schedule, with X = 5.7 ≈ 6 minutes, or departures at 07:20, 07:34, and 08:00. The third departure will add this difference of 20 passengers at the hourly Max load point. The even-headway setting assures enough vehicles to accommodate the hourly demand, but it cannot guarantee balanced loads for each vehicle at the peak point. In order to avoid this imbalanced situation, the following principle should be exploited.

Principle 2: Construct a curve representing the cumulative loads observed on individual vehicles at the hourly Max load points. Move horizontally per each d_{oj} for all j, until intersecting the cumulative-load curve, and then vertically; this results in the required departure times.

Proposition 2: Principle 2 results in departure times such that the average Max load on individual vehicles at the hourly j^{th} Max load point approaches the desired occupancy d_{oj}.

Proof: Figure 8 illustrates Principle 2. Method 2 will be used in the upper part of Figure 8 in which the derived departure times are unevenly spaced to obtain even loads at stop 3 for j = 1 and at stop 2 for j = 2. These even loads are constructed on the cumulative curve to approach d_{o1}=50 and d_{o2}=60. If we assume a uniform passenger-arrival rate between each two observed departures, it can be shown that the load (at stop 3) of the first derived departure (6:23) consists of the arrival rate between 6:00 and 6:15 (35/15 = 2.33) and the rate between 6:15 and 6:50 (65/35 = 1.86). Thus, $2.33 \cdot 15 + 1.86 \cdot 8 \approx 50$. In the transition between j = 1 and j = 2 (in the upper part of Figure 8), the value of d_2 = 60 is considered, since the resultant departure comes after 7:00. The load of the vehicle departing at 7:07 at its hourly Max load point, stop 2, is simply $17 \cdot (90/25) = 61.2$ from rounding off the departure time to the nearest integer. That is, $(10+y) \cdot (90/25) = 60$ results in y = 6.67 minutes. This completes the proof-by-construction of Proposition 2.

6.2 Further studies

Two further works worth mentioning. The first by Hassold and Ceder (2012) uses two simultaneous objectives: minimizing the expected passenger waiting time and minimizing the discrepancy from a desired occupancy level on the vehicles. A network-based procedure is used to create timetables with multiple vehicle types to solve this bi-objective problem. The methodology developed was applied to a case study in Auckland, New Zealand and results in a saving of more than 43% of passenger waiting time where, at the same time, attaining an acceptable passengers' load on all vehicles.

The second work by Ceder et al (2013) proposes a multi-objective methodology to create bus timetables using multiple vehicle sizes, and has two objectives carried out simultaneously: First, minimize the deviation of the determined headways from a desired even headway and second, minimize the deviation of the observed passenger loads from a desired even-load level of the vehicles at the maximum-load point. The suggested methodology uses a graphical heuristic approach to examine different strategies in the creation of the optimal timetables.

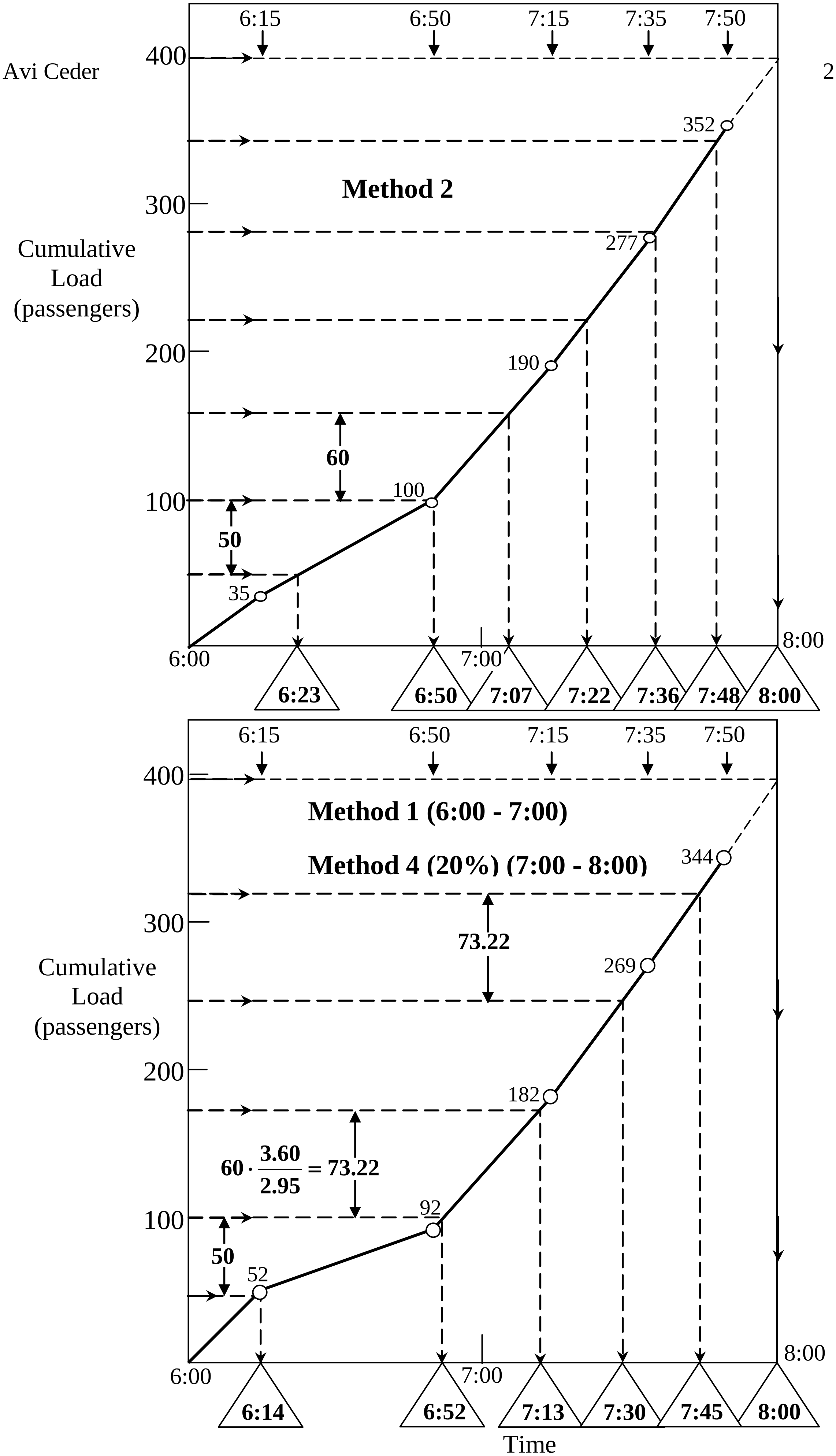

Figure 8 Determination of departure times with even loads

7. Optimal Vehicle Schedules

7.1 Background

Figure 1 presents the public transport operations planning framework as a multistep process. Due to the complexity of this process each step is normally conducted separately, and sequentially fed into the other. In order for this process to be cost-effective and efficient, it should embody a compromise between passenger comfort and cost of service. For example, a good match between vehicle supply and passenger demand occurs when vehicle schedules are constructed so that the observed passenger demand is accommodated while the number of vehicles in use is minimized. Following the construction of an adequate public timetable above, the next step is to determine vehicle schedules or chains of trips carried out by individual vehicles so as to reach the minimum number of vehicles required to cover the entire timetables. It is assumed that each vehicle has the same number of seats and same capacity (seats plus standees). This section provides an overview on exact solutions to the vehicle scheduling problem and describes a graphical heuristic procedure for the determination of minimum fleet size and its lower bound.

7.2 Deficit Function (DF) approach

Following is a description of a step function approach described by Ceder and Stern (1981), for assigning the minimum number of vehicles to allocate for a given timetable. The step function is called **Deficit Function (DF)** as it represents the deficit number of vehicles required at a particular terminal in question in a multiterminal transit system. That is, DF is 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. To construct a set of deficit functions, the only information needed is a timetable of required trips. The main advantage of the DF is its visual nature. Let $d(k,t,S)$ denote the DF for the terminal k at the time t for the schedule S. The value of $d(k,t,S)$ represents the total number of departures minus the total number of trip arrivals at terminal k, up to and including time t. The maximal value of $d(k,t,S)$ over the schedule horizon $[T_1,T_2]$ is designated $D(k,S)$.

Let t_s^i and t_e^i denote the start and end times of trip i, $i \in S$. It is possible to partition the schedule horizon of $d(k,t,S)$ into sequence of alternating hollow and maximal intervals. The maximal intervals $[s_i^k, e_i^k], i = 1,...,n(k)$ define the interval of time over which $d(k,t)$ takes on its maximum value. Note that the S will be deleted when it is clear which underlying schedule is being considered. Index i represents the *i*th maximal intervals from the left and $n(k)$ represents the total number of maximal intervals in $d(k,t)$. A hollow interval H_l^k, l=0,1,2,…,$n(k)$ 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 $\left(T_1 \text{ or } T_2\right)$, the graphical representation of $d\left(k,t\right)$ is emphasized by clear dot.

If the set of all terminals is denoted as *T*, the sum of $D\left(k\right)$ for all $k \in T$ is equal to the minimum number of vehicles required to service the set *T*. This is known as the fleet size formula. Mathematically, for a given fixed schedule *S*:

$$D\left(S\right) = \sum_{k \in T} D\left(k\right) = \sum_{k \in T} \max_{t \in \left[T_1, T_2\right]} d\left(k,t\right) \qquad (6)$$

where $D\left(S\right)$ is the minimum number of buses to service the set *T*.

When Deadheading (DH) or empty trips are allowed, the fleet size may be reduced below the level described in Equation (6). Ceder and Stern (1981) described a procedure based on the construction of a Unit Reduction DH Chain (URDHC), which, when inserted into the schedule, allows a unit reduction in the fleet size. The procedure continues inserting URDHCs until no more can be included or a lower boundary on the minimum fleet is reached. The lower boundary *G(S)* is determined from the overall deficit function defined as $g\left(t,S\right) = \sum_{k \in T} d\left(k,t,S\right)$ where $G(S) = \max_{t \in \left[T_1, T_2\right]} g\left(t,S\right)$. This function represents the number of trips simultaneously in operation. Initially, the lower bound was determined to be the maximum number of trips in a given timetable that are in simultaneous operation over the schedule horizon. Stern and Ceder (1983) improved this lower bound, to $G\left(S'\right) > G\left(S\right)$ based on the construction of a temporary timetable, *S'*, in which each trips is extended to include potential linkages reflected by DH time consideration in *S*. This lower bound is further improved in this work.

The algorithms of the deficit function theory are described in detail by Ceder and Stern (1981). However, it is worth mentioning the next terminal (NT) selection rule and the URDHC routines. The selection of the NT in attempting to reduce its maximal deficit function may rely on the basis of garage capacity violation, or on a terminal whose first hollow is the longest, or on a terminal whose overall maximal region(from the start of the first maximal interval to the end of the last one) is the shortest. The rationale here is to try to open up the greatest opportunity for the insertion of the DH trip. In the URDHC routines there are four rules: R=0 for inserting the DH trip manually in a conversational mode, R=1 for inserting the candidate DH trip that has the minimum travel time, R=2 for inserting a candidate DH trip whose hollow starts farthest to the right, and R=3 for inserting a candidate DH trip whose hollow ends farthest to the right. In the automatic mode (R=1,2,3), if a DH trip cannot be inserted and the completion of a URDHC is blocked, the algorithm backs up to a DH candidate list and selects the next DH candidate on that list.

Figure 9 presents an example with 9 trips and four terminals (a, b, c, and d). In its upper part the 9 trips are shown with respect to time with departure and arrival terminals. Note that trip 4 starts and ends in the same terminal. Four DFs are constructed along with

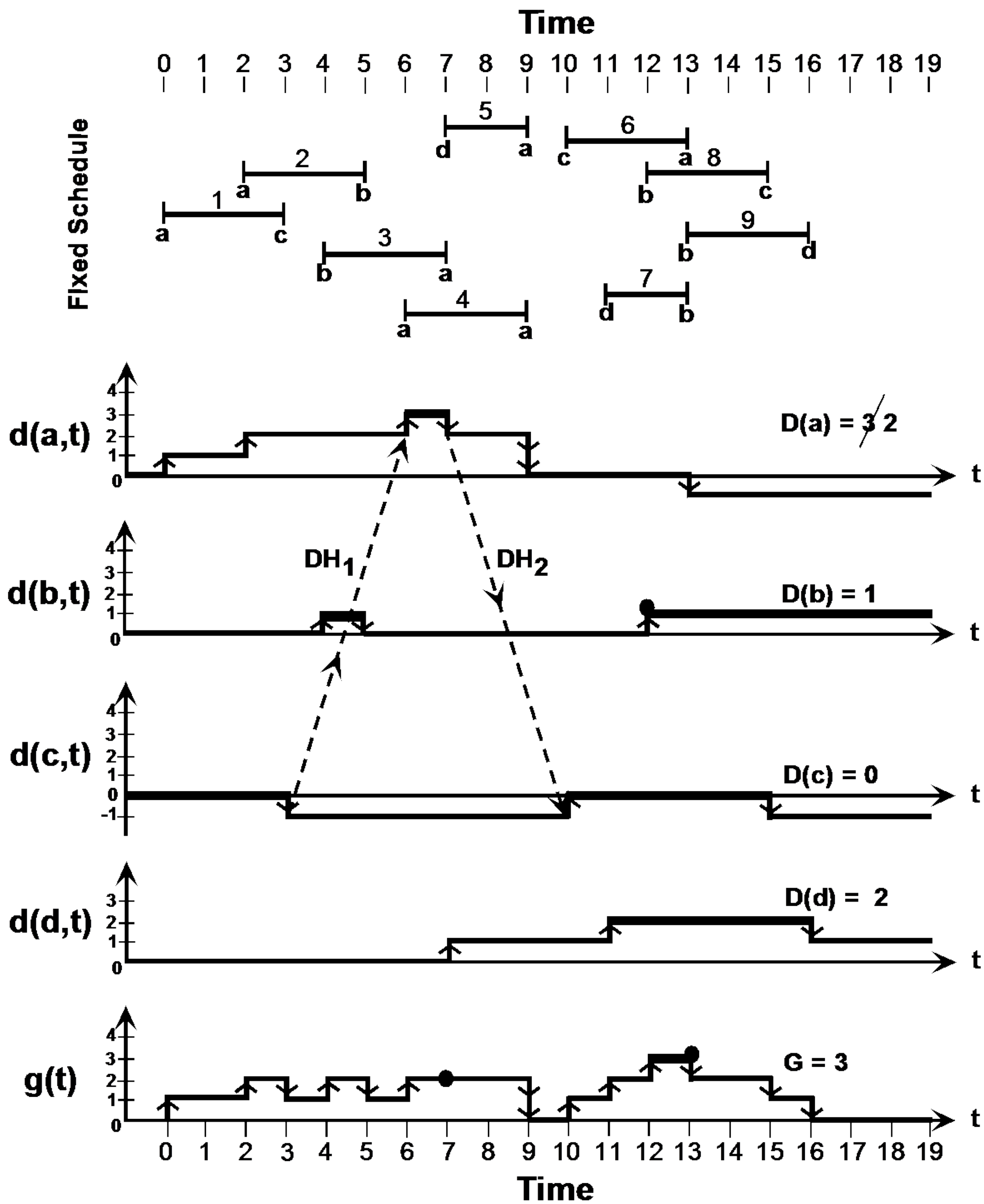

Figure 9. Nine-trip example with deadheading trip insertion for reducing the fleet size.

the overall DF. The maximal areas of the DFs are emphasized with a heavy line. Based on the NT procedure terminal a (whose maximal region is the shortest) is selected for possible reduction in *D(a)*. Given that all DH times are 3 units of time, and using R=2, a DH trip is inserted from terminal c to a, DH_1. This will increase *d(c,t*) at t=3 from -1 to 0, reduce *d(a,t*) at t=6 from 3 to 2, but will also increase *d(c,t)* at t=10 from 0 to 1. In order to eliminate the increase of *D(c)* from 0 to 1 another DH trip is inserted, DH_2 from a to c. The result is that *D(a)* is reduced from 3 to 2, and the DFs of a and c are updated with *d(a,t*)=2 between t=6 and t=7, and *d(c,t*)=0 between t=3 and t=10. One can see that no more DH trips (with trip time of 3 units) can be further inserted to reduce *D(k*), k=a,b,c,d. Hence *D(S*)=5. The sum of all the DFs, *g(t*), is illustrated at the bottom of Figure 9 and has *G*=3 (maximal number of vehicles in simultaneous operation). It will be used in a following section for the lower bound improvement.

Finally, all of the trips, including the DH trips, are chained together for constructing the vehicle schedule (blocks). Two rules can be applied for creating the chains: first in-first out (FIFO), and a chain-extraction procedure described by Gertsbach and Gurevich (1977). The FIFO rule simply links the arrival time of a trip to the nearest departure time of another trip (at the same location), and continues to create a schedule until no connection can be made. The trips considered are deleted and the process continues. The chain-extraction procedure allows an arrival-departure connection for any pair within a given hollow (on each deficit function). The pairs considered are deleted and the procedure continues. Figure 10 illustrates for clarity one hollow (between two peaks of the deficit function) with arrivals of trips 1, 2, 3 and departures of trips 4, 5, 6. Below the figure there is the FIFO chain (within this hollow) as well as other alternatives, where in all- the minimum the fleet size is maintained.

The initial lower bound on the fleet size with DH trip insertions was proved by Ceder and Stern (1981) to be *G.(S)*. An improved lower bound of this problem was established and proved later by Stern and Ceder (1983), and Ceder (2007) using the following procedure:

1. extend each trip's arrival time to the time of the first feasible departure time of a trip with which it may be linked to T_2 (the ending time of the finite time horizon).
2. given that the extended schedule is *S'*, construct the overall DF, *g'(t,S')*, and determine its maximum value as *G'(S')*.

While creating *S'* it is possible that several trips' arrival points will be extended forward to the same departure point being their first feasible connection. Nonetheless in the final solution of the minimum fleet size problem only one of these extensions will be linked to the single departure point. This observation opens an opportunity to look into further artificial extensions of certain trips' arrival points without violating the generalization of all possible combinations needed to prove that the resultant boundary on the fleet size is its lower bound. A stronger lower bound than *G'(S'*) is found and proved in Ceder (2007). The stronger the lower bound is, the closer it is to the minimum fleet size required. Also, the stronger the lower bound is the better it serves the public transport decision makers on how far the fleet size can be reduced via DH trip insertions.

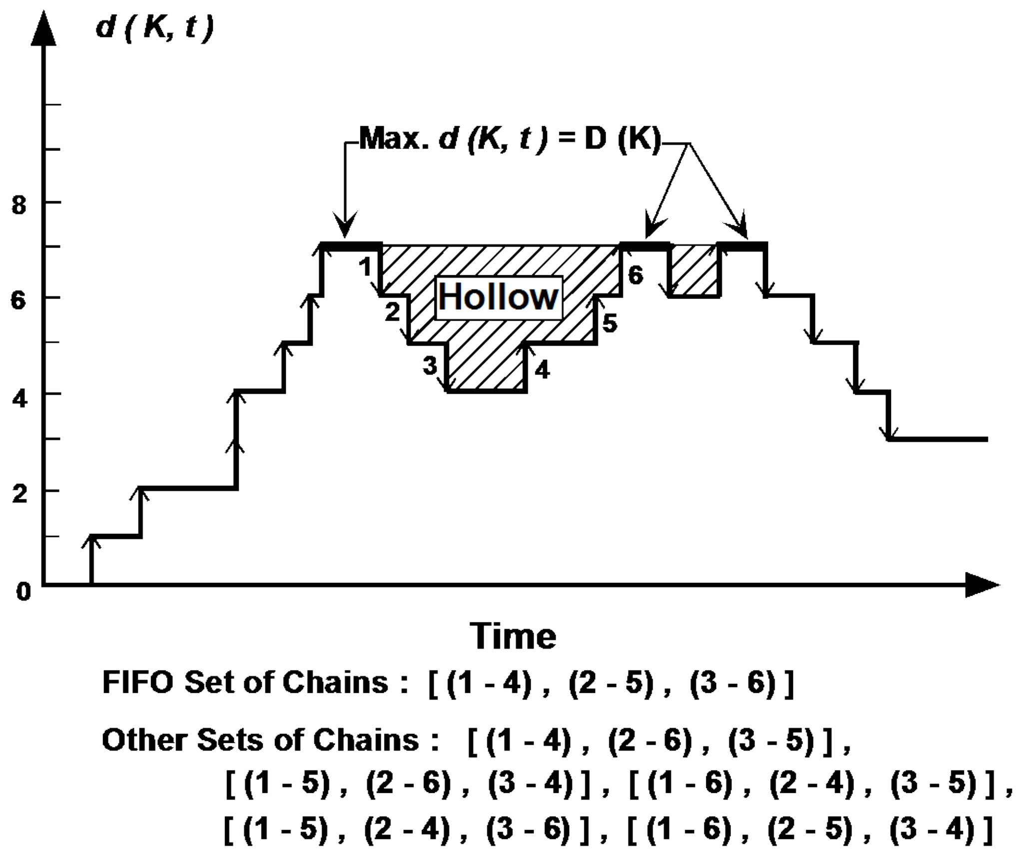

Figure 10. An example of creating chains of trips within a hollow using FIFO rule and all other possibilities.

Figure 11 presents the schedule of Figure 9 with *S'* in its upper part and two overall deficit functions: *g(t)* and *g'(t)*. All trips in *S'* are extended either to their first feasible connection (with all DH times are 3 units of time) or to the time horizon, t=18. The improved lower bound is therefore *G'(S'*) = 4.

7.3 Shifting Departure Times with Given Tolerances

Another factor considered in a manually produced public transport schedule is related to the shifting of trip departure times. A general description of a technique to reduce the fleet size for a variable departure time scheduling problem can be found in Gerstbach and Stern (1978). This technique for job schedule utilizes the deficit function representation as a guide for local minimization in maximal intervals, $M_r^u \forall u \varepsilon T$. However when considering variable departure times along with a possible insertion of DH trips, the problem becomes more complex. The scheduler who performed shifting in trip departure times is not always aware of the consequences which could arise from these shifts. Ceder (2007) describes a formal algorithm to handle the complexities of shifting departure times. The algorithm is intended for both automatic and man-computer conversational modes.

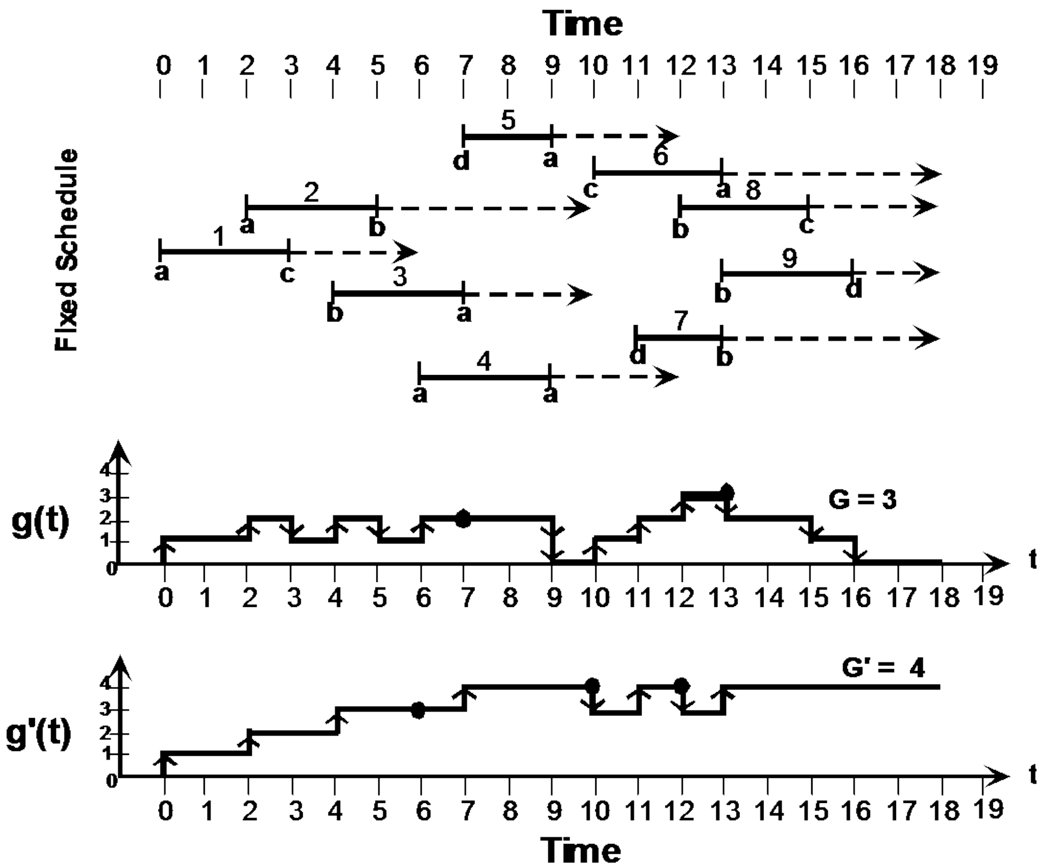

Figure 11. The example of figure 5 with artificial extensions of each trip to its first feasible connection which results with the improved lower bound, G'=4

Figure 12 illustrates an example of two terminals and seven trips using the DH representation in part (i). Part (ii) shows how to reduce the fleet size using shifting tolerances of ½ time unit (forward or backward) where the shifts are shown with small arrows and the update DF is marked by a dashed line. Part (iii) shows how to apply only the URDHC procedure with DH times of 2 time units, and Part (iv) presents a modified URDHC (mixed with the shifting) procedure. As can be seen in Figure 12, Part (i), the fixed schedule without DH considerations requires 5 vehicles. Using shifting allows for reducing the number of vehicles to 3. The use of URDHC allows for reducing it to 4, and the use of a combined approach requires 3 vehicles.

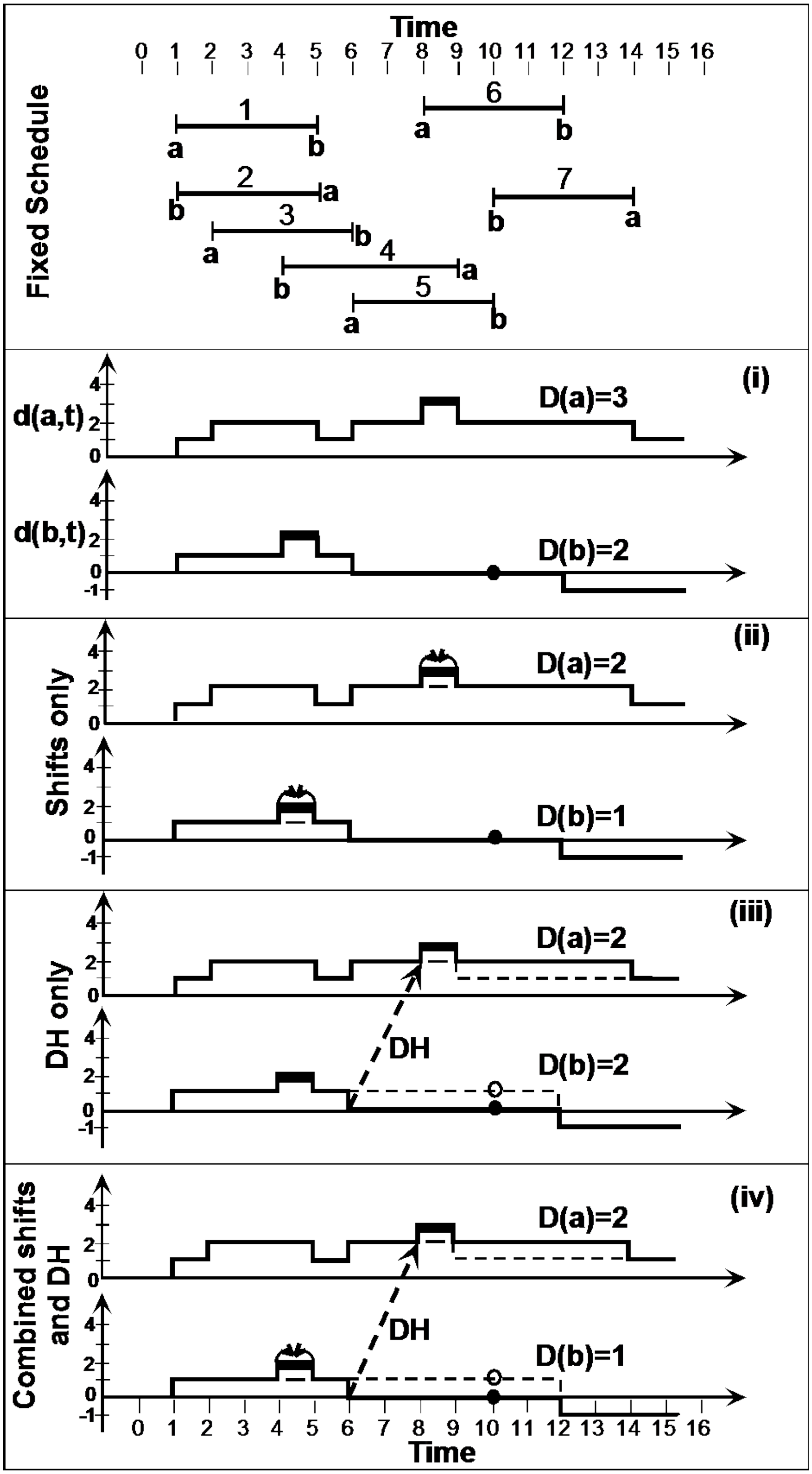

Figure 12. An example with seven trips and two terminals using three procedures.

The viewpoint of the public transport operator will lead to use first the shifting procedure while wishing to minimize the operational cost (reducing DH mileage). However there is also the issue of passenger comfort while trying to accommodate the observed demand. Changes in departure times may result in imbalance passenger loads and reduction in the service reliability. Past experience in applying the DF approach at several bus properties shows that best is to first identify small shifts in departure times, enabling the reduction of the fleet size, without noticeable changes in the timetable. Second is to apply the combined approach of URDHC and shifting departure times.

8. References

Ceder, A. (2007). *Public Transit Planning and Operation: Theory, Modeling and Practice.* Elsevier, Butterworth-Heinemann, Oxford, UK.

Ceder, A. and Stern, H.I. (1981). Deficit function bus scheduling with deadheading trip insertion for fleet size reduction. *Transportation Science,* **15 (4)**, 338-363.

Ceder, A., Hassold, S. and Dano, B. (2013). Approaching even-load and even-headway transit timetables using different bus sizes. *Public Transport – Planning and Operation*; DOI: 10.1007/s12469-013-0062-z**.**

Daduna, J.R. and Wren, A. (eds.) (1988). *Computer-Aided Transit Scheduling*. Lecture Notes in Economics and Mathematical Systems, **308**, Springer-Verlag.

Daduna, J.R., Branco I., and Paixao, J.M.P. (eds.) (1995). *Computer-Aided Transit Scheduling.* Lecture Notes in Economics and Mathematical Systems **430**, Springer-Verlag.

Desrochers, M. and Rousseau, J.M. (eds.) (1992). *Computer-Aided Transit Scheduling*. Lecture Notes in Economics and Mathematical Systems **386**, Springer-Verlag.

Gertsbach, I. and Gurevich, Y. (1977). Constructing an optimal fleet for transportation schedule. *Transportation Science,* **11**, 20-36.

Gertsbach, I. and Stern, H.I. (1978). Minimal resources for fixed and variable job schedules. *Operations Research*, **26**, 68-85.

Hassold, S. and Ceder, A. (2012). Multiobjective Approach to Creating Bus Timetables with Multiple Vehicle Types. *Journal of the Transportation Research Board*, **2276**, 56-62.

Hickman, M., Voss, S. and Mirchandani, P. (eds.) (2008). *Computer-Aided Systems in Public Transport*. Lecture Notes in Economics and Mathematical Systems, Vol. 600, Springer-Verlag.

Proceedings, The 10th International Conference on Advanced Systems for Public Transport (CASPT*)* (2006). *Chair of organizing committee*: Raymond Kwan. Leeds, UK, July 21-23.

Proceedings, The 11th International Conference on Advanced Systems for Public Transport (CASPT), (2009). *Organizing committee*: Hong K. Lo (Chair), William Lam, SC Wong, Janny Leung. Hong Kong, July 20-22.

Proceedings, The 12th International Conference on Advanced Systems for Public Transport (CASPT), (2012). *Organizing committee*: Juan Carlos Muñoz (Chair), Ignacia Torres, Patricia Galilea, Felipe Delgado, Joaquín De Cea, Juan Carlos Herrera, Ricardo Giesen. Santiago, Chile, July 23-27.

Rousseau, J.M. (ed.) (1985). *Computer Scheduling of Public Transport 2*. North-Holland.

Salzborn. F. J. M. (1972). Optimum bus scheduling. *Transportation Science,* **6**, 137-148.

Stern, H.I. and Ceder, A. (1983). An improved lower bound to the minimum fleet size problem. *Transportation Science,* **17 (4)**, 471-477.

Voss, S. and Daduna, R. (eds.) (2001*). Computer-Aided Scheduling of Public Transport.* Lecture Notes in Economics and Mathematical Systems **505**, Springer-Verlag.

Wilson, N.H.M. (ed.) (1999). *Computer-Aided Scheduling of Public Transport*. Lecture Notes in Economics and Mathematical Systems **471**, Springer-Verlag.

Wren, A. (ed.) (1981). *Computer Scheduling of Public Transport*: *Urban Passenger Vehicle and Crew Scheduling.* North Holland.

End of subject