A DECISION SUPPORT SYSTEM FOR TIMETABLE ADJUSTMENTS

João Mendes-Moreira
Dept. of Informatics Engineering, Faculty of Engineering, University of Porto, Portugal
LIAAD-INESC Porto L.A., Portugal
E-mail: jmoreira@fe.up.pt

Eva Duarte
Department of Informatics, School of Engineering, University of Minho, Portugal
E-mail: evamcduarte@gmail.com

Orlando Belo
Department of Informatics, School of Engineering, University of Minho, Portugal
E-mail: obelo@di.uminho.pt

1 INTRODUCTION

For a public transport company, the process of defining trips offers is a central task because trips are the main product they have to offer to their clients. As in other business areas the offer should maximize clients' satisfaction at a minimum cost. Traditionally, timetables were defined assuming a deterministic travel time. However, with the investments done in the last decade in Advanced Public Transportation Systems, a large amount of current data obtained from Automatic Vehicle Location systems is now available. This data can be used to enhance travel time modelling for timetable definition. This is the subject of this paper, which imposed us the study of how to use current data in order to better define timetables aiding public transport companies to accomplish their mission. We present a Decision Support System (DSS) for the special case of timetable adjustments, assuming therefore that the schedule under study is not new, i.e., there are actual trips for that schedule.

Firstly, in this paper, we present a brief state of the art review on defining travel times for timetabling (Sect. 2) then we discuss the reasons for developing a DSS for timetable adjustments (Sect. 3). In Sect. 4 we describe the DSS, and in Sect. 5 how to use it in typical situations. We conclude with a discussion and guidelines for future research (Sect. 6).

2 RELATED WORKS ON TRAVEL TIMES FOR TIMETABLING

An important difference between the existing approaches on timetable creation concerns the variables used. These variables depend on the purpose. If, for instance, a timetable for a new line is needed, a variable such as the population density of the served area is important [3]. However, if the goal is to make small adjustments to the timetables, this variable is not relevant. We describe the variables for the latter case, i.e., for regular planning tasks (line creation is a sparse event in a public transport company). Some typical variables are: scheduled travel time (STT), slack time (SlT), scheduled headway (SH), fleet size (N) and scheduled departure time. Let us assume that SCT is the scheduled cycle time,

$$SCT = STT_g + SlT_g + STT_r + SlT_r, \quad (1)$$

where the indexes $_g$ and $_r$ represent the go and return trips (Fig. 1).

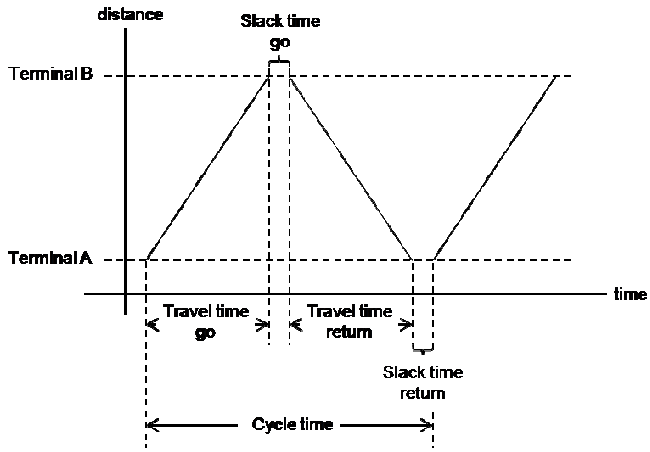

Figura 1: Time-distance chart on different concepts of travel time.

For urban areas, the departure times are usually defined by headway instead of irregularly spaced. Irregularly spaced departure times are typically used for long distance trips or trips in rural areas. We focus on the definition of the timetables' departure times by headway.

According to [10],

$$SCT = N \times SH. \quad (2)$$

Fixing N, and assuming that travel times are exponentially distributed, slack times can be optimized [10]. Using this approach, the shorter the slack time is, the shorter the scheduled headway is. The objective function used is the passengers' expected waiting time function,

$$E[\omega] = \frac{SH}{2} \times \left(1 + \frac{2Var[l]}{SH^2}\right), \quad (3)$$

where l represents the delay of the bus and $Var[l]$ the variance of l. This function assumes that passengers arrive uniformly at the bus stops, which is acceptable for short headways. Headways equal or shorter than 10 minutes are usually defined as short headways in the literature [10].

However, Zhao et al. [10] argue that by using the function defined in [1] for the passengers' arrival at the bus stops, it is possible to adapt the solution to problems with large headways. In this case,

$$f_A(t) = \exp\big(U(t)\big) \, / \int_0^{SH} \exp\big(U(\tau)\big)\, d\tau, \tag{4}$$

where $U(t) = aE[\omega(t)]^b$ is the utility function, $E[\omega(t)]$ is the expected waiting time of an arrival at time t and a and b are constants that must be defined from empirical data. The derivation of the passengers' expected waiting time for large headways has not yet been done.

Using an economic perspective, Carey defines as objective function, the cost expressed in terms of STT, lateness and earliness unit costs [2]. Using this approach it is possible to define the optimal STT and SlT for given ratios between the STT unit cost and both the lateness and earliness unit costs. Another contribution of this work is the inclusion in the model of the effect of relaxation when the SlT is larger, i.e., it is known that when the schedule is tight, the actual travel time is shorter than when it is large. Carey calls it the behavioural response. What Carey shows is that the timetable definition should be neither too tight, to avoid delays in departures, nor too large, to avoid behavioural inefficiency.

In all the studies on the definition of travel times, it is not explicit how global cost is defined, i.e., the cost for the passengers and for the company. In Carey's approach, these costs are implicit in the unit costs, but the author does not explore how to estimate them (it is not the goal of the paper). The work by Zhao et al. uses just the passengers' cost, i.e., the expected time the passengers must wait at the bus stop. The operational costs are not considered.

The above mentioned approaches assume that the purpose is to adjust STT and SlT, i.e., the timetable is already defined (even if roughly) and the goal is just to tune it. However, there are several studies on methods for the creation of bus timetables, with different purposes. For instance, Ceder [3] addresses the definition of the frequency related to the problem of the efficient assignment of trips to running boards (i.e., bus duties). Input variables such as the population density of the area served by the line, bus capacity and single mean round-trip time, including slack time, are used [3]. This work was extended in order to address the synchronization of certain arrivals [4]. In [6] the goal is to minimize total schedule delay costs for the users. In [7], the goal is to define the bus departure rate as a function of passengers' arrival rate. In all these works [3,4,6,7], it is assumed that the travel time is deterministic.

The approaches used by Carey and by Zhao et al. benefit from the existence of abundant archived data from AVL systems, in particular the one by Zhao et al.. This work has the appeal of being an analytical approach. However, for the schedulers, rather than a method that solves the (partial) problem in a deterministic way, they need a tool to give them insights into the best solution, at least while there are no answers to questions such as “what are the optimal ratios between STT and lateness unit costs and between STT and earliness unit costs (in Carey's approach)?”, or, “when should the scheduler put on an additional bus, i.e., how does passengers' waiting time compare with the operational cost of an additional bus?”, or even, “what is the impact of reducing the SCT on operational costs?”.

3 A DSS FOR TIMETABLE ADJUSTMENTS

In the previous section we presented the limitations of the existing methods for the creation of bus timetables with regard to the value of travel time. We pointed out that one of the difficulties is the inherent multi-objective nature of the problem of finding the optimal value, namely the minimization of both the expected passengers' waiting time at the bus stop and the operational costs. In this paper we propose a DSS, which allows the person in charge of timetable planning to assess the impact of different scenarios in both objectives. He or she can test different values for the scheduled travel time, slack time and headway, and obtain a set of descriptive statistics that allow this person to evaluate the impact of this scenario using data from a past period similar to the one that the timetable is going to cover.

The reason for using this approach is that the existing ones can give optimized solutions when some of the variables are fixed but do not allow the planner to easily evaluate the sensitivity of the solution to each decision variable. Furthermore, the objective functions used by these approaches do not simultaneously cover the two objectives above mentioned. This DSS can be seen as an integrated environment for analysis and is compatible with the use of optimized solutions like the one described in [10]. In fact, such solutions can always be developed and integrated in this DSS as a default solution for fixed given values.

4 SUPPORTING DECISION IN TIMETABLE ADJUSTMENTS TASKS

The person in charge of planning (the planner) starts by selecting the data to be used for the analysis of travel times. It is expected that the planner will choose past data that might be representative of the period (in the future) that is going to use the new version of the timetable. The planner is able to choose the characteristics of the analysis, such as period of time (days), the time of the day and the line/route. There are three types of analysis that depend on the characteristics of the line and the objectives of the analysis the planner wants to perform: single direction, double direction and circular route analysis.

4.1 Single direction analysis

In a single direction analysis (Fig. 2) the information provided is:

- A time plot of travel times: it provides visual information about travel times. It allows the user to observe the dispersion of travel times during the period of analysis. This can show, for instance, the difference in travel times along the seasons of the year or days of the week. It is also possible to identify outliers and to obtain information on the trips by clicking the mouse.

Figure 2: Single direction analysis.

- A chart of the accumulated relative frequency of travel times: it allows the user to have an idea of the route performance, just by looking at it. The steeper is the slope of the plot, the less is the variability of travel times. We can observe that the duration of the majority of the selected past trips for line 602 is between 50 and 62.5 minutes.

- Information about the sample ('amostra' in Portuguese): it characterizes the sample past data being used in the analysis.

- A table with statistical information on travel times: it provides the user with statistical information for the 25th, 50th and 75th percentile. For these three travel times, say travel time t, one presents the percentage of travel times in the intervals: $]-\infty, t-5\,min[$, $[t-5\,min, t+5\,min[$, $[t+5\,min, t+10$ min $[$, $[t+10\,min, +\infty[$. The value of t that maximizes the percentage of travel times in the interval $[t - 5min, t + 5min[$ is also calculated and presented in the table. This is the initial information provided in the table. The percentiles and the intervals being used were chosen by the planners we have been working with. This information complements the plot of the accumulated relative frequency of travel times. However, the user may add lines to the table by choosing the values of percentile or duration he wants to analyze. The information provided in this table is very useful for the transit planners because it helps them to estimate the effects on delays and early arrivals, when choosing the duration for a trip.

- A table with the STT used in the selected past period: it provides information about the STT in the time interval and dates selected. It is useful to compare the STT being used with the actual past travel times, which allows the detection of STT that are not correctly defined.

4.2 Double direction analysis

In a double direction analysis (Fig. 3), all the information provided for single direction is also shown separately for each of the two directions, as described in the previous section. One also provides an additional tab with statistical information of travel times, possible slack times and the minimum number of vehicles needed, according to the scheduled headway (a value to be introduced by the planner). This kind of information couldn't be shown in a single direction analysis because the definition of slack times and the scheduling of vehicles depend on the entire cycle (go and return).

Fig. 3 shows a screenshot of the application for a double direction analysis containing:

- A plot of the accumulated relative frequency of travel times for both routes (go and return): it is identical to the plot of the accumulated relative frequency of travel times described in Sect. 4.1 but with two data series, one for each route.

- A table with statistical information on cycle times: the information provided in this table is applied to the lines with period trips in a defined period of time, in which vehicle scheduling is done together for both directions. For each travel time presented in the table (go and return), one presents three possible slack times and the percentile for the sum of travel time and slack time. The table also presents the minimum number of vehicles needed to cover the trips with those travel and slack times. The total slack time is calculated based on the headway of the service. Imagine the case of an 11 minutes headway, STT of the first trip equal to 57 minutes and STT of the returning trip equal to 55 minutes. The sum of both durations is 112 minutes. This means that the three lowest values for the duration of the cycle are 121, 132 and 143 minutes, because the duration of the cycle has to be a multiple of the headway (Eq. 2). In the first case we have 121 - 112 = 9 minutes of slack time to distribute by the two trips. The appropriate slack time to add to each trip is calculated based on the aim of minimizing the difference between the percentiles of the total times for each direction (travel time plus slack time). In this example, the solution is to give 6 minutes of slack time for the first trip and 3 minutes of slack time for the returning trip. The percentile values for the sum of STT and slack time are 79.76% and 80.65%, for the go and return trips respectively. This is the solution that minimizes the differences between them. The minimum number of vehicles needed to cover this demand is 11. However, if this solution is applied, we can expect that around 20% of trips, in each direction, won't be finished in a time less or equal to the duration of travel time plus slack time.

This may cause an important number of delays in the departures of sequential trips and poor line performance. So, the transit planner should analyze all the possibilities suggested in the table, trying to maximize the percentage of trips covered by the travel time and slack time, but at the same time, minimizing the costs for the company. The planner is also able to add lines to the table in order to analyze different solutions. He just has to introduce the desired travel times or percentile values, for both directions.

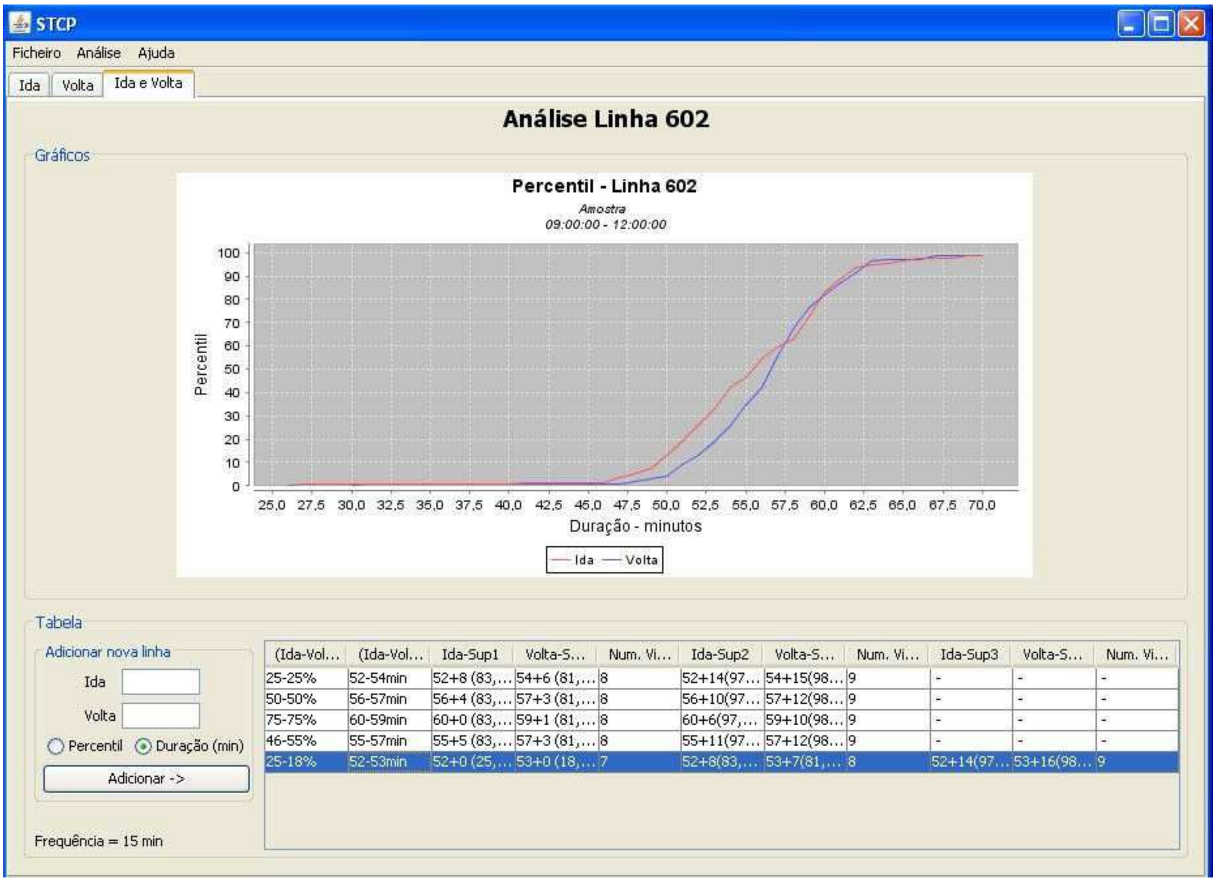

Figure 3: Double direction analysis.

4.3 Circular route analysis

In circular routes there is no slack time added to trips because the bus must always be in service. Besides, the cycle is composed of just one trip. Because of this, the information provided is similar to that in single direction analysis with the addition of the minimum number of vehicles needed for each tested scenario.

5 USING THE DSS

The criteria for defining the timetable depend necessarily on the headway [10]. In this section we describe how to use the DSS for timetable adjustments for typical situations: short scheduled headways, large scheduled headways and circular routes. This section is based in our experience using this tool at the Sociedade de Transportes Colectivos do Porto, SA (STCP), the largest urban public transport operator in greater Oporto, Portugal.

5.1 For short headways

It is known that for short headways (those shorter to 10 minutes) customers arrive roughly uniformly at the bus stops. They do not care about the timetable because they know that they will wait no more than the headway time. Additionally, there is an inherent instability of headways when they are short. This happens because when a bus is delayed, it has, typically, to pick up more customers, increasing the delay after the previous bus even further. At the same time, the following bus tends to go faster because it will stop less and less time since the number of passengers tends to decrease. In other words, for short headways, the actual travel times are very sensitive to the maintenance of the headways. If the actual headways are irregular, the buses tend to bunch [5,10]. Fig. 4 presents a real example of this situation visible for the return trips (the top-down ones). In this situation, i.e., for short headways, the concern of both the planners and the controllers should be to guarantee that the time between two buses on the same route is, as much as possible, equal to the scheduled headway [10].

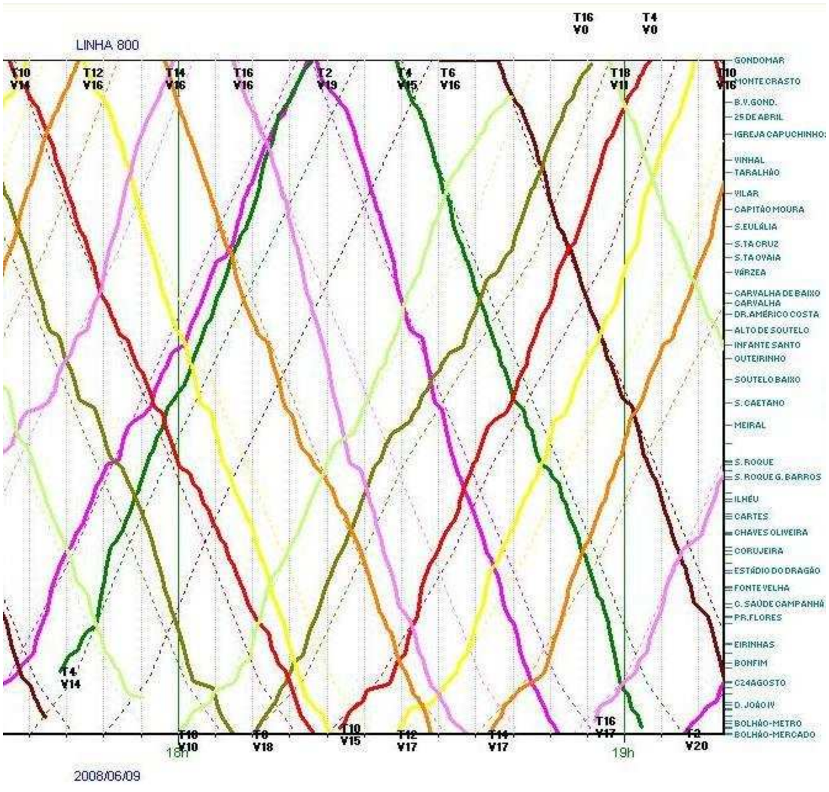

Figure 4: An example of a bunch situation [1].

For the planners the aim is to minimize the variance of the headways. It is advisable to use the mean travel time (or the median, which is more robust to outliers). The sum of STT and the slack time for each direction should be a high percentile value $p.max$. This value is,

[1] The source of this image is a software application owned by STCP.

necessarily, strongly conditioned by the satisfaction of Eq. 2. The possible values are represented in the table with statistical information on cycle times (Fig. 4).

5.2 For large headways

For large headways (those of over 10 minutes), the behaviour of customers is different. They tend to arrive sometime before the scheduled passing time. In this case, the planner should guarantee the adherence of the actual trips to the scheduled ones and the controllers should act accordingly. In this case it is important to be aware that, from the passengers' point of view, an offset between the actual and the scheduled passing times is accepted as in time if in the interval $[-1\ min, 5\ min]$ [8]. Another important issue is that a delay is more acceptable than an advance of the same amount of time.

In this case, a low percentile should be chosen for the STT (represented by $p.min$) in order to reduce the number of trips passing ahead of schedule. The slack time should be the difference between $p.max$ and STT (the $p.min$). Again, this value should respect Eq. 2 and, consequently, the options are limited.

5.3 For circular routes

For circular routes, whatever the scheduled headway is, the STT should comprise both the expected travel and the slack times. The problem is that on circular routes it is not possible to have slack times because the buses are always in service. The time to accommodate delays must be incorporated in the STT. Consequently, a high percentile value $p.max$ should be chosen, in the same way as previously described. In this case, controllers tend to have more intervention in order to guarantee that the actual travel time respects the STT.

6 DISCUSSION AND FUTURE WORK

An important advantage of this DSS is that it allows planners to evaluate different scenarios. This is achieved by providing them information about the two business objectives, namely, the minimization of both the expected passengers' waiting time at the bus stop and the operational costs. With this purpose different indicators are used:

- Number of vehicles needed: an important indicator of the operational costs;
- Estimated percentage of trips passing ahead of schedule: an indicator of the level of passengers' satisfaction;
- Estimated percentage of trips starting delayed: another indicator of the level of passengers' satisfaction.

However, this set of indicators is neither complete for the case of large headways nor adapted for short headways or circular routes. A study on the best set of indicators for each case is needed. In [8, 9] many indicators are suggested.

It is expected that the values of $p.min$ and $p.max$ discussed along Sect. 5 can be obtained empirically by the planners. Their values can be different for different routes but it is expected that, at least for routes with identical characteristics, they will not be too different. Anyway, this DSS can also integrate analytical approaches, like the one suggested in [10]. This is a natural step forward for this research.

REFERENCES

[1] Bowman, L. A. and M. A. Turnquist (1981). "Service frequency, schedule reliability and passenger wait times at transit stops." Transportation Research Part A 15: 465-471.

[2] Carey, M. (1998). "Optimizing scheduled times, allowing for behavioural response." Transportation Research Part B 32(5): 329-342.

[3] Ceder, A. (1986). "Methods for creating bus timetables." Transportation Research Part A 21: 59-83.

[4] Ceder, A., B. Golany, et al. (2001). "Creating bus timetables with maximal synchronization." Transportation Research Part A 35: 913-928.

[5] Newell, G. F. (1974). "Control of pairing vehicles on a public transportation route, two vehicles, one control point." Transportation Science 8(3): 248-264.

[6] Palma, A. and R. Lindsey (2001). "Optimal timetables for public transportation." Transportation Research Part B 35: 789-813.

[7] Salzborn, F. J. M. (1972). "Optimum bus scheduling." Transportation Science 6(2): 137-147.

[8] Strathman, J. G., K. J. Dueker, et al. (1998). “Automated bus dispatching, operations control and service reliability: analysis of Tri-Met baseline service date.” University of Washington - U.S.A.

[9] Wang, P., J. J. Lin, et al. (2006). “Data envelopment analysis of bus service reliability using automatic vehicle location data.” Annual Meeting of Transportation Research Board.

[10] Zhao, J., M. Dessouky, et al. (2006). "Optimal slack time for schedule-based transit operations." Transportation Science 40 (4): 529-539.