AN ALGORITHM FOR THE EXPRESS SERVICE DESIGN PROBLEM ON A CORRIDOR.

HomeroLarrain. Pontificia Universidad Católica de Chile. hlarraii@uc.cl

Juan Carlos Muñoz. Pontificia Universidad Católica de Chile. jcm@ing.puc.cl

ABSTRACT

In public transit systems with high demand levels, the use of express services is a promising alternative given the benefits they offer to both users and operators.For users, express buses offer shorter travel times due to fewer stops and higher between-stop speeds. For system operators, express buses allow demand to be met with fewer vehicles because of shorter bus cycles.

Leiva et al. (2010) introduced a methodology for the design of express services on a corridor that works for a given set of services. In order to develop an algorithm that doesn't rely on a predefined set of services, we propose a new way to represent services when working on a bidirectional corridor, separating them by direction, which expands the model's domainusing lesser input. We also improved the capacity adjusting heuristic to make it considerably faster.

Working with this new version of the frequency setting model, we developed two heuristics to generate express services to feed the model. We introduce a specific format of express services we called superexpress service, and we propose heuristic methods to design these services for congested (i.e. with active bus capacity restrictions) and uncongested scenarios. The second heuristic is of particular interest because it is an example of a particular instance of the express service design problem that can be solved to optimality (for a specific set of services) analytically considering both active capacity constraints and rational user behavior, without involving the use of computational tools or iterative methods.

The two service generating heuristics and the modified frequency setting model were applied on a forty stop bidirectional corridor (20 stop per direction), where we were able to obtain social cost reductions around 10% for both congested and uncongested scenarios.

Keywords: BRT, bus rapid transit, public transport, express services, network design, limited-stop services, route design.

INTRODUCTION

During the last decades, bus operated public transport systems have flourished around the whole world, and they keep becoming more and more popular. This can be explained by two factors: on one hand, experience has shown that public transport and buses in particular are a vital element in the sustainable development of an urban area.On the other hand, bus systems can be very cost effective and beneficial for both users and operators. Since the early 2000's the concept of a bus system with high quality standards has been referred as Bus Rapid Transit, BRT. It's estimated that at this moment there are 147 cities around the world that have implemented BRT on their streets. Around 75% of these systems were inaugurated after the year 2000, and about half of the total systems in the world are less than 6 years old (Global BRT Data, *www.brtdata.org*).

There are several elements that are associated to BRT. Amongst many others, we can mention the use of segregated bus lanes that reduce travel times, centralized fare collection systems, and modern bus stops and stations with unobstructed boarding (Levinson et al, 2002). Another defining element of BRT is the inclusion of express bus services that serve only a subset of stops along certain routes, which especially designed to deal with higher demands and to improve the level of service.

In public transit systems with high demand levels, the use of express services is a promising alternative given the benefits they offer to both users and operators. For users, express buses offer improved service levels in the form of shorter travel times due to fewer stops and higher between-stop speeds. For system operators, express buses allow demand to be met with fewer vehicles because of shorter bus cycles. In practice, express services in systems such as Transmilenio (Bogota, Colombia), Transantiago (Santiago, Chile), and Metro Rapid (Los Angeles, California) have proved to be highly appealing.

The express service design problem involves answering two main questions: which services to provide (i.e. which stops should a service skip) and on what frequency or schedule should each service operate. Due to the complex nature of the design problem, it seems natural to approach these two questions individually, at least as a first stage. In this work we show some results that can be useful to answer the first question. To be able to design good service configurations, it is necessary to have a way to optimize their frequencies in order to compare the quality of different solutions. This second problem has been studied by Leiva et al. (2010), where they provide a continuous mathematical programming model for solving the frequency setting problem for a given set of express services, assuming the operation is frequency based and not schedule based, which holds true in systems where the frequencies are high.

In this work we present a method for the design of express services. We take a specific format of service we called superexpress service, that serves a set of consecutive stops on the beginning of its route, then it skips a set of consecutive stops, to finally serve the remaining stops of its route. This kind of service can be very appealing in terms of time savings to users, it can be very useful to alleviate the burden on an all-stop service when working with capacity, and it also presents some properties that make it possible analytically optimize their configuration.

Ceder and Wilson (1986) give a framework for the general public transport system design problem, dividing it in five levels, where the first one is network design. The express service design problem is very close to this problem: both consist, in general terms, in defining routes that connect nodes of demand. There are a great number of examples on the literature that address thenetwork design problem. Some examples are the works of LeBlanc (1988), Baaj and Mahmassani (1992, 1995), Fan and Machemehl (2006), Mauttone and Urquhart (2009), and reviews from Ceder (2003), Dasaulniers and Hickman (2007), and Guihaire and Hao (2008). In general, these works do not take capacity into account for the analysis. When capacity is not reached by a system, since every user can choose freely which services to use, minimizing social costs will result in a solution that satisfies user equilibrium. However, when capacity constraints are included in the models and some of the services reach capacity, social optimum may differ to user optimum, and it becomes necessary to assure that the model will yield solutions consistent to rational user behaviour. In Fernandez et al. (2003, 2008) a bi-level formulation is introduced which allows to design bus services for a network dealing with capacity and user behaviour at the same time.

The main difference between the network design problem and the express service design problem is that the first one assumes that services stop in every bus stop along their way (or they just don't deal with bus stops at all). Both problems could be seen as different levels of detail for a broader design problem: on a higher level we have the route definition problem, defining where the services operate, and on a lower level we would have the service optimization problem, where bus stops for every service are designed in more detail.

One of the first works that propose express services as an alternative to improve the performance of public transport systems is Furth and Day (1985), where three different established planning strategies for high-demand corridors are discussed:

1. Short turn: Some buses serving a route make shorter cycles in order to concentrate on areas of greater demand. This is useful when it is desired to bolster capacity along a given stretch of the route.

2. Deadheading: Empty vehicles return to the route starting point in the low-demand direction in order to begin another run as quickly as possible in the high-demand direction, thus increasing the latter's frequencies. This is advantageous when demand along the corridor is imbalanced.

3. Express services: Services that visit only a subset of the stops on a route.

Although various works in the literature focus on the first two approaches, such as Furth (1987) on short turning and Ceder and Stern (1981) on deadheading, there appears to be no published research that explores optimization techniques for designing express services on high demand corridors and evaluating their benefits. Turnquist (1979) studies an express service optimization problem and solves it using dynamic programming. However, he assumes for simplicity that all trips share the same origination or destination. This assumption prevents dealing with transfers or user behavior, as every user have just one way to get to their destination. Unfortunately, such an assumption is too strong to be useful in the context of urban corridors like the one in the case in study.

Leiva et al. (2010) propose a design model for express services in public transit corridors with capacity constraints that minimizes social costs. It assumes first of all that the network topology representing the corridor is known and fixed, implying that we know a set of stops and have some notion of the between-stop distances so as to estimate the corresponding travel times. A second assumption of the model is that there exists an exogenous trip matrix for the corridor that is also known. For simplicity, this value is assumed to be fixed, and its effect on modal share of express services will be neglected. Finally, it is assumed that the set of possible corridor servicesfrom which the services to be used will be selected are known a priori. These services are defined by the set of stops they serve. This means that the planner has a basic intuition of what kind of lines would be of interest to take into account, and this is why in the context of this work we will refer to this model as a frequency optimization model instead of as a design model.Larrain et al. (2010) used this model to make some experiments to study which demand patterns on a corridor where more likely to yield more savings when using express services, concluding that corridors with longer trips and decreasing (or increasing) load profiles were the most attractive.

The remainder of this paper is organized as follows: On the next chapter, we introduce Leiva's model and propose some improvements and simplifications that facilitate the service design problem. These improvements consist in a new way of modeling services separating the directions of a corridor, and a modification on the capacity adjustment heuristic that makes it noticeably faster. In chapter three we propose two different algorithms that generate superexpress services. The first one optimizes a service assuming bus capacity is not reached and can be used as a criterion for generating services to feed the frequency optimization model. The second one does a similar analysis assuming capacity constraint is active. In chapter four an experiment is presented to test how the heuristics perform on a simple scenario. Finally, the conclusions for this work are summarized in the final chapter of this document.

FREQUENCY OPTIMIZATION MODEL

In this section we present the model we used to deduce the optimal frequencies for a set of given services operating on a corridor. As mentioned before, the model we used for this work is based on the model proposed by Leivaet. al. (2010), which can be stated as follows:

$$Min\left[\sum_{l\in\mathcal{L}} c_l f_l + \theta_{wt}\sum_{w\in\mathcal{W}}\sum_{s\in\mathcal{S}} V_s^w \frac{\lambda}{\sum_{l\in\mathcal{L}} f_l^s} + \theta_{tt}\sum_{w\in\mathcal{W}}\sum_{s\in\mathcal{S}} V_s^w \frac{\sum_{l\in\mathcal{L}} t_l^s f_l^s}{\sum_{l\in\mathcal{L}} f_l^s} + \theta_{tr}\left\{\sum_{w\in\mathcal{W}}\sum_{s\in\mathcal{S}} V_s^w - \sum_{w\in\mathcal{W}} T_w\right\}\right] \quad (1)$$

Subject to:

$$f_l^s \leq f_l, \qquad \forall l \in \mathcal{L}, \forall s \in \mathcal{S} \tag{2}$$

$$f_l^s \geq 0, \qquad \forall l \in \mathcal{L}, \forall s \in \mathcal{S} \tag{3}$$

$$\sum_{s \in \mathcal{S}_i^+} V_s^w - \sum_{s \in \mathcal{S}_i^-} V_s^w = \begin{cases} T_w, & \forall i \in \mathcal{P}/i = O_w \\ -T_w, & \forall i \in \mathcal{P}/i = D_w \\ 0, & e.o.c. \end{cases} \tag{4}$$

This problem optimizes three sets of variables: first we have f_l, the frequency for service $l \in \mathcal{L}$, where $\mathcal{L}$ is a set of services for the corridor that the modeler has to define. Second, there is variable f_l^s, with $s \in \mathcal{S}$, where $\mathcal{S}$ is the set of route sections that can be used as stages for a trip that transfers, and can be assumed to be equivalent to $\mathcal{W}$, the set of OD pairs. This variable makes it possible to model behavior, allowing users on a specific section s to choose which services to user in order to minimize their expected travel time. Finally, variables V_s^w represent the total flow from users travelling on pair $w \in \mathcal{W}$ that will use section s as part of their trip.

The objective function of this optimization problem corresponds to the total social costs, consisting of:

- Operator costs, estimated as the product between frequencies and the operational cost c_l of each service. Operational cost can be obtained as the sum of the operational costs that are proportional to the service cycle length (gas, tires, etc.) and costs that depend on cycle duration or, indirectly, fleet size (wages, depreciation of the vehicles, etc.). Accordingly, costs for each service can be estimated as:

  $$c_l = L_l \cdot c_L + T_l \cdot c_T \tag{5}$$

  In expression (5) L_l and T_l stand for cycle length and time, while c_L and c_T represent the operational costs per distance and time, respectively.
- Total waiting time cost (second term on the objective function) is computed assuming that the expected waiting time will be proportional by a constant λ to the average headway (Mohring, 1972, Jansson, 1980). In this expression parameter θ_{wt} stands for the users' value of waiting time. Note that for the estimation of the average interval, only the frequencies of attractive services are taken into account.
- Total travel time cost (third term) is obtained as a weighted average, considering attractive services and the travel time for each one of them,t_l^s. Parameter θ_{tt} stands for the value of travel time.
- Transfer times. Since this term is not relevant for this work, it will not be explained here. The reader can refer to Leiva et al. (2010) or Larrain et al. (2010) for more detail on this model.

Restrictions (2) and (3) for the model ensure that the attractive lines' frequencies take valid values. Restriction (5) ensures continuity of flow, and is only important when working with transfers. Since we are assuming that there are no transfers being made on the network, we can restate the objective function of the model as:

$$Min\left[\sum_{l\in\mathcal{L}} c_l f_l + \sum_{w\in\mathcal{W}} T_w \left\{\theta_{tt}\frac{\lambda}{\sum_{l\in\mathcal{L}} f_l^w} + \theta_{wt}\frac{\sum_{l\in\mathcal{L}} t_l^w f_l^w}{\sum_{l\in\mathcal{L}} f_l^w}\right\}\right] \tag{6}$$

Furthermore, an absenceof transfers also means that constrain (4) can be dropped from the model, and that there is no need to work with route sections, so restrictions (2) and (3) can be written as follows:

$$f_l^w \leq f_l, \qquad \forall l \in \mathcal{L}, \forall w \in \mathcal{W} \tag{7}$$

$$f_l^w \geq 0, \qquad \forall l \in \mathcal{L}, \forall w \in \mathcal{W} \tag{8}$$

If our corridor is bidirectional, we can treat service design for each direction independently, feeding the model with services for each direction instead of cyclic services as the ones used in the original model. This gives the model additional flexibility to explore more solutions using fewer variables, and it is also very convenient analytically, as we will see in the next chapter. If every service to be taken into account spans the whole length of corridor for one of its directions, we must add the following restriction to assure that the number of buses required in each direction is the same, and that buses aren't accumulating in one end of the corridor.

$$\sum_{l\in\mathcal{L}_1} f_l = \sum_{l\in\mathcal{L}_2} f_l \tag{9}$$

In this expression sets $\mathcal{L}_1$ and $\mathcal{L}_2$ represent the services belonging to each direction of the corridor. This equation can be understood as a continuityrestriction for the frequencies.

So, in summary, the model that we used to solve the frequency optimization problem was to minimize social costs as show in expression (6), subject to restrictions (7), (8) and (9). This model can be solved using any commercial solver for non-linear optimization problems available in the market, such as MINOS or LANCELOT just to mention a few.

The model we have introduced does not deal with capacity constraints on the buses. In other words, the solution this model gives can be unrealistic in terms of bus occupation when demand is high. Unfortunately, adding a capacity constraint to the model would not solve this problem satisfactorily, because this imposition would lead to solutions where the users are forced to behave irrationally. This problem is further explained in Leiva et al (2008).In the same work the authors propose a heuristic that deals with this problem, looking for a new solution that is feasible in terms of bus occupation, taking the solution from the model as a starting point. This heuristic can be used as well with our version of the model, and goes as follows:

1. Add lower bound restrictions for the frequencies to the model and set this lower bound for each frequency initially as $f_l^{lb} = 0$.

2. Solve the optimization problem. Check if the solution is feasible in terms of capacity using equation (10):

$$cap \geq \max_{l \in \mathcal{L}, a \in \mathcal{A}} \frac{1}{f_l} \cdot \sum_{w \in \mathcal{W}} T_w \varsigma_a^w \frac{f_l^w}{\sum_{i \in \mathcal{L}} f_i^w}, \tag{10}$$

   In the last expression the binary parameter ς_a^w takes the value of 1 when arc a is included somewhere in between the origin and destination of pair w. If this inequality holds, then the current solution is a feasible one and the algorithm comes to an end.

3. For the current solution, identify the service$\hat{l}$where the capacity deficit is higher, i.e.:

$$\hat{l} = \underset{l \in \mathcal{L}}{\operatorname{argmax}} \left[\max_{a \in \mathcal{A}} \left(\sum_{w \in \mathcal{W}} T_w \varsigma_a^w \frac{f_l^w}{\sum_{i \in \mathcal{L}} f_i^w} - cap \cdot f_l \right) \right] \tag{11}$$

4. Increase the lower bound for the frequency of line$l$ in a predefined step Δ, i.e., make $f_{\hat{l}}^{lb} \rightarrow f_{\hat{l}}^{lb} + \Delta$.

5. Go back to step 2.

This heuristic has two main drawbacks: first, while it manages to get to a feasible solution, it gives no certainty that this solution is optimal for the provided set of services. Second, this algorithm can get very slow sometimes because it has to solve an instance of the optimization problem in every iteration it makes.

The first issue is not easy to tackle, because it would imply reformulating the model in a way that somehow deals simultaneously with capacity and user behavior at the same time. An example of a particular case where the problem can be solved to optimality for a given se of services would be the second heuristic we present on chapter three, but unfortunately the set of services it considers is so reduced that it is not difficult to find better solutions working with larger sets of services (even when these solutions are not necessarily optimal).However, even though we have found instances where this algorithm leads to suboptimal solutions, it usually finds an improvement for the system when compared to the basic solution where no express services are provided.

Regarding the second issue, there is a very simple improvement that can be made in order to speed up the algorithm. It consists in replacing the fixed step Δfrom step number 4 for a variable one, computed as the required frequency needed to meet capacity. In other words, the lower bound actualization in step number 4 from the heuristic can be replaced for $f_{\hat{l}}^{lb} \rightarrow P_{\hat{l}}^{max} / \, cap$, where $P_{\hat{l}}^{max}$ is the load of the critical arc in line $\hat{l}$, and can be computed using the right hand side of expression (10).

EXPRESS SERVICE DESIGNHEURISTICS

The model presented so far obtains the optimal frequencies (or near-optimal, when capacity restrictions are active) for a set of given services for a corridor. However, the quality of this solution depends largely on the quality of the services fed to the model. In this chapter we willintroduce a methodology for designing a particular format of service we are calling superexpress services thatcan be very beneficial for the system in some occasions.

For this analysis, we will focus on one direction of the corridor at the time. Later on we will discuss a way to extend the results to both directions of the corridor. We'll define superexpress service as a service that runs along one direction of a corridor, skipping a subset of successive stops in the middle section of it. This type of service can greatly reduce travel times for certain passengers and it is also cheaper to operate, but these benefits come at the cost of a greater waiting time for the passengers of the skipped middle section. To simplify the designof this type of service, we assumed that a single one of them was operating simultaneously with an all-stop service (i.e. a service that attends every stop on its route), and separated the analysis in two cases: no congestion and under congestion (i.e. when capacity restrictions become active).

Before examining each case it is necessary to make some definitions. We will define $\mathcal{L}_E$ as the set of superexpress services that exist for a direction of a corridor under study. We will denote the first and last stops that are skipped in the middle section of service $e \in \mathcal{L}_E$ as $i_e$and j_e. Every trip on the corridor will fall under one of three categories: trips that will only use the superexpress service, trips that only use the all stop service, and trips that will use the first service to come. Let's call $\mathcal{W}_E^e$, $\mathcal{W}_A^e$ and $\mathcal{W}_{AE}^e$ the subsets of $\mathcal{W}$ that contain the trips for each category, respectively. If the time savings due to the skipped stops are attractive enough, we can assure that every trip that goes through the skipped section of the superexpress will prefer it, i.e., $w \in \mathcal{W}_E^e \Leftrightarrow (O(w) < i_e \wedge D(w) > j_e)$ (in this expression,$O(w)$ and $D(w)$ represent the origin and destination nodes of pair w). It is possible to determine under what conditions this situation holds, as we will show later in this paper. The set $\mathcal{W}_{AE}^e$ can be defined as $w \in \mathcal{W}_{AE}^e \Leftrightarrow (D(w) < i_e \vee O(w) > j_e)$, or, in other words, a trip will be indifferent between the two options if it's contained in either of the extreme sections of the superexpress, where it operates as an all-stop service. The last set, $\mathcal{W}_A^e$, can be simply defined as the complement of the other two ones, i.e. $\mathcal{W}_A^e = \mathcal{W} - \mathcal{W}_E^e - \mathcal{W}_{AE}^e$. Using these sets, that can be easily computed for every $e \in \mathcal{L}_E$, we can obtain the total number of trips falling under each category:

$$T_E^e = \sum_{w \in \mathcal{W}_E^e} T_w \tag{12}$$

$$T_{AE}^e = \sum_{w \in \mathcal{W}_{AE}^e} T_w \tag{13}$$

$$T_A^e = \sum_{w \in \mathcal{W}_A^e} T_w \tag{14}$$

These values are needed to estimate the effect on the social costs of the system of any superexpress service, but in order to correctly estimate them it is vital to know the congestion level of the corridor. We will focus on each case on the following two sections of this document.

Superexpress service design for an uncongested corridor

The objective function to minimize for this case is:

$$SC = (f_a c_a + f_e c_e) + \left(\frac{\lambda \theta_{wt} T_A^e}{f_a} + \frac{\lambda \theta_{wt} T_E^e}{f_e} + \frac{\lambda \theta_{wt} T_{AE}^e}{f_a + f_e} \right) + \left(\theta_{tt} \sum_{w \in \mathcal{W}} \sum_{l \in \mathcal{L}} t_l^w T_w - \theta_{tt} T_E^e N^e t_{stop} \right) \tag{15}$$

The first term in bracketson (15) corresponds to the operator costs and can be obtained directly from equation (6). The second part represents users waiting costs and is also very straightforward: every type of user will wait accordingly to the services they are willing to use. The last term of the objective function represents the travel time costs, and is computed as the total travel time when every user takes the all-stop service minusthe savings that superexpress users will face. In this expression the parameters N^e and t_{stop} are the number of stops that the superexpress service skips and the travel time each skip saves.

Given that the benefits from using superexpress services come primarily from travel time savings and operator cost savings, we will simplify the objective function assuming that we can neglect the effect on waiting times the superexpress service has on trips belonging to $\mathcal{W}_{AE}^e$. In simpler words, we will assume that the users for the uncongested case will either wait for the all-stop or the express service, which is an overestimation of waiting times, but still can give a good approximation of the social costs in order to generate superexpress services to feed the model. We will also exclude the constant total travel time term from the social costs, since they don't affect the optimal solution.Thus, the modified objective function for the problem now will be:

$$\widehat{SC} = f_a c_a + f_e c_e + \frac{\lambda \theta_{wt} (T_A^e + T_{AE}^e)}{f_a} + \frac{\lambda \theta_{wt} T_E^e}{f_e} - \theta_{tt} T_E^e N^e t_{stop} \tag{16}$$

Now we can calculate the derivatives of the objective function for $f_a$and f_e:

$$\frac{\delta \widehat{SC}}{\delta f_a} = 0 \rightarrow f_a^* = \sqrt{\frac{\lambda \theta_{wt} (T_A^e + T_{AE}^e)}{c_a}} \tag{17}$$

$$\frac{\delta \widehat{SC}}{\delta f_e} = 0 \rightarrow f_e^* = \sqrt{\frac{\lambda \theta_{wt} T_E^e}{c_e}} \tag{18}$$

Replacing these values in (16) we can obtain the optimal modified social costs for any given superexpress service e:

$$\widehat{SC}_e^* = 2\sqrt{\lambda \theta_{wt}(T_A^e + T_{AE}^e)c_a} + 2\sqrt{\lambda \theta_{wt} T_E^e c_e} - \theta_{tt} T_E^e N t_{stop} \tag{19}$$

We can compare these costs with the ones we would obtain if the system was served by just an all-stop service operating at optimal frequency. This will give us a way to tell if any given superexpress servicecan be beneficial. It is easy to show that the optimal social costs for this case, leaving out the travel time costs because they would cancel out the term we left out in the objective function, are:

$$\widehat{SC}_a^* = 2\sqrt{\lambda \theta_{wt} c_a \sum_{w \in \mathcal{W}} T_w} \tag{20}$$

These values allow us to formulate a simple heuristic for generating services to feed the frequency optimization problem: we can calculate $\widehat{SC}_e^*$ for every $e \in \mathcal{L}_E$ and compare it to $\widehat{SC}_a^*$: if the costs using the superexpress service represent an improvement versus the only all-stop case, then this service should be included in the frequency optimization problem. For a bidirectional corridor it is possible to apply the heuristic on both directions of the corridor to independently generate some superexpress services to feed the model.However, because of the approximation made on the social costs, and because of the frequency continuity constraints of the model, frequency optimization should be always left to the model in this case, and the values of f_a^* and f_e^* should be only used for the estimation of the social costs.

It was noted before that in order to the behavior assumption to hold, we have to check that the time savings of the superexpress service are high enough to make this service the only attractive option for long trips. Nevertheless, since we are using this method to generate services to feed the frequency optimization model, it's not that important to be absolutely sure that the computed social costs are exact. This is because the frequency optimization model will be able to leave out by itself the services that are not beneficial for the system. Still, when working on a congested corridor, it is indispensable to check this condition, as it will be shown on the next section.

Superexpress service design for a congested corridor

When designing a superexpress service for a congested corridor (or, more specifically, for the direction of the corridor that contains the critical section), it is necessary to reconsider the effect on social costs of the new service. Since in this case the system frequencies are set to meet capacity requirements on the critical arc $a^*$and these frequencies are higher than the ones that would minimize social costs, we can expect the capacity restriction to remain active even when the new service is taken into account. If we focus on the critical arc, we can be sure that the total frequency of buses running through this arc will have to be at least of $f_0 = P_{a^*}/cap$, where P_{a^*} is the load on the critical arc of the corridor. Given that we already know that the frequencies are higher than the optimal ones if there was no congestion, it would be desirable that the frequencies of the all-stop and superexpress services add up to f_0. This can only be achieved when every bus from every service is fully loaded on the critical arc.

$$f_a + f_e = f_0 \tag{21}$$

The load on the critical arc P_{a^*} can be written as the sum of the three terms, PM_E^e, PM_A^e, andPM_{AE}^e, which represent the fraction of the critical load for each type of trip as defined earlier, and can be computed as follows:

$$PM_E^e = \sum_{w \in \mathcal{W}_E^e} \varsigma_{a^*}^w T_w \tag{22}$$

$$PM_{AE}^e = \sum_{w \in \mathcal{W}_{AE}^e} \varsigma_{a^*}^w T_w \tag{23}$$

$$PM_A^e = \sum_{w \in \mathcal{W}_A^e} \varsigma_{a^*}^w T_w \tag{24}$$

The average loads on a bus for each type of service can be estimated as:

$$LM_a^e = \frac{PM_A^e}{f_a} + \frac{PM_{AE}^e}{f_0} \tag{25}$$

$$LM_e^e = \frac{PM_E^e}{f_e} + \frac{PM_{AE}^e}{f_0} \tag{26}$$

As we already mentioned, in the optimal solution we can expect both services to be operating at full capacity. Imposing that$LM_a^e = cap$ and $LM_e^e = cap$ we can obtain the following expressions for the optimal frequencies:

$$f_a^* = \frac{f_0 PM_A^e}{PM_A^e + PM_E^e} \tag{27}$$

$$f_e^* = \frac{f_0 PM_E^e}{PM_A^e + PM_E^e} \tag{28}$$

For the objective function we will use, for simplicity, the variation of social costs using the all-stop solution as reference instead of the total social costs. For the operator, the savings will come from the reduction in the cycle time of the superexpress service. If c_T is the cost the operator faces for unit of time of providing a service the savings the operator will get from the superexpress will be:

$$\Delta OC = f_e c_T N t_{stop} \tag{29}$$

Waiting times are going to change for users belonging to $\mathcal{W}_A^e$and $\mathcal{W}_E^e$. This variation can be calculated as follows:

$$\Delta TWC = \theta_{wt} T_A^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_a}\right) + \theta_{wt} T_E^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_e}\right) \tag{30}$$

The effect on travel times is the same as in the uncongested case. Therefore, the total savings due to the new services will be:

$$S_e = f_e c_T N t_{stop} + \theta_{tt} T_E^e N t_{stop} + \theta_{wt} T_A^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_a}\right) + \theta_{wt} T_E^e \left(\frac{\lambda}{f_0} - \frac{\lambda}{f_e}\right) \tag{31}$$

We can now substitute the values for f_a^* and f_e^* in (31) to find which superexpress service leads to greater savings. However, there is one additional check we should make before choosing the service, and it is to make sure that the superexpress service will indeed be fast enough to be attractive by itself. This condition will be met when the expected travel time considering waiting and in-vehicle travel times for using just the express service is lower than the one obtained when taking the first service to arrive:

$$\frac{\theta_{tt} f_a t_a^s + \theta_{tt} f_e t_e^s + \theta_{wt}\lambda}{f_a + f_e} \geq \theta_{tt} t_e^s + \frac{\theta_{wt}\lambda}{f_e} \tag{32}$$

Solving equation (32) for f_e we obtain the condition for f_e^*:

$$f_e^* \geq \frac{\theta_{wt}\lambda}{\theta_{tt} N t_{stop}} \tag{33}$$

Therefore, the methodology for choosing the optimal superexpress service for a congested scenario would be to calculate the social savings for every $e \in \mathcal{E}$ that satisfies equation (33), and take the one with greater savings. Then, the optimal solution will be to operate with an all-stop and a superexpress service using the computed frequencies on the congested direction under consideration. For the opposite direction of the corridor we can use the heuristic for uncongested scenarios to generate services to feed the model. So, to obtain a solution for the whole corridor with its two directions, we have feed the model with just the all-stop and the best superexpress for the congested direction and with the all-stop and all the services found by the heuristic for the opposite direction, and solve the optimization problem fixing the frequencies of the congested direction at the computed optimal values.

This version of the algorithm has the advantage that it can find a solution analytically, and more important, it can assure that, for at least the congested direction and for the set of services under consideration, the solution is optimal. However, since the number of services is very reduced, this optimal solution could be beaten by suboptimal solutions for larger service sets. The issue here is that we have to be careful when adding new services to the congested direction because it might invalidate the analytical results by changing the configuration of sets $\mathcal{W}_E^e$, $\mathcal{W}_A^e$ and $\mathcal{W}_{AE}^e$. However, it is easy to see that if a service does not compete for demand with the express service on its middle section (like, for example, any short turning service that doesn't contain the entire middle section in its route), then the load on this service on the critical arc will remain unchanged if the critical arc is contained by the middle section, as it most likely will be. This means that, if we choose a superexpress service that contains the critical arc in its middle section, we can add some services to let the optimization model look for a better solution, using the capacity algorithm presented on chapter 2. In this case the only frequency that should be fixed from the beginning should be the one of the superexpress service, because the new services might steal some trips from the all-stop service on the critical section.

TESTING THE HEURISTICS

The two heuristicswere tested on a forty stop bidirectional corridor (20 stops on each direction). On this corridor we considered an OD demand matrix we adapted from the real OD matrix from AvenidaPajaritos corridor in Santiago. The OD matrix load profile and the demand on each bus stop ($O_i$and D_i) for each direction are shown in Figures 1 and 2.

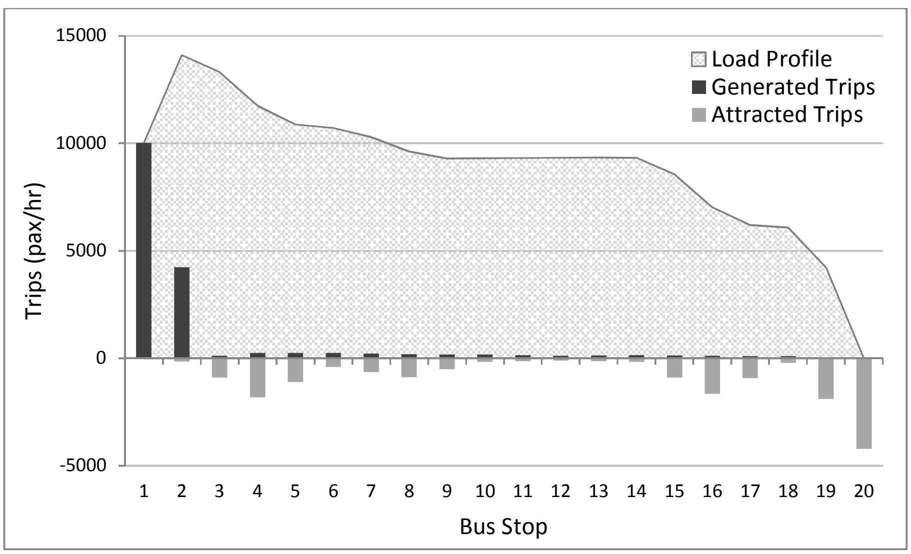

Figure 1 – Demand of the test corridor on the N-S direction

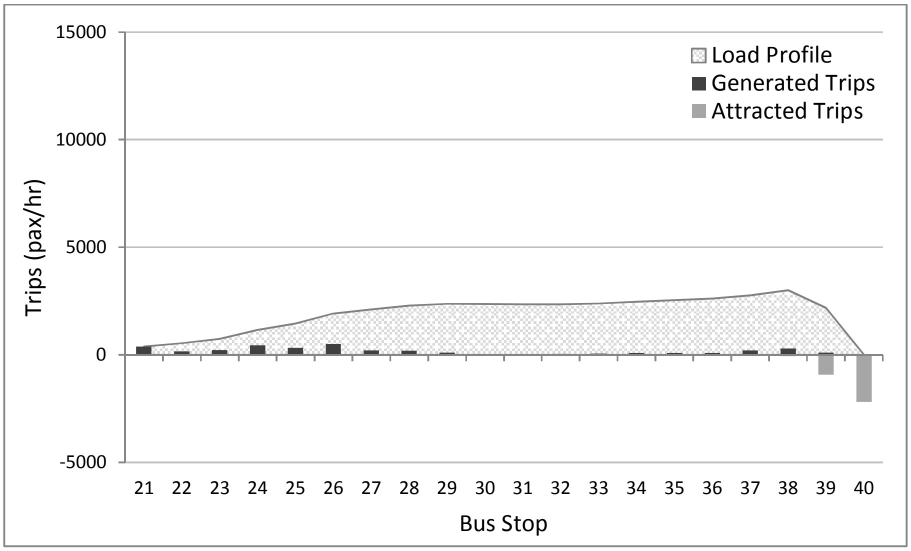

Figure 2 – Demand of the test corridor on the S-N direction

From these figures it can be observed that this is a corridor with a decreasing load profile on its most congested direction, like the ones one would expect to find in corridors that begin on the central business district of a city during the afternoon peak hour, for example. Some other relevant characteristicsof the corridor and its demand matrix are summarized in Table I.

Table I – Corridor and demand attributes

Parameter	Value
Number of stops	40
Number of OD pairs	380
Average distance between bus stops	400 m
Bus running speed	25 Km/hr
Total trips N-S	16,903 pax/hr
Total trips S-N	3,550 pax/hr
Average trip lenght N-S	4.23 Km
Average trip lenght S-N	4.28 Km

For this experiment we assumed that bus stops were all evenly spaced. The value bus running speed stands for the speed buses reach between bus stops and it is used to calculate fixed travel times. It is worth noting, anyway, that these two values have no effect on the optimal services and their frequencies, since the fixed travel times add up to a constant on the objective function, and it is the variable part of travel times that depends on dwell times the one that can be saved.

The average trip lengths for this corridor are considerably high: the average trip goes through more than ten stops, meaning that there nearly ten minutes that could be saved on each trip on average, which, according to Larrain et al. (2010), is a sign that we can expect this corridor to greatly benefit from the use of express services.

The values for the parameters from the model used for this experiment are shown in Table II:

Table II – Parameters for the experiment

Parameter	**Value**
$\boldsymbol{\lambda}$	1
Value of in-vehicle travel time, $\boldsymbol{\theta_{tt}}$	1.91 USD/hr
Value of waiting time, $\boldsymbol{\theta_{wt}}$	5.73 USD/hr
Stop time, $\boldsymbol{t_{stop}}$	1 min
Operator costs per distance, $\boldsymbol{c_L}$	750 $/bus-Km
Operator costs per time, $\boldsymbol{c_T}$	7500 $/bus-hr
Bus capacity, $\boldsymbol{cap}$	160 pax/bus

The experiment consisted in solving the design problem for this corridor using the algorithms introduced in this paper for service design and for frequency optimization under congestion. We defined six stages, consisting on different congestion levels and solution algorithms. These stages are:

1. All-stop, no capacity constraint. This stage was solved using the frequency optimization model.

2. All-stop and superexpress services, no capacity constraint. We used the superexpress service design algorithm for uncongested corridors to identify the ten most promising services (the ones where the difference between (19) and (20) is higher), and then we fed them (jointly with all-stop services) to the frequency optimization model.

3. All-stop, with active capacity constraint. The optimal frequency for this case is determined by the load on the critical arc, i.e. $f_a^* = P_{a^*}/cap$.

4. All-stop and superexpress services, with active capacity constraint. In this stage the model was fed with the same services generated in stage 2, and capacity restrictions were dealt with using Leiva's algorithm.

5. Same as 4, but using the proposed correction for the step in Leiva's algorithm.

6. All-stop and superexpress services, with active capacity constraint, using the superexpress service design algorithm for congested corridors.

The resulting services and their frequencies and loads are shown in Table III for each stage of the experiment.

Table III – Resulting services for each stage

Stage	**Service**	**N-S Direction**	**S-N Direction**	**Frequency (bus/hr)**	**Max load (pax/bus)**
1	1	oooooooooooooooooooo	--------------------	51.64	273.06
1	2	--------------------	oooooooooooooooooooo	51.64	58.11
2	1	oooooooooooooooooooo	--------------------	34.83	197.72

2	2	--------------------	oooooooooooooooooooo	69.82	42.98
2	3	oo------------oooooo	--------------------	12.47	210.51
2	4	oo-------------ooooo	--------------------	16.77	195.22
2	5	oo----------------oo	--------------------	1.52	128.84
2	6	oooo-----------ooooo	--------------------	4.23	264.71
3	1	oooooooooooooooooooo	--------------------	88.13	160
3	2	--------------------	oooooooooooooooooooo	88.13	34.05
4	1	oooooooooooooooooooo	--------------------	33.00	157.98
4	2	--------------------	oooooooooooooooooooo	89.00	33.72
4	3	oo------------oooooo	--------------------	19.00	158.23
4	4	oooo-----------ooooo	--------------------	37.00	158.95
5	1	oooooooooooooooooooo	--------------------	32.48	160.15
5	2	--------------------	oooooooooooooooooooo	88.01	34.1
5	3	oo------------oooooo	--------------------	18.48	160.32
5	4	oooo-----------ooooo	--------------------	37.05	160.23
6	1	oooooooooooooooooooo	--------------------	41.55	160
6	2	--------------------	oooooooooooooooooooo	88.13	34.05
6	3	oo------------oooooo	--------------------	46.58	160

Table IV summarizes the resulting social costs for each stage.

Table IV – Resulting social costs per stage

Stage	Social costs ($/hr)	Social cost reduction	Iterations
1	8,507,920	-	-
2	7,852,817	7.7%	-
3	8,820,851	-	-
4	7,977,375	9.6%	90
5	7,973,753	9.6%	14
6	7,915,484	10.3%	14

For the uncongested stages the heuristic was able to find good services to feed the model, reducing 7.7% of the total social costs. As it can be seen in Table III, the services that the model creates for the unrestricted problem (stage 2, services 3 to 6) are very similar to each other, and one of those services (service 5) has a low frequency, suggesting that a very similar solution can be obtained using fewer services. To find this out, we made one additional run of the model, considering only services 1, 2 and 4. The results for this scenario are show in Table V:

Table V – Resulting services for each stage

Stage	Service	N-S Direction	S-N Direction	Frequency (bus/hr)	Max load (pax/bus)
2b	1	oooooooooooooooooooo	--------------------	37.54	196.35
2b	2	--------------------	oooooooooooooooooooo	70.37	42.98
2b	4	oo-------------ooooo	--------------------	32.82	205.06

The social costs for this solution are of 7,864,773.88 $/hr, that still represent a reduction of 7.6% of the social costs. It can be concluded that in fact the solution from the model can be simplified without losing that much savings. It also can be observed for the uncongested scenarios that the maximum load on buses greatly exceeds the capacity we defined for this experiment, so it was to be expected that the solutions from the scenarios with active capacity would have higher frequencies and social costs.

For the congested scenarios it can be seen that every heuristic was able to improve the base case solution by around 10%. Solutions for stages 4 and 5 are practically the same (slight differences on the solutions comes from working with a fixed step), but differ considerably in the number of iterations involved, confirming that the modification on the capacity adjusting heuristic makes it noticeably faster. For this experiment, stage 6 found an improvement to the solution of stages 4 and 5. Anyway, if we take a look at the services involved in the solution, it is possible to see that the solution from stage 6 was achievable by the capacity algorithm but it took another direction at one point, leading to a different solution, providing an example of one instance where Leiva's algorithm (and its modified version) leads to an suboptimal solution. Anyway, the solution that these algorithms find is still significantly better than the base case solution.

It is also worth noting that the costs from solutions 4 to 6 are lower than the cost from solution 1. This means that in this case express services could be used both as a way to lower costs and increase capacity at the same time.

CONCLUSION

In this work we introduced some improvements to the frequency optimization problem for the express service design problem, and two new algorithms for the service design problem. We showed an implementation of the model on a 40 stop bidirectional corridor, where the model was able to generate the services and find solutions that yielded savings in social costs around 10% respect to the solution where no express services were used.

We also showed that it is possible to tackle the capacity issue in an analytic way (and respecting the way users behave) for specific service configurations, creating indicators to set frequencies and design services simultaneously. In this scenario, this methodeven beat the other iterative heuristics that worked on a broader set of services, which mean that in general it is worth trying both approaches to solve the problem. This also suggests that it is worth exploring other services configurations to solve the capacity problem analytically. We are currently studying short turning services and the combination of these services with superexpress services, and preliminary results look very promising.

The service design problem can be faced in many other ways than the ones presented here. We are currently exploring several approaches. One of these approaches consists in computing an indicator for every bus stop measuring the effect of skipping it on a given service, and then building services heuristically using this indicator. Another approach would be to start with a given solution and its travel times for each pair, and then generate a new indicator based on the time savings that every skipped stop could yield given these current times. One more idea that has given good preliminary results is to identify the beneficial short turning services for the corridor, and then using them as starting points with the other methods to build shorter express services.

The following steps for this ongoing research would be to design and execute a bigger experiment, consisting on a greater number of scenarios to optimize, using a generation algorithm that combines many heuristics for the service design problem in order to tell apart which parts of the algorithm give better results, and in which cases they are worth being used. This algorithm will then be generalized to deal with transfers and with gradually more complex network configurations.

REFERENCES

Baaj, M.H. and H.S. Mahmassani (1992).AI-Based System Representation and Search Procedures for Transit Route Network Design. Transportation Research Record, 1358, 67-70.

Baaj, M.H. and H.S. Mahmassani (1995).A Hybrid Route Generation Heuristic Algorithm for the Design of Transit Networks. Transportation Research C, 3C(1), 31-50.

Ceder, A. and H. I. Stern (1981).Deficit Function Bus Scheduling with Deadheading Trip Insertion for Fleet Size Reduction. Transportation Science, 15(4), 338-363.

Ceder, A. andN.Wilson (1986).Bus Network Design. Transportation Research B, 20(4), 331-44.

Ceder, A. (2003).Designing Public Transport Network and Routes. In: Lam, W.H.K. and M.G.H. Bell (eds.). Advanced Modeling for Transit Operations and Service Planning. Pergamon, 59-92.

Desaulniers, G. and M. Hickman (2007).Public Transit. In: Barnhart, C. and Laporte, G., (eds.). Handbooks in Operations Research and Management Science, Vol. 14, Transportation. Amsterdam, North-Holland, 69-128.

Fan, W. and R. Machemehl (2006). Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem.Journal of Transportation Engineering, 132(2), 122-32.

Fernández, J. E., J. De Cea and H. Malbran (2008). Demand Responsive Urban Public Transport System Design: Methodology and Application. Transportation Research A, 42(7)A, 951-972.

Fernández, J.E., J. De Cea and I. Norambuena (2003).UnaMetodologíapara el DiseñoTopológico de Sistemas de TransportePúblicoUrbano de Pasajeros.In Proceedings of the XI CongresoChileno de Ingeniería de Transporte, Santiago de Chile.

Furth, P.G. and F.B. Day, (1985). Transit Routing and Scheduling Strategies for Heavy Demand Corridors. Transportation Research Record, 1011, 23-26.

Furth, P.G. (1987). Short Turning on Transit Routes. Transportation Research Record, 1108, 42-52.

Guihare, V. and J. K. Hao (2008). Transit Network Design and Scheduling: A Global Review. Transportation Research A, 42A(10), 1251-1273.

Jansson, J. O. (1980). A Simple Bus Line Model for Optimization of Service Frequency and Bus Size.Journal of Transport Economics and Policy, 14, 53-80.

Larrain H., R.Giesen and J.C.Muñoz (2010). Choosing the Right Express Services for Bus Corridor WithCapacity Restrictions. Transportation Research Record, 2197, 63-70.

Leblanc, L. (1988). Transit System Network Design. Transportation Research B, 22B(5), 383-390.

Levinson, H. S., S. Zimmerman, J. Clinger and C. S. Rutherford (2002). Bus Rapid Transit: An Overview. Journal of Public Transportation, 5(2).

Leiva, C., J. C. Muñoz, R. Giesen, and H. Larrain.(2010) Design of Limited-Stop for an Urban Bus Corridor with Capacity Constraints. Transportation Research B, 44B(10), 1186-1201.

Mauttone, A. and M. E. Urquhart (2009).A Route Set Construction Algorithm for the Transit Network Design Problem.Computers& Operations Research, 36(8), 2440-2449.

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation.American EconomicReview, 62, 591-604.

Turnquist, M. A. (1979). Zone Scheduling of Urban Bus Routes. Transportation Engineering Journal, 105(1), 1-13.