A NEW SOLUTION FRAMEWORK FOR THE LIMITED-STOP BUS SERVICE DESIGN PROBLEM

Guillermo Soto, Pontificia Universidad Católica de Chile gbsoto@uc.cl
Homero Larrain, Pontificia Universidad Católica de Chile homero@ing.puc.cl
Juan Carlos Muñoz, Pontificia Universidad Católica de Chile jcm@ing.puc.cl

ABSTRACT

In this study, we present a framework for addressing the limited-stop service design problem over a corridor. Using a bi-level optimization approach, we introduce a method of designing these services while considering bus capacity, transfers, and two behavioral models for passengers: deterministic and stochastic. The algorithms were tested on nine scenarios with up to 80 stops. Working with deterministic passenger assignment, our model solved the problem in a small fraction of the time required by a benchmark algorithm. We finally show that although it makes the problem much harder, working with stochastic assignment leads to more realistic and robust solutions.

Keywords: limited-stop service; transit network design; stochastic passenger assignment.

1. INTRODUCTION

As demand and modal share of trips on public transport keeps rising throughout the world, the need for fast and reliable public transport systems with high quality standards for its users becomes more important. Bus Rapid Transit (BRT), which can be defined as a "high-quality, customer-orientated transit that delivers fast, comfortable and cost-effective urban mobility" (Wright and Hook, 2007), is a mass transportation alternative which has gained attention and popularity particularly during the last decade. Besides the emblematic cases of Curitiba and Bogotá, there are currently more than 200 cities around the world which have adopted BRT systems on their main arteries. Furthermore, new BRT corridors keep popping up in every corner of the world every day: nearly one third of these cities launched their BRT systems after the year 2010 (www.brtdata.org).

A key element in BRT allowing to provide fast rides while making an efficient use of a bus fleet, is the correct utilization of limited-stop services (services that skip a set of stops in the corridor). Thus, counting on a reliable tool for designing efficient limited-stop services is of special importance in the light of the rise of BRT systems. During the last decade, many authors have proposed different methodologies for the design of limited-stop services (see Sun et al, 2008; Leiva et al, 2010; Chiraphadhanakul and Barnhart, 2013, Larrain et al, 2015 and Chen et al, 2015). Different models and methodologies work under a wide range of assumptions, however, there is one particular assumption that all of these authors seem to share: the deterministic nature of passenger assignment.

Deterministic passenger assignment is a very convenient assumption which simplifies the formulation of the problem. In absence of capacity limitations (such as the bus capacity), it allows to formulate a mathematical programming problems that determines the optimal frequencies of a set of services while assigning passengers to the minimum cost route. However, this type of assignment (also known as “all or nothing”) has one important shortcoming: it assumes that every passenger will always take his/her shortest path, meaning that a slightly worse option will not carry any passengers at all. This triggers stability problems with the optimal solutions of the model.

To deal with this instability in the assignment caused by small changes in frequency, we propose using a stochastic assignment. However, this poses some new challenges. Since there is no longer a natural way to solve the network design problem simultaneously, we separate the problem and solve it using the methodology suggested by Larrain (2013). We propose a framework for the limited-stop service design problem (LSDP) over a corridor where the problem is divided into the limited-stop service generation problem (LSGP), and the capacitated frequency optimization and passenger assignment problem (CFOAP).

2. SOLUTION FRAMEWORK FOR THE LIMITED-STOP SERVICE DESIGN PROBLEM

The CFOAP can be presented in a generic way as the problem of minimizing a social cost function, subject to three groups of constraints: structural constraints, capacity constraints, and passenger behavior constraints. The input for this problem is a set of services previously defined by solving the LSGP or alternatively by an expert, an origin-destination (O-D) trip matrix, and some other key parameters.

In this framework, structural constraints encompass all the constraints that make the frequencies of a solution feasible: non-negativity and frequency conservation. The role of capacity constraints is to ensure that bus capacity (and possibly other types of capacity limitations) is not exceeded by the solution. The last group, passenger behavior constraints, guarantees that passengers are assigned to routes that are consistent to a selected behavioral model, and not just the routes that minimize social costs. In the deterministic case where passengers travel through their minimum cost option, behavioral constraints can be dropped only when i) there are no capacity constraints (or no capacity is binding) and ii) the user cost function in the total cost being minimized coincides with the cost individuals minimize to reach their destinations.

The solution of the CFOAP can tackled by first solving the uncapacitated version of the problem (FOAP) and then applying some heuristic approach to find a solution where capacity constraints are not violated, and where passenger behavior constraints are also met. Two greedy heuristics for capacity, which rely on solving the FOAP iteratively setting lower bounds to the frequencies, are proposed in Leiva et al (2010) and Larrain (2013). In this work we solve the FOAP using a bi-level formulation that makes it possible to model passenger assignment as a stochastic process. The problem is divided into a frequency optimization stage (FOP) where passenger assignment is fixed, and a passenger assignment problem (PAP) where frequencies are fixed. This separation of the problem in two levels makes it possible to test different behavioral models for passengers.

In this context the FOP consists on minimizing a social cost function subject only to structural constraints. This is a non-linear problem, but it is not hard to solve for instances inspired in real sized corridors. The PAP, on the other hand, can be formulated as deterministic or stochastic. In this work, we consider that in the presence of parallel services passengers will choose a set of attractive services (which is a subset of the services connecting a given O-D pair) to perform their journey. It was proven by Chriqui and Robillard (1975) that if a service A belongs to the set of attractive services for a desired trip, a service B which connects the same O-D pair with a lower travel time will also belong to the set of attractive services, regardless of its frequency. In simple words: if you are willing to take service A, but a faster service B shows up first, you will take it no matter how infrequent this service is.

To solve the deterministic PAP, first we compute for each O-D pair the expected travel time associated to every possible set of attractive services. Then, we find the set of attractive services that minimizes this value using the methodology proposed by Larrain (2013), and build an auxiliary network where every O-D pair is connected by an arc with travel time equal to this value plus a transfer cost. Finally, trips are assigned to their minimal route over this network, and passengers are split among the attractive services on every arc in proportion to the frequency of the services.

The stochastic PAP, in turn, is solved by including randomness in two steps of the algorithm just described. First, for a given O-D pair, passengers will choose one of the reasonable sets of attractive lines following a multinomial logit model. The cost of an arc in the auxiliary network will be represented by the expected maximum utility of the pair. Then, we assume that passengers' route choice over this network can be also modeled as a multinomial logit. This last step of the assignment process can be performed using Dial's algorithm, which does not rely on route enumeration (Dial, 1971). The assigned trip over the network are finally translated into bus loads.

3. METHODOLOGY

In this section, we explain our CFOAP solution framework. This problem seeks to determine cost-minimizing frequencies for a given set of services on a bidirectional corridor, subject to bus capacity constraints and where users choose their most convenient routes. To solve this problem, we take a solution that omits capacity (FOAP), and then we iteratively adjust this solution to account for this constraint. The FOAP solution, meanwhile, is reached by decomposing the problem into two parts: passenger assignment in a corridor with a given set of services and frequencies (PAP), and the optimization of frequencies, given a specific pattern of passenger assignment throughout the network (FOP). In Section 3.1, we describe the FOAP and our proposed solution algorithm. Within this section, we formally present both the deterministic and stochastic versions of the two subproblems (FOP and PAP) and explain the proposed solution algorithms for each case. In Section 3.2, we explain how to obtain a feasible solution for the CFOAP using an iterated FOAP solution.

3.1 The frequency optimization and assignment problem

The dynamics underlying the FOAP, in its general version, can be understood as a Stackelberg game, where the leader (in our case the planner) defines the frequencies of all available services, and followers (i.e., system users) choose which services to use. Thus, to solve the FOAP, we divide the problem into two parts: first, the problem that a planner solves, which we have called the FOP, and second, the problem faced by users, which we have called the PAP.

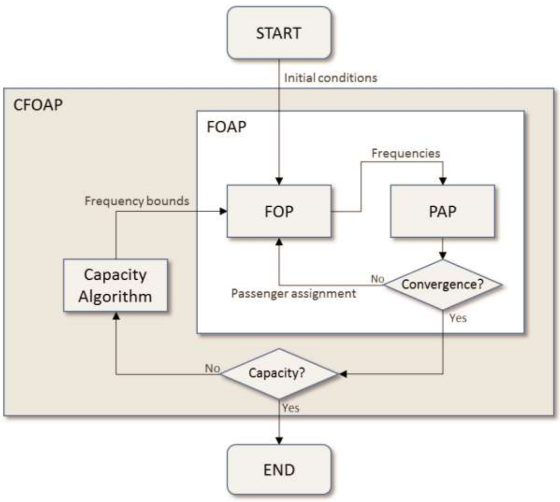

Fig. 1. CFOAP solution algorithm diagram.

The structure of the FOAP and its subproblems suggest the adoption of a bi-level solution scheme, in which the FOP and PAP are solved alternately, forming a feedback loop. This proposed solution is illustrated in Figure 1, inside the box corresponding to the FOAP. Given a set of services, the FOP optimizes frequencies, assuming that passenger assignment is known. In this context, assignment is understood as the sequence of trip legs followed by each passenger of the given O-D matrix. The PAP meanwhile, receives frequencies from the FOP and closes the cycle by using these frequencies to determine the passenger assignment. This cycle continues until the convergence of frequencies is reached. As mentioned above, it is possible to model a passenger's route choice process (i.e., the process of assigning passengers to different services) in different ways. In this paper, we study two types of passenger assignment: deterministic and stochastic. These types of assignments affect the models of both FOAP subproblems and are detailed in the sections that follow.

3.1.1 The passenger assignment problem

Predicting how users choose their routes to reach their destinations, given a set of services and frequencies, can be approached in many ways. Passenger assignment models can be classified as either deterministic or stochastic. In models of the first type, we assume perfect knowledge of users' perceived utility, and therefore, we assume that it is possible to deterministically predict a user's choice. In stochastic models we acknowledge limited and imperfect information, and hence the selection process is modeled probabilistically. That is, instead of predicting which alternative every user will choose, we calculate the probability that a user might choose a given alternative.

The decision process of a passenger can be modeled in several ways. This means that its stochasticity can be also incorporated in various ways. In this study, we assume that the decision-making process of a passenger follows two stages:

- Service selection. This stage determines which services will be utilized by different passengers taking a direct trip in a given route section. In this context, a route section is any combination of origin and destination nodes that is connected by an existing service. Suppose a rider travels on a given route section. To make this trip, there is a set of services that the rider could use, each with its own associated travel time and frequency. Which services should the rider be willing to take when they arrive? In the presence of limited-stop services, the answer to this question is not easy, as the rider might avoid boarding some slower services to wait for a faster one to come.
- Route selection. All trips on any given O-D pair will be structured as a sequence of trips over route sections. Given the expected travel costs of the possible route sections obtained from solving the previous stage, passengers choose a route that minimizes their expected total costs.

In what follows, we will propose a formulation and solution method for each of these two stages, both for the deterministic case and the stochastic case.

3.1.1.1 Service selection stage

Given a route section $s \in \mathcal{S}$, let $\mathcal{A}_s$ be the set of services that allows a direct trip (i.e., with no transfers) between the origin and destination of s. For each service $l \in \mathcal{A}_s$, travel time is t_l^s and frequency is f_l. Travel time can be computed using the length of the route section, the operating speed of buses, and the stop configuration of the service. There are two outcomes of the service selection stage: the expected travel cost for route section s, ETC_s, and the probability $\Pr\{l, s\}$ that any user traveling over s uses the service $l \in \mathcal{A}_s$. We will detail two approaches to estimate these values deterministically and stochastically.

Deterministic case: Let ETC_a^s be the expected travel cost for the route section s by considering the set of services $a \subseteq \mathcal{A}_s$. The value of the expected travel cost can be estimated using the following equation:

$$ETC_a^s = \theta_{TT} \frac{\sum_{l \in a} f_l \, t_l^s}{\sum_{l \in a} f_l} + \theta_{TW} \frac{\kappa}{\sum_{l \in a} f_l} \tag{1}$$

The first term of this equation corresponds to the expected travel time, modeled as the weighted average of the travel time of the alternative options with respect to their frequencies, multiplied by the value of travel time θ_{TT}. The second term refers to the expected waiting time for the value of waiting time θ_{TW}. Here, we assume the expected wait times to be proportional to the inverse of the total frequency for services in a. The proportionality factor, κ, takes the value of 1 when bus arrivals follow a Poisson process.

In the study by Chriqui and Robillard (1975), an efficient method for determining the set of lines that minimizes ETC_a^s is proposed. This method is based on the observation that if a rider is willing to use a service, this rider should also be willing to take any faster service, if it arrives at the bus stop earlier. The output of Chriqui and Robillard's algorithm for each route section s is a set of attractive services, a_s^*, and the value of the expected travel cost associated, $ETC_s = ETC_{a_s^*}^s$. Probabilities $\Pr\{l, s\}$ are calculated by assigning trips proportionally to the frequency of each service (i.e $\Pr\{l, s\} = f_l / \sum_{l' \in a_s^*} f_{l'}$ when $l \in a_s^*$, and $\Pr\{l, s\} = 0$ for 0, $l \notin a_s^*$.

Stochastic case: In the stochastic version of the service selection problem, there is uncertainty in the users' perceptions of travel cost components. Different users may choose different sets for the same trip. This implies that instead of predicting with certainty which subset of services will minimize these costs, we will estimate the probability that a subset is the most attractive.

Suppose there is a set of possible sets of attractive services, $A_s = \left\{a_s^1, \dots, a_s^{K_s}\right\}$. Associated with each alternative a_s^k from this set there is a utility U_s^k, which corresponds to the expected cost of selecting it as the set of services. This utility can be modeled as $U_s^k = V_s^k + \varepsilon$, consisting of a deterministic element V_s^k, which can be measured by the planner, and a random component ε, which represents the uncertain part of the utility. We model the deterministic part as $V_s^k = ETC_{a_s^k}^s$, which corresponds to the expected travel cost of using subset a_s^k, calculated using (1). Assuming that errors ε are random *i.i.d.* variables that follow a Gumbel distribution, the probability $\Pr\{k, s\}$ of users choosing an alternative k can be estimated using a multinomial logit model:

$$\Pr\{k, s\} = \frac{\exp\left(-\lambda \cdot V_s^k\right)}{\sum_{k'=1}^{K_s} \exp\left(-\lambda \cdot V_s^{k'}\right)} \tag{2}$$

In this equation, λ is the scale parameter of the logit model, and is inversely proportional to the standard deviation of ε. To ensure that variables ε are in effect independent, the set A_s must be constructed in a way that ensures the independence of the options. One way of doing this is to define A_s as a partition of $\mathcal{A}_s$, the set of services connecting s. In the model by De Cea and Fernandez (1993), passenger assignment in the presence of congestion is modeled by grouping services into subsets that satisfy the requirement of independence of options. These sets of services, which we implement in our model, are built using the following procedure:

1. Define a counter $k = 1$ and an auxiliary set of services $A' \leftarrow \mathcal{A}_s$.
2. Build a_s^k by applying the Chriqui and Robillard algorithm on the set of services A_s.
3. Remove the new subset from A', i.e., do $A' \leftarrow A' \setminus a_s^k$, and update the counter $k \leftarrow k + 1$.
4. Repeat steps 2 and 3 until there are no more remaining services.

By applying this algorithm, we construct the set A_s for every route section s. The cost associated with the route section s should be calculated based on the expected maximum utility of the route section, which is given by the following equation:

$$ETC_s = -\frac{1}{\lambda}\ln\left(\sum_{k=1}^{K_s} \exp\left(-\lambda \cdot ETC_{a_s^k}^s\right)\right) \tag{3}$$

The probability of use for a service $l \in a_s^k$ (given that the elements of A_s are disjointed) will correspond to the probability of choosing the set a_s^k, $\Pr\{k, s\}$ (which can be computed using (2)) times the ratio of the corresponding frequencies: $\Pr\{l, s\} = \Pr\{k, s\} \cdot f_l / \sum_{l' \in a_s^k} f_{l'}$.

3.1.1.2 Route selection stage

For a directed network over the set of nodes $\mathcal{N}$ of a public transport corridor, let $\mathcal{W}$ be the set of O-D pairs from that corridor. For each route section $s \in \mathcal{S}$, the expected travel cost for each route section, ETC_s, and the demand for each O-D pair, T_w, are known. The route selection problem seeks to predict what route sections and transfers are made by users for each O-D pair $w \in \mathcal{W}$. The outcome of this stage is H_s, the flow over route section s.

Since bus capacities are relaxed, the problem of route selection can be solved separately for each O-D pair w, with origin and destination nodes denoted as O_w and D_w. We build an auxiliary graph containing an arc for each route section $s \in S_w$, with $S_w = \{s = (i,j) : O_w \leq i < j \leq D_w\}$. In other words, the auxiliary graph contains all the route sections s that could be used as trip segments for the pair w without changing the direction of the trip. In this auxiliary graph, the cost of a route section s is ETC_s if s originates in O_w, and $ETC_s + \theta_{Tr}$ in any other case. In the latter equation, θ_{Tr} corresponds to a fixed cost incurred for each transfer.

Deterministic case: To model the route choice deterministically, it is sufficient to find the shortest paths for each O-D pair. This can be done by using a shortest path algorithm like Dijkstra's, which finds a shortest path tree to all the remaining nodes, once from every node. This process can be made more efficient by using an efficient version of the shortest path algorithm that takes advantage of the acyclic structure of the auxiliary graphs.

From the results of the shortest path algorithm, it is possible to obtain the resulting flow over section s, H_s. Let H_s^w be the flow of users that use route section s when making a trip on O-D pair w. This value can be easily obtained: it takes on the value of T_w when s belongs to the optimum route of w, and 0 if it does not. With this, H_s corresponds to $H_s = \sum_{w \in W} H_s^w$.

Passenger flow H_s^l associated with each line l and route section s can simply be calculated as the flow H_s times the probability (computed as in the deterministic case) that line l is used for section s using the equation $H_s^l = \Pr\{l, s\} \cdot H_s$.

<u>Stochastic case</u>: We can model stochastic route choices assuming that users make these choices according to a random utility logit model similar to the one used for the selection of services, but with a scale parameter β. Dial's algorithm (1971) solves this problem for a specific O-D pair w without requiring the enumeration of routes. This saves a great deal of computational effort and allows for a direct calculation of H_s^w. Flows H_s and H_s^l are obtained in the same way as in the deterministic.

3.1.2 Frequency optimization problem

This problem estimates, for a given passenger assignment, the frequencies that minimize system costs. Additionally, since the model will be embedded into the algorithms addressing bus capacity (as described in the following section), we add lower bound constraints on the frequencies to the model. This problem can be expressed as follows:

$$\min_{\boldsymbol{f}} Z = \left[\sum_{l \in \mathcal{L}} c_l f_l + \sum_{s \in S} H_s \cdot ETC_s(\boldsymbol{f}) \right] \tag{4}$$

s.t

$$\sum_{l \in \mathcal{L}_i^+} f_l = \sum_{l \in \mathcal{L}_i^-} f_l, \qquad \forall i \in \mathcal{N} \tag{5}$$

$$f_l \geq f_l^{min}, \qquad \forall l \in \mathcal{L} \tag{6}$$

The objective function of this problem corresponds to an approximation of the social costs for a known passenger assignment, which is estimated as the sum of two components. The first term corresponds to operator costs, which are obtained by assuming that each service ($l \in \mathcal{L}$) has an operational cost proportional to its frequency. The second term corresponds to the total travel costs in the route section s. The term $ETC_s(\boldsymbol{f})$ corresponds to the expected travel cost of segment s when the frequencies are $\boldsymbol{f}$, and it can be computed using both stochastic and deterministic assignments. It should be noted that even though transfers are included in the FOAP, transfers are a constant in this subproblem (as passenger assignment is known) and are therefore left out of the objective function.

The set of constraints (5) imposes the conservation of bus frequency at bus stops. Sets $\mathcal{L}_i^+$ and $\mathcal{L}_i^-$ contain services that begin and end their journey in a bus stop $i \in \mathcal{N}$. The second set of constraints (6) ensures that the frequency of service is greater than a certain given lower bound. This constraint is used in the capacity adjustment algorithms that solve the CFOAP.

In the objective function, the term $ETC_s(\boldsymbol{f})$ must be consistent with the type of passenger assignment assumed. As this term is a function of the frequency of the system, it is necessary to incorporate its definition as a constraint to the models (4)–(6). Specifically, if the assignment is deterministic, the formulation of ETC_s is determined using equation (1) and set a_s^* obtained as a result of the assignment. If however, the assignment criterion is stochastic, ETC_s must be computed using equation (3).

3.2 Addressing bus capacity

Our algorithm for solving the version of the problem considering bus capacity constraints (CFOAP) consists of an iterative procedure based on the solution of its uncapacitated version (FOAP). An outline of the algorithm is shown in Figure 1. In each iteration, after reaching convergence on the FOAP, it determines whether the current solution exceeds the maximum bus capacity somewhere. For each unidirectional line segment $q \in Q$ of the corridor (a line segment is defined as a pair of consecutive stops on the corridor), the load profile is calculated as $P_q^l = \sum_{s \in S} \varsigma_q^l H_s^l$ (where ς_q^l takes value 1 if the route section s passes through line segment q), and then the maximum load for each service is obtained with the equation $P_l^{max} = \max_{q \in Q} P_q^l$.

The capacity deficit of a service can be obtained from its maximum load, frequency, and bus capacity k as $D_l \coloneqq k - P_l^{max}/f_l$. A solution for the CFOAP is feasible in terms of capacity if $D_l \geq 0$ for every service $l \in \mathcal{L}$. If the solution at any iteration is not feasible, the algorithm adds a minimum frequency constraint associated to a service in deficit. The criteria by which this service is chosen lead to different capacity algorithms.

Larrain (2013) proposes a greedy algorithm consisting in simply picking the service l' with the largest deficit, and fixing it by imposing on the next iteration the following lower bound (6) on its frequency, as follows: $f_{l'}^{min} = P_{l'}^{max}/k$. This greedy approach yields a single solution that is not guaranteed to be optimal. One way of improving this approach is to add randomness in the selection of the service to be selected. This type of algorithm, known as greedy randomized adaptive search procedure (GRASP), explores the domain of the problem in the search of better solutions. In our implementation, the probability of choosing a bus service was proportional to the deficit (only for services with $D_l > 0$). Naturally, this strategy relies on executing the algorithm several times to beat the greedy solution; therefore, it is much more time consuming than its counterpart.

4. COMPUTATIONAL EXPERIMENTS

In this section we describe the computational experiments we performed to test our algorithm. We start by describing in Section 4.1 the scenarios we constructed using data from Santiago, Chile and Bogota, Colombia. The results of these experiments are detailed in Section 4.2.

4.1 Corridors and scenarios

To test and compare the CFOAP solution algorithms, we have created nine scenarios based on three real transit corridors. The first corridor is on Av. Pajaritos in Santiago, Chile. The second one is on Av. Grecia, also in Santiago, Chile. The third corridor is on Av. Caracas, the corridor with the highest frequency and demand of the Transmilenio system in Bogota, Colombia.

To compare the performance of the different versions of our algorithm in different-sized problems, the original O-D matrices for the three corridors were adapted to fit 20, 40, and 80 total bus stops (counting both directions) while keeping the length of each of the three corridors constant. The nine scenarios we used to test our algorithms and their main attributes are summarized in Table 1.

Table 1. Descriptions of experimental scenarios.

Scenario	P20	P40	P80	G20	G40	G80	C20	C40	C80
Corridor name	Pajaritos	Pajaritos	Pajaritos	Grecia	Grecia	Grecia	Caracas	Caracas	Caracas
Number of stops	20	40	80	20	40	80	20	40	80
Number of O-D pairs	90	380	1,560	90	380	1,560	90	380	1,560
Corridor length* (Km)	8	8	8	10	10	10	30	30	30
Operational speed (Km/h)	22	22	22	25	25	25	26	26	26
Total trips (pax/h)	20,546	20,527	20,529	37,728	37,719	37,810	43,562	43,564	43,551
Maximum load (pax/h)	14,119	14,163	14,372	13,392	13,223	13,335	19,179	19,082	19,004
Number of services	11	27	56	26	31	31	17	22	41

*per direction.

The name of each scenario is a combination of the initial of the corridor name and the number of total stops, which are their defining attributes. It is important to note that the maximum load for Av. Caracas is close to 19,000 pax/h, far from the 48,000 pax/h reported in Global BRT Data (BRT Centre of Excellence, 2017).

The last row in Table 1 indicates the number of a priori services considered for the solution to the CFOAP. These services come from a solution of the LSGP (the service generation problem) obtained by using the algorithms reported by Larrain (2013) and Larrain et al. (2015). Table 2 presents the parameters used in the passenger assignment and optimization for each corridor.

Table 2. Parameters for the experiments.

Parameter	Pajaritos	Grecia	Caracas
Value of in-vehicle travel time ($/min), θ_{TT}	0.15	0.15	0.08
Value of waiting time ($/min), θ_{WT}	0.15	0.15	0.05
Transfer cost ($), θ_{Tr}	0.60	0.60	0.15
Scale parameter λ	0.00176	0.00176	0.00340
Scale parameter β	0.00224	0.00224	0.00440

Service regularity parameter, κ	1	1	1
Dwell time (min), τ	1	1	1
Operating costs - distance ($/bus-km), c_L	0.75	0.75	0.75
Operating costs - time ($/bus-h), c_T	7.52	7.52	7.52
Bus capacity (pax/bus), k	80	72	120

The values of parameters θ_{TT}, θ_{WT}, and θ_{Tr} were obtained from the study by Batarce et al. (2015), which calibrated them for public transit route choice models for both Santiago and Bogota. In this study these parameters depended on a comfort measure, calculated as bus passengers per square meter, which in our case we assumed to be of 5–6 pax/m^2 and 5 pax/m^2 in Santiago and Bogota. As we could not find values for the scale parameters λ and β calibrated for scenarios comparable to ours, they were obtained by setting a reasonable deviation for the error. More precisely, λ was calibrated to ensure that given two services to choose from with a difference of 5 min in travel time, 70% of passengers would opt for the fastest service and 30% for the other. Parameter β was calibrated to ensure that given a difference of 10 min in travel time between two routes, 90% of passengers would opt for the fastest route and 10% for the other route.

4.2 Experimental results

This section presents the results of the experiments carried out for each scenario. The experiment shows how the algorithms work in both the stochastic and deterministic versions.
The bi-level solution algorithms were coded using the Visual Studio C# language. The FOP sub-problems within these algorithms were solved in AMPL, as was the model associated with algorithm A0. All algorithms were run on a personal computer with the following processor: Intel Core i7-4510, CPU @ 2.00 GHz, 2.6 GHz with 8.00 G 0 RAM. The tolerance for convergence was set at $\alpha = 10^{-7}$ for the algorithms.

4.2.1 Performance of the algorithms

The scenarios defined above were optimized using the different types of algorithms described in this paper. These algorithms are summarized in Table 3. We use the A0 algorithm as a benchmark in this experiment. It solves the FOAP directly as an optimization problem without separating it into subproblems and applies a greedy heuristic for capacity adjustment. This algorithm and the optimization model behind it are explained in detail in the study by Larrain (2013). As this algorithm relies heavily on the capacity of a nonlinear solver, solutions for the case allowing transfers were not very reliable; therefore, we report the solution of the case with no transfers.

Table 3. Algorithms.

Algorithm	Optimization strategy	Passenger assignment	Capacity heuristic	Transfers allowed
A0	Simultaneous	Deterministic	Greedy	No
D / g / N	Bi-level	Deterministic	Greedy	No
D / G / N	Bi-level	Deterministic	GRASP	No
D / g / T	Bi-level	Deterministic	Greedy	Yes
D / G / T	Bi-level	Deterministic	GRASP	Yes
S / g / T	Bi-level	Stochastic	Greedy	Yes

S / G / T	Bi-level	Stochastic	GRASP	Yes

The other six algorithms use the bi-level algorithm described in Section 3 for the solution of the scenarios. These algorithms are named using three letters, denoting the nature of the assignment model (deterministic or stochastic), the type of capacity algorithm (lowercase g for greedy, uppercase for GRASP), and if transfers were allowed in the solution (transfers or no transfers). In this experiment, the deterministic and stochastic versions of the algorithm were tested in combination with two different capacity heuristics. As the benchmark algorithm A0 does not allow for transfers, we examine a version that excludes transfers for the deterministic case.

To compare the performance of the algorithms, we use the corrected total cost of the solutions (as defined by Larrain and Muñoz (2016)), which can be calculated as:

$$CTC = \sum_{l \in \mathcal{L}} c_l f_l + \sum_{w \in \mathcal{W}} \sum_{s \in S} H_s^w \, ETC_s + \theta_{Tr} \left(\sum_{w \in \mathcal{W}} \sum_{s \in S} H_s^w - \sum_{w \in \mathcal{W}} T_w \right) - \theta_{TT} \sum_{w \in \mathcal{W}} T_w \, t^w \tag{7}$$

The first two terms of equation (7) are the operator costs and the expected travel costs (considering in-vehicle travel times and waiting times), which are the same as in equation (4). The third term corresponds to the transfer costs, which are computed as the difference between the number of trip legs and demand for trips in the system times the unit cost θ_{Tr}. The last term is a fixed cost term corresponding to the fixed travel time portion that every passenger will bear independent of the solution of this problem. Since it constitutes a high proportion of the total costs, we subtract it to highlight the part of the cost where limited stop services can make a difference. This fixed term is the minimum in-vehicle travel time costs for each O-D pair, which are computed assuming that every passenger uses a direct non-stop express service with a travel time t_w, which can be estimated from the distance of the O-D pair w, the operating speed of the buses, and the dwell time τ.

Note that in equation (7), the value of the expected cost of traveling in each route section s, ETC_s, must be calculated in accordance with the appropriate assignment model, whether it be deterministic or stochastic. Table 4 presents the results for the different algorithms. As a reference, we present for each scenario the CTC of the optimal solution associated to using a single all-stop service in each direction. This solution can be obtained using either the A0 algorithm or using Mohring's square root formula (Mohring, 1972). For each algorithm, the corrected percentage savings (CPS) with respect to the all-stop CTC is shown.

Table 4. Results for deterministic and stochastic assignment algorithms.

	CTC ($/h)	Corrected percentage savings (CPS)						
Scenario	All-stop	A0	D / g / N	D / G / N	D / g / T	D / G / T	S / g / T	S / G / T
P20	16,920	48.5%	48.5%	48.5%	48.5%	48.5%	65.6%	65.6%
P40	33,847	56.2%	56.5%	56.5%	56.5%	56.5%	48.8%	48.8%
P80	68,353	63.7%	63.8%	63.8%	63.8%	63.8%	48.2%	**
G20	18,070	22.2%	22.2%	22.2%	22.2%	22.2%	48.7%	50.0%
G40	36,589	31.3%	31.2%	31.2%	31.3%	31.3%	27.6%	27.6%
G80	73,873	35.7%	35.6%	35.6%	38.0%	38.0%	22.8%	**
C20	22,142	13.2%	17.8%	17.8%	14.7%	18.3%	0.0%*	0.0%*

C40	36,540	30.2%	31.3%	31.6%	34.4%	34.5%	1.6%	2.6%
C80	172,995	79.4%	79.3%	79.6%	81.4%	81.4%	71.0%	**

*The algorithm was not able to improve the all-stop solution.
**The algorithm was not implemented because runtimes were too high.

Looking at the results from the deterministic cases, it can be observed that the savings increase as the corridors grow in number of stops. This result is consistent with what was observed by Larrain and Muñoz (2016), because as the number of stops increases, so does the amount of travel time limited-stop services can save. The D / g / N algorithm, yields results that are very close to the A0 algorithm. This result validates the bi-level approach, which despite not providing any guarantees of global optimality, obtains good solutions and even surpasses the benchmark algorithm in two instances (C20 and C40).

Comparing the greedy and GRASP solutions for the deterministic algorithms (i.e., D / g / N versus D / G / N and D / g / T to D / G / T), no major differences can be detected. This means that in general, the simple greedy algorithm is a decent approach for dealing with capacity. However, an exception occurs when comparing D / g / T to D / G / T in scenario C20, where GRASP gains an additional 3.6% in savings. This suggests that further improvement to the capacity heuristics is still worth looking into, possibly by performing a local search around the FOAP solutions.

Looking at the best deterministic solutions for algorithms with and without transfers (D / G / N and D / G / T), some scenarios (G80, C20, C40, and C80) show that additional savings (up to 2.9% in the best case) can be obtained when optimizing with transfers. Also, transfers appear to play a more relevant role in the design as the corridor demand grows. As for the stochastic algorithms listed in Table 4, the solutions present some unexpected trends, for example, in the way savings decrease with corridor size in Pajaritos but increase in Caracas. This is probably a symptom of the algorithm converging to suboptimal solutions. However, this algorithm still manages to find savings with respect to A0 in most cases.

Table 6 presents the runtimes of the seven modeled algorithms, measured in seconds. These results also show that the deterministic greedy algorithms (D / g / N and D / g / T) are considerably faster than algorithm A0 for medium and long corridors. This means that for the longest corridors, the proposed algorithms provide a similar output to A0 but in just a fraction of the time. For instance, in the toughest scenario, C80, D / g / T finds a slightly better solution than A0 in only 8% of the time. It is worth noting that including transfers in the deterministic assignment does not increase the runtime. On the contrary, is also clear that stochastic assignment algorithms require considerably more time to run than deterministic algorithms.

Table 6. Execution times for the different algorithms.

Scenario	A0	D / g / N	D / G / N*	D / g / T	D / G / T*	S / g / T	S / G / T*
P20	1	3	59 (10)	4	88 (10)	9	189 (10)
P40	32	11	232 (10)	11	378 (10)	161	973 (5)
P80	984	48	1,318 (10)	8	197 (10)	1,912	**
G20	3	28	222 (10)	9	267 (10)	127	1005 (10)
G40	31	4	86 (10)	3	79 (10)	40	349 (5)
G80	169	28	411 (10)	9	235 (10)	9,289	**
C20	17	10	128 (10)	3	135 (10)	10	135 (10)

C40	219	59	2,127 (10)	22	191 (10)	298	1573 (5)
C80	1,382	149	4,427 (10)	110	1364 (10)	17,241	**

*For the GRASP scenarios, the number of iterations is given in parenthesis.
**Algorithm not implemented because runtimes were too high.

4. CONCLUSIONS

In this study, we have formally presented the LSDP for a corridor, and proposed a solution framework. The solution procedure involves splitting the LSDP into two subproblems: a service generation problem (LSGP) for which we have previously proposed some solution algorithms in previous studies, and the capacitated frequency optimization and assignment problem (CFOAP). In this paper, we have focused on studying and developing an efficient solution method for the CFOAP, which is key to the LSDP solution.

One of the main accomplishments of this study was to improve the existing solution algorithms for the CFOAP, significantly reducing the solution times in the deterministic assignment case. This improvement is due to separating the FOAP (i.e., the uncapacitated version of the CFOAP) into two subproblems, and solving it using a bi-level approach. The implementation of GRASP heuristics for the capacity algorithm showed not only a significant improvement on the solution to the problem, but also revealed some room for future improvement. We believe that a local search algorithm would be a reasonable way to further improve the current capacity algorithms for the problem.

The new solution approach opens the possibility of addressing larger problems with a greater number of services to optimize, and to do so in significantly less time. For the scenario inspired in the Caracas Av. corridor in Bogota, our new algorithm reduced the runtime from 18 to 2 min. This opens the door for the development of new, more ambitious algorithms aimed at improving the LSDP solution and eventually extending this algorithm to the design of limited-stop services on networks of corridors, which is one of the most significant limitation of existing approaches for this design problem.

Another important contribution of this study is that it presents the first attempt at allowing for stochastic passenger assignment without limiting the structure or number of services. This extension results in more realistic and robust design solutions. Nevertheless, the solution algorithm we report does not work very efficiently on the stochastic case: this poses the challenge of finding a better solution algorithm tailored for the stochastic setting that would allow improvement on the robustness of the solution to the LSDP.

The improvements on the CFOAP algorithm also enable the optimization of scenarios in which users transfer between services. The experiment we report in this paper shows that allowing transfers in the solution leads to better designs, meaning that neglecting transfers can lead to suboptimal solutions to the problem and underestimating the benefits of limited stop services.

Future research should expand this study by incorporating the capacity limitations of bus stops to accommodate multiple buses, which has been ignored here for the sake of simplicity. This limitation is crucial to consider since they often become the bottleneck of the entire system as demand grows. The methods we introduced in this paper, possibly combined with some local search procedures, may open the way for a solution to this important variant of the LSDP. In the context of BRT systems facing a growing demand and increasingly challenging scenarios, well-designed limited-stop services are key to delivering the promise of a metro-like level of service. The models and solution algorithms presented in this paper contribute to the understanding of these complex systems.

ACKNOWLEDGMENTS

The authors wish to thank FONDECYT for its funding, project 11,140,443. This study was also supported by the BRT Center funded by the Volvo Research and Educational Foundations (VREF) and the Center for Sustainable Urban Development (CEDEUS), Conicyt / Fondap / 15110020.

REFERENCES

Batarce, M., Muñoz, J. C., Ortúzar, J. D. D., Raveau, S., Mojica, C., & Ríos Flores, R. A., 2015. Evaluation of Passenger Comfort in Bus Rapid Transit Systems. **Inter-American Development Bank.**

BRT Centre of Excellence, EMBARQ, IEA and SIBRT. "**Global BRTData**." Version 3.16. Last modified: August 4, 2016. Available at: http://www.brtdata.org

Chen, X., Hellinga, B., Chang, C. y Fu, L., 2012. Optimization of Headways for Bus Rapid Transit System with Stop-Skipping Control. **Transportation Research Board Annual Meeting 2012** Paper #12-1999.

Chiraphadhanakul, V. and C. Barnhart., 2013. "Incremental Bus Service Design: Combining Limited-Stop and Local Bus Services." **Public Transport**, 5(1-2), 53-78.

Chriqui, C. and P. Robillard., 1975. "Common bus lines". **Transportation Science**, 9, 115-121.
De Cea, J. and Fernández, E., 1993. "Transit assignment for congested public transport systems: an equilibrium model". **Transportation science**, 27(2), 133-147.

Dial, R. B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. **Transportation research**, 5(2), 83-111.

Larrain, H., 2013. "**Diseño de Servicios Expresos para Buses.**" PhD dissertation, Pontificia Universidad Católica de Chile.

Larrain, H., J.C. Muñoz and R. Giesen., 2015. "Generation and Design Heuristics for Zonal Express Services." **Transportation Research Part E**, 79, 201 – 212.

Larrain, H., J.C. Muñoz, 2016. "When and where are Limited-stop Bus Services Justified?" Transportmetrica A: **Transport Science**, 12(9), 811-831.

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**, 44(10), 1186-1201.