CIRRELT
Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport
Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

The Design Problem for Single-Line Demand-Adaptive Transit Systems

Fausto Errico
Teodor Gabriel Crainic
Federico Malucelli
Maddalena Nonato

October 2011

CIRRELT-2011-65

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

The Design Problem for Single-Line Demand-Adaptive Transit Systems

Fausto Errico[1,*], Teodor Gabriel Crainic[1], Federico Malucelli[2], Maddalena Nonato[3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

[2] Dipartimento di Elettronica e Informazion, Politechnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy

[3] Dipartimento di Ingegneria, Università di Ferrara, Via Saragat 1, 44122, Ferrara, Italy

Abstract. When demand for transportation is low or sparse, traditional transit cannot provide efficient and good-quality level service, due to their fixed structure. For this reason, mass transit is evolving towards some degree of flexibility. Although the extension of Dial-a-Ride systems to general public meets such need of adaptability, it presents several drawbacks mostly related to the extreme flexibility. Consequently, new transportation alternatives, such as Demand Adaptive Systems (DAS), combining characteristics from both the traditional transit and Dial-a-Ride, have been introduced. For their twofold nature, DAS require careful planning. We focus on tactical aspects of the planning process by formalizing the Single-line DAS Design Problem (SDDP) and proposing two alternative hierarchical decomposition approaches for its solution. The main motivation behind this work is to provide with a general methodology based on realistic assumptions and suitable to be used as a tool to build the tactical DAS plan in real-life conditions. We provide an experimental study where the two proposed decomposition methods are compared and the general behavior of the systems is analyzed when altering some design parameters.

Keywords. Public transit, demand-responsive systems, dial-a-ride, semi-flexible transportation, demand adaptive systems, planning, design.

Acknowledgements. Funding for this project has been provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Industrial Research Chair and Discovery Grants programs, by our partners CN, Rona, Alimentation Couche-Tard, the Ministry of Transportation of Québec, and by the Fonds de recherche du Québec - Nature et technologies (FQRNT) through its Team Research grants program.

---

* Corresponding author: Fausto.Errico@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2011

1 Introduction

When the demand for transportation is consistently high during a given time period, traditional transit operates well and efficiently as it naturally allows for high degrees of resource sharing and consequent good level of service. In contrast, when the demand for transportation is low or sparse, the resource sharing levels drastically drop, particularly because of the fixed structure of traditional transit services. For this reason, mass transit evolved towards some degree of flexibility. A first attempt in such a direction was made by extending the well-known Dial-A-Ride systems (DAR), originally designed to serve people with reduced mobility, to general customer service. With respect to traditional transit, DAR provides a more *personalized* service by modifying itineraries, schedules and stop locations according to the transportation needs of users at a given time. At the same time, it still guarantees a certain degree of resource sharing by serving requests *collectively*.

The adaptation of DAR to general public displays, however, a number of drawbacks, some of which follow from the extreme flexibility inherent in the system definition. Thus, for example, because the supply of transportation services changes according to needs expressed for particular time periods, neither the transit operator nor the users may predict the vehicle itineraries, stop locations, and associated schedules. As a consequence, users are obliged to book the service well in advance of the actual desired time of utilization and the actual pick up time is very much left to the discretion of the operator. For similar reasons, it is difficult to integrate DAR with traditional transit services.

With the purpose of addressing the above mentioned issues, a system different from DAR, denoted *Demand-Adaptive System* (*DAS*), was introduced by Malucelli et al. (1999) and then treated in more general contexts (Crainic et al. 2001, 2005). DASs are transit services displaying features of both traditional fixed-line bus services and purely on-demand systems such as DAR. A DAS bus line serves, on the one hand, a given set of *compulsory* stops according to a predefined *master schedule* specified by the time windows associated with each stop. This provides the traditional use of the transit line without in-advance reservations. On the other hand, similarly to DAR services, passengers may issue requests for transportation between two *optional* stops which induces detours in the vehicle routes. The fundamental idea behind DASs is that the time windows mechanism, introducing a degree of flexibility, provides a certain *regularity* of the service thus allowing users to plan their trips, simplifies integration with other transportation modes, and makes the service accessible also without reservation (at compulsory stops). Errico et al. (2011b) show that DAS is a general model for a large class of transit systems usually called *semi-flexible*.

Transportation systems dedicated to service several demands with the same vehicle generally require complex planning activities. The planning phase is very important because it deeply influences the overall behavior of the system. DAS, combining charac-

teristics of both traditional and on-demand system, requires both a service-design phase and an operational management that adjusts vehicle routes and schedules depending on actual user requests. In this paper, we focus on *tactical* aspects of the planning process, which we formalize as the *Single-line DAS Design Problem* (*SDDP*). The SDDP assumes that the territory to be covered by the DAS line has been determined, together with a set of potential stops, the road network and travel times between stops are known, and a measure of transportation demand among potential stops is available. For a given time horizon, which usually is seasonal, the SDDP is made up of several interrelated decisions regarding the selection of compulsory stops among all the potential stops in the territory, their sequencing, and the determination of the master schedule.

In the literature a few works related to the SDDP can be found, mostly focusing on some partial aspects, mainly the scheduling, (see, for e.g. Fu 2002; Quadrifoglio et al. 2006). Moreover, such works usually assume very simplified operating frameworks with the effect that they are mainly useful studies giving insights about the influence of certain planning parameters and providing basic approximations about the system behavior, rather than actual tools to adequately address the variety and complexity of real situations. The scope of the present paper is to fill this gap by providing the necessary methodological support to build a single DAS line without relying the methodology on restrictive assumptions.

More specifically, we provide a formal description of the SDDP and show that the ability to solve it for a time period with homogeneous characteristics (i.e., with constant demand), called time *slice*, plays a prominent role, representing the most challenging aspect of the overall solution process. We will denote the single slice SDDP as S-SDDP. For this reason our discussion and experimentation focuses on the S-SDDP. Given the extreme complexity of the S-SDDP, we propose a solution strategy based on the hierarchical decomposition of the associated decisions. The decomposition yields several *core* problems that need to be addressed (see Errico (2008), Crainic et al. (2010) and Errico et al. (2011a) for the details). We propose two possible hierarchical decompositions, called *Sequence Compulsory* and *Sequence All*. We then make an experimental study aiming at evaluating the impact of the tactical level decisions on the operational management measuring the performance parameters and comparing the two decomposition strategies.

After recalling the basic concepts of DAS comparing them with traditional transit and DAR and Semi-Flexible systems (Section 2), we present the main contributions of the paper. Namely, the formalization of the SDDP, the identification of the S-SDDP as the basic tool to build the design of a single DAS line in Section 3 where also a brief summary on related works is reported. Then the introduction of two different hierarchical decompositions for the S-SDDP (Section 4) and the experimental comparison underlining advantages and disadvantages of both methods, and a characterization of the system behavior in terms of several design parameters (Section 5). In Section 6, we show how

the methodology developed for the S-SDDP can be used to solve the SDDP. We finally report some concluding remarks in Section 7.

2 Public Transit, DAS and Planning Issues

In this section, after reviewing the main features of traditional transit, DAR, and semi-flexible systems (Section 2.1), we introduce the details of DAS (Section 2.2) and briefly recall the planning issues related to the deployment of DASs (Section 2.3).

2.1 Traditional transit, DAR and semi-flexible systems

Traditional transit services are particularly suited to handle situations where the demand for transportation is *strong*, i.e., when there is a consistently high demand over the territory and for the considered time period. The high degree of resource sharing by a large number of passengers makes it possible to efficiently provide high quality services, i.e., frequent, usually operating high-capacity vehicles over fixed routes and schedules. Routes and schedules may and do vary during the day, but, in almost all cases, they are not dynamically adjusted to the fluctuations of demand. In contrast, when the demand for transportation is *weak*, e.g., during out of rush-hour periods or in low-population density zones, operating a good-quality traditional transit system is very costly. The fixed structure of traditional transit services cannot economically and adequately respond to significant variations in the demand. In the presence of weak demand, itineraries and timetables may perfectly meet the transportation needs of the population at a specific moment, but might be completely inadequate at another time. On the one hand a traditional frequent service would be extremely expensive. On the other hand, reducing service frequency would make the service unattractive. For this reasons, mass transit services evolved towards some degree of flexibility.

Demand Responsive Systems are a family of mass transportation services which, as the name suggests, are *responsive* to the actual demand for transportation in a specific time period. Such responsiveness evolves towards a personalization of the services: itineraries, schedules and stop locations are variable and determined according to the needs for transportation as they change in time. Demand Responsive Systems were introduced under the name *Dial-a-Ride* (DAR). The first implementations of DAR consisted in door-to-door services for users with particular needs or reduced mobility, such as handicapped and elderly people (see Wilson et al. (1971) and for a recent survey Cordeau and Laporte (2003)). The *flexibility* of DAR systems allows to respond to the fluctuation of demand and provides the means to offer personalized services, while still maintaining a certain degree of resource sharing. This has led certain transportation or city authorities to

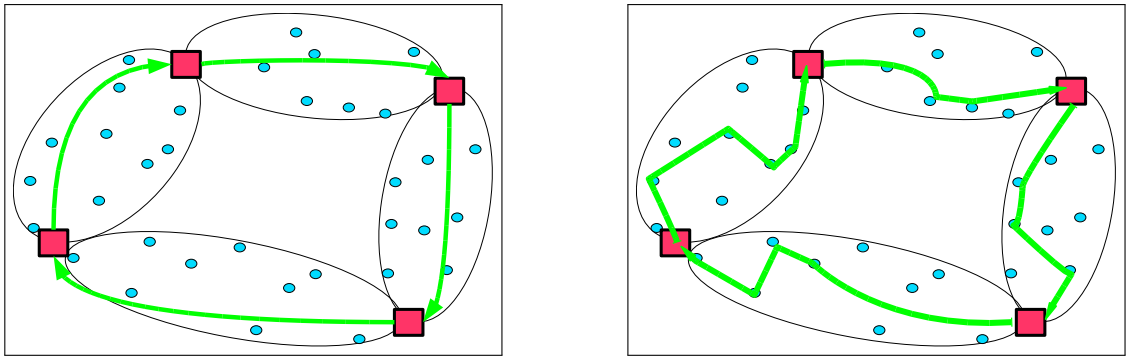
extend DAR services to more general transportation settings.

As mentioned earlier, DAR systems display a number of drawbacks, some of which follow from the extreme flexibility inherent with the system definition. Consequently, practitioners started to experiment and implement new transportation paradigms in order to reintroduce a certain structure in the service. In particular, the combination between the regular traditional transit systems with pure on demand services. Such systems are commonly denoted *semi-flexible* systems. Koffman (2004) and more recently Potts et al. (2010) report practical experiences with semi-flexible systems undertaken in North America and testifies to the importance of semi-flexible systems and their potentially very large impact in terms of cost reductions and quality of service improvement in weak-demand scenarios. These systems, while differing in terms of organization, fleet management, policies, and so on, they all present a number of basic common features. Similarly to traditional transit, they have a set of stops with a fixed predetermined timetable. Moreover, part of the service is flexible and users can ask for service at optional locations. The fundamental idea behind semi-flexible systems is that the regularity of the service is by itself a valuable property of a transportation system because, for example, it helps users in planning their trips, facilitates integration with other transportation modes, and makes it possible to access the service without booking. At the same time, such systems try to inject flexibility by considering some slack time that can be used to possibly deviate from a basic path to operate in a demand-responsive framework. Errico et al. (2011b) show that DAS (Malucelli et al. 1999) is a model sufficiently general to suitably represent all the variants of semi-flexible systems described in Koffman (2004) and Potts et al. (2010). Consequently, most of the considerations and developments done for DAS also apply to other more specific semi-flexible systems.

2.2 Demand Adaptive Systems

In its most general form, a DAS is made up of several lines and is interconnected with the traditional transit system. Several vehicles operate on each DAS line providing service among a sequence of *compulsory* stops. Each compulsory stop is served within a predefined *time window*. The collection of time windows corresponding to the compulsory stops, including the start and end of the line, forms what we call the *master schedule* of the DAS line. This defines the traditional part of a DAS. Additional flexibility is provided by allowing customers to request service from and to *optional* stops within a given area. Such stops are visited only if a request is issued and accepted when operating the service. We denote users requesting service at an optional stop as *active users*, while users moving only between compulsory stops are called *passive users*.

To serve optional stops, the vehicle must generally deviate from the shortest path joining two successive compulsory stops. The region, and consequently the set of optional stops, that is possible to visit between two consecutive compulsory is defined in advance



(a) DAS serving only compulsory stops and passive users only

(b) DAS serving compulsory and optional stops, i.e. passive and active users

Figure 1: A Single-Line DAS

and it is called *segment*. Figure 1a depicts the basic DAS service visiting only compulsory stops, while Figure 1b illustrates the same DAS line when user requests for optional stops are present.

The combination of compulsory stops and time windows is one of the main feature of DAS and is the "backbone" providing a regular service for stops that are particularly demand attractive and simple to access. In particular, the regularity of service encourages some customers to use the service in passive mode, indeed users that for some reason did not book the service, can still access the service by walking to the nearest compulsory stop. If compulsory stops coincide with stops of a the regular transit line or other DAS lines and timetables are synchronized, users can transfer from one line to another. This way the attractiveness in terms of coverage and flexibility can drastically increase. In this context, time windows play an important role because they determine the possible synchronization among lines. The time windows in the master schedule also influence the flexibility the service may provide for user requests at optional stops. Notice finally, that the time windows and the segment specification provide an *a priori* guarantee for the longest user travel time. The detours induced by the activation of optional stops must be such that the time windows at compulsory stops are met.

From the operational point of view, the time window associated to a compulsory stop defines the earliest and the latest vehicle departure time (EDT and LDT, respectively) for that compulsory stop. In practice, the vehicle is allowed to arrive at any time before the LDT. But if it arrives at the compulsory stop before the EDT, it will have to wait, experiencing what we call *idle* time periods. Because the vehicle might leave at any time after the EDT, passive users need be present at the desired compulsory stop not later than the EDT. As a consequence, if the vehicle arrives at the compulsory stop later then the EDT, passive users will experience what we call *passive* waiting times.

The DAS model is actually more general than what is usually intended as semi-flexible system. In fact it displays a more complex schedule mechanism. Semi-flexible systems usually associate to compulsory stops a one-point-in-time schedule. Flexibility is obtained by allocating some additional (usually denoted *slack*) time to the shortest time needed to reach two consecutive compulsory stops. From the operations point of view, if vehicles arrive earlier than the scheduled time, they must wait and idle times will be experienced. As a consequence there is no possibility to transfer the available flexibility among consecutive segments. In DAS this is not necessarily true. In fact, as described earlier, if vehicles arrive within the time window, they are allowed to leave at any time, gaining in this way time that can be possibly spent in the following segment. Observe finally that by fixing all the time windows to zero-width, the DAS schedule reduces to the schedule type commonly used in semi-flexible systems.

2.3 Planning Issues

Transportation systems dedicated to serve several demands with the same vehicle generally require complex planning activities. For traditional transit, the design of the system in terms of line routes is determined during the so-called strategic planning phase, timetables and vehicle schedules and routes are part of the tactical planning phase, and crew schedules are built during operational planning (Ceder and Wilson 1997). Comparatively, purely on-demand services such as DAR, need little strategic design, mainly to define service areas and the composition of the fleet (see, e.g., Diana et al. 2006; Quadrifoglio et al. 2008). The most important planning process for DAR is at the operational level, however, when routes and schedules are determined little time before each departure and are possibly dynamically modified once service has begun.

DAS, inheriting features of both traditional transit and DAR, requires a careful planning phase. Errico et al. (2011b) reported a classification of the planning decisions in the traditional strategic, tactical and operational hierarchy, together with a detailed review of the related literature. Summarizing, at the strategic level, the region to be served is identified and partitioned into subregions, each corresponding to the area a single DAS line will service. The desired quality of service and frequency of service is established for each service area. An initial set of compulsory stops (possibly empty) to be used as transfer points is also determined. At the tactical level, the DAS line is built for every subregion identified at the strategic level, this process including decisions about the location of additional compulsory stops and their sequence, time windows, and segments. This *backbone* of the line plays the same role for the transit authority and the users of a DAS as the schedule in traditional transit systems. It defines only partial itineraries and schedules, however. The backbone may be completed at operational level, when the actual requests for transportation become known. Thus, for each departure time, the actual itinerary and schedule is built to incorporate the additional optional stops corresponding to the accepted active-user requests, while respecting the constraints imposed

by the line backbone.

3 The SDDP and related literature

In the present section we introduce and formalize the SDDP (Section 3.1) and briefly review the related literature (Section 3.2).

3.1 Definition of the SDDP

When considering the SDDP, we suppose that all decisions belonging to the strategic planning level have been already taken. We assume, in particular, that the area to be serviced by the single DAS line has been selected and that frequencies of service have been established according to vehicle capacities, transportation demand in the area, and target level of service. In the tactical planning the horizon is typically seasonal. The input coming from the strategic planning is the service area, a probabilistic knowledge of the demand in the planning horizon and the desired frequency of service. Regarding the service area, recall that it is common in public transit to represent the regions of interest partitioned into several smaller zones, each characterized by approximately homogeneous features, and to identify such a zone with a single point that we call *demand point*. Attributes of a specific zone, such as demand, population density, etc, are then formally referred to the demand point. Depending on the particular planning needs, the representation of the service area may be more or less refined, according to the number of demand points and level of data aggregation. For the scope of the SDDP, we assume that the service area is represented by a set of demand points and that we are given the traveling times between any pair of demand points. Observe that we do not make any particular assumption on the shape of the service area, nor on the level of data aggregation or the road network.

Demand for transportation is usually modeled as trip volumes associated to each possible pair of demand points, called *origin-destination* pair (O/D). In our specific context, decisions have to be taken when the actual requests are not known yet. For this, we assume that a certain knowledge of the demand is available (e.g., from history or surveys) so that, together with the service frequencies, it is possible, for a given period in the time horizon and for all possible O/D, to extract significant demand information such as probability distributions, expected demand volumes, etc.

The objectives associated to the SDDP take into account both efficiency of the system and level of service offered to users. As an indirect measure of the operational costs, we adopt the widely accepted practice of considering those as strongly related to the vehicle

traveling times. On the other side, we measure the quality of the service offered to the users as the time spent by users on the vehicle. In the following, we use the term *Latency* to refer to the latter measure. Observe that such objectives are conflicting as to a higher degree of efficiency usually corresponds worse latency values.

Given the above assumptions, input data and objectives, the SDDP is the problem requiring to take, for every occurrence of the line in a certain planning horizon, the following decisions: the choice of the sites where compulsory stops may be located, the definition of the main structure of the line by sequencing the compulsory stops, the partition of the service area into segments, and the definition of time windows at compulsory stops. As detailed in Section 4, the SDDP is an extremely complex problem, encompassing several difficult sub-problems and characterized by demand that may fluctuate significantly during the planning horizon.

3.2 Related Literature

Errico et al. (2011b) report a comprehensive literature review on planning DAS and we refer to that work for details on the topic. In this section, we restrict our review to contributions focusing on tactical planning aspects.

As mentioned earlier, relations among service area and the amount of slack time to allow deviations is a typical tactical planning issue and it is the most studied in literature. In particular Smith et al. (2003), based in turn on a more practical work by Durvasula et al. (1998), considers two fixed lines and applies the method described in Welch et al. (1991) to obtain a DAS line with a candidate service area. The service area is defined as the maximum allowable deviation from the shortest path joining consecutive compulsory stops. The authors consider, as in Durvasula et al. (1998), three possible values of deviation not necessarily equal for all segments and two possible slack time policies (all the segments must have the same policy). The authors build a multi-objective nonlinear choice model where the contrasting objectives are the maximization of the feasible deviations and the minimization of slack time. The problem is solved by a gradient method where the evaluation of the objective function at each iteration is performed by a heuristic GIS based tool described in Durvasula et al. (1998). Given the very few variables and possible considered values, the authors are able to solve the model but computing times or efficiency study are not reported.

Several works consider a simplified operational framework with the aim of providing closed-form analytic relations between the main design parameters of the DAS. Although the notation used in those works is not uniform, the assumptions are mostly the same. The service area is represented as a rectangle with length considerably higher than width. The service is performed along the horizontal direction (the length) and compulsory stops are located in the middle of the two vertical edges of the rectangle. The demand is

modeled as a set of locations, either pickups or drop-offs, uniformly and continuously distributed on the service area, with a specific per unit density. The vehicle is considered to have constant speed, and move on an infinitely dense grid road network according to linear paths parallel or orthogonal to the edges of the service area.

Fu (2002) addresses the problem of defining the optimal amount of slack time needed for the service of a single DAS segment while optimizing an objective function made up of three components: the operator cost, the service benefit, and user costs. The resulting model is a linear program in one variable with a feasible region bounded by three constraints, which can be trivially solved analytically. The author completed the study by simulating the operations by a tool called SimParatransit (Fu 2001) originally devised for simulation of paratransit operation and adapted for the scope. The simulation considers an operational framework very close to that used in the analytical model, and focused on estimating the effects of slack time changes on idle times and number of feasible deviations. Though the model catches some general tendencies, it substantially fails in capturing the details of the system behavior.

Quadrifoglio et al. (2006) also considers a setting similar to the previous one, where additionally vehicles have infinite capacity. The scope of the paper is to derive upper and lower bounds on the expected vehicle speed along the main direction (the length). Assuming the so called no-backtracking policy introduced in Daganzo (1984) (the vehicle is not allowed to move backward with respect to the main direction), a lower bound on the expected longitudinal velocity is obtained, then it is shown that, under certain conditions, the no-backtracking policy is optimal. This fact is used to derive an upper bound on the expected longitudinal velocity by considering subsets of requested points satisfying such conditions and characterized by the fact that the Hamiltonian path through such subset is certainly shorter than the one on all the requested stops. To compute expected values, the authors need to compute the probability distribution of the number of points belonging to such subsets and, as this computation is very complex, it is approximated. A second upper bound is obtained by considering the total travel time as the summation of the time to travel from each requested stop to its closest neighbor. This is equivalent to considering a relaxation of the Hamiltonian path and thus provides an upper bound of the longitudinal velocity. The authors consider also a simulation study where the operational policy adopted is an insertion heuristic algorithm as described in Quadrifoglio et al. (2007) and results are compared with the approximated upper and lower bounds. The authors claim that, with some exceptions, the approximated values fit the simulated ones sufficiently well. Finally, based on the previous results, the authors give an estimated relation between longitudinal velocity and service *capacity*, defined as number of optional location the vehicle is able to service in a given time.

Similar aspects, but different methods, are investigated in Zhao and Dessouky (2008) where the system capacity is treated. The authors analyze the relationship between service cycle time, and the length and width of the service area. They consider the same

simplified setting as in Quadrifoglio et al. (2006). Considering the programmed time duration of the DAS line, T, and the distribution probability of the actual arrival time T_R, and under the assumption that $E(T_R) < T$, the authors call on queueing-theory results and derive a Wiener-Hopr integral equation to represent the delay distribution. From this, and by approximating the travel-time distribution with an exponential distribution, they derive approximated analytical relations among the length and the width of the square and T. Finally the authors test the correctness of their analytical model by simulating a non back-tracking nearest insertion operational strategy. The authors claim that the experimental results obtained are in line with the analytical approximations derived.

According to the planning classification given earlier, all the above works address tactical issues. However, due to the simplifying hypothesis, they appear more as fast and approximated evaluation tools of tactical plans suitable to be used at the strategic level, rather than tools to build the design of DAS lines when considering the complexity of actual applications, where the demand is not uniformly and continuously distributed on the service area, the service areas do not have nice geometric shapes, and vehicles do not move along infinitely dense grid road networks.

The main motivation of the present work (and of Errico (2008)), is in fact to fill this gap and to provide a general methodology and algorithmic framework based on more realistic assumptions and suitable to build the design of DAS lines in general, real-life environments.

4 A framework to address the SDDP

The SDDP is very complex because it integrates several classes of interdependent decisions that are very difficult even when addressed separately, such as, Traveling Salesman Problem-like (sequencing of compulsory stops), location (position of compulsory stops), timing (definition of time windows), and partition (definition of segments). We propose a decomposition strategy in the decision space, creating thus subproblems where only part of the decisions have to be considered simultaneously, and in the time space, so that demand fluctuations can be ignored.

Due to the particular structure of the problem, the latter decomposition is not a critical issue. In fact, it is usually possible to identify in the planning horizon several *homogeneous* time periods, i.e. periods where demand variations are sufficiently small to be ignored and demand to be assumed constant. Thus, given the targeted level of service in terms of departure frequency, homogeneous periods can be considered as being made up of several identical and consecutive time *slices*, defined a the time length corresponding to two consecutive occurrence of the line. Observe that the *S-SDDP*, defined as the SDDP

over a single time slice, is a time-independent problem. Once it is possible to address the S-SDDP, a solution for the SDDP on a given time horizon can be easily built, as shown in Section 6.

The most challenging aspect when addressing the SDDP is the combination of decisions, that is, the S-SDDP and we focus the next subsection on it. In particular, we propose in Section 4.1 two hierarchical decomposition approaches for the S-SDDP, which generate a number of *core* subroblems. We briefly recall the core problems in Sections 4.2 and 4.3, together with the methodology proposed to address them. The parameters of these problems and algorithms are important because by influencing the solutions obtained for the particular problems, they influence the global S-SDDP design.

4.1 Addressing the S-SDDP

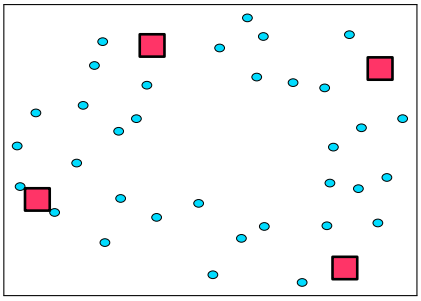
We propose to decompose the S-SDDP into simpler, self-contained hierarchical subproblems. The possible decompositions strategies vary depending on the order of the chain of decisions and the level of aggregation among decisions. A broader discussion on possible strategies can be found in Errico (2008). In this paper, we propose two strategies called *Sequence Compulsory* (SC) and *Sequence All* (SA), schematically represented in Figure 2.

The SC strategy first selects the compulsory stops, and sequences them. It then defines the segments by a geometric closeness criterion: a demand point is assigned to the closest pair of compulsory stops. Finally, SC determines the set of time windows.

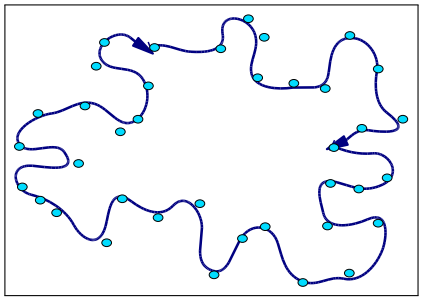
In the SA strategy, a sequence of all stops is determined first. The compulsory stops are determined in a second step. The sequence determined at the first step, together with the definition of the compulsory stops, induces their sequence and is used to implicitly determine the segments. Finally, similarly to the SC strategy, time windows are determined.

The two hierarchical decompositions give rise to a number of *core* problems, namely, the selection of compulsory stops, the sequencing of compulsory stops (for SC) or all demand points (in SA), and the definition of time windows. The process is depicted in Figure 3, where both decompositions can be encompassed. The leftmost block is composed of two sub-blocks. Different interactions among these sub-blocks correspond to SA or SC strategy. In both cases, the output of this design phase is given by the set of compulsory stops, their sequence, and the segment definition. We call this set of decisions the *topological* DAS line design. The last step of the design is the definition of the time windows (*Master Schedule*) and it is common to both decompositions.

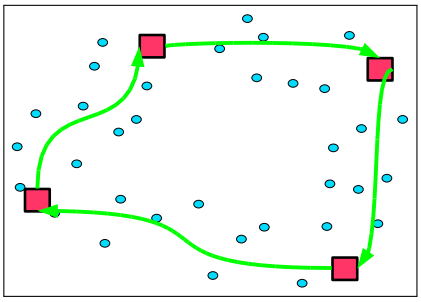
Regarding the selection of compulsory stops, the main idea is to select them in lo-



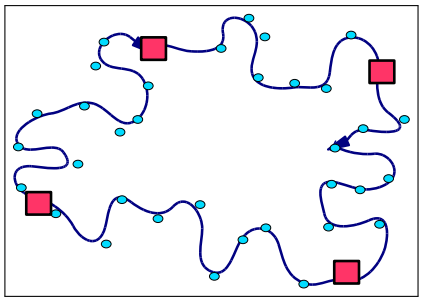
(a) SC: Compulsory stops are selected



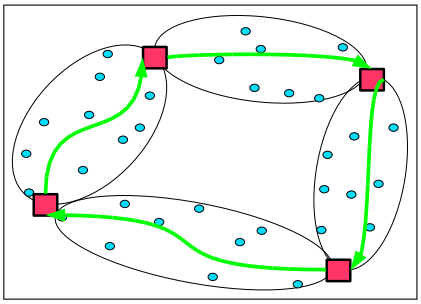
(b) SA: A sequence among all the stops is found



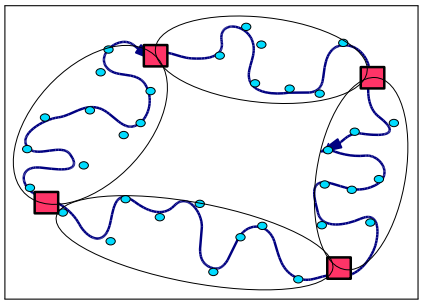
(c) SC: A sequence among compulsory stops is found



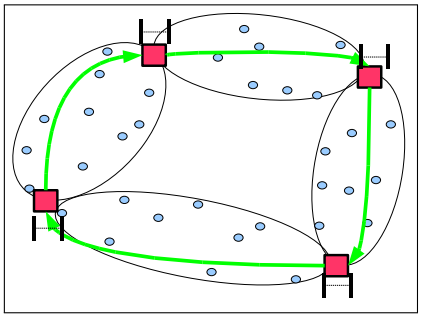
(d) SA: Compulsory stops are selected



(e) SC: Segments are defined by geometric closeness criterion



(f) SA: Segments are deduced by relaxing the original sequence



(g) SC and SA: Time windows are defined

Figure 2: Comparison between the two hierarchical decomposition strategies: SC and SA

cations where demand is naturally high, such as hospitals, schools, libraries, railway stations, etc. From a mathematical point of view, this may be quantified through the probability of a given location to be requested for service. An easy, and in our opinion reasonable, way to choose the compulsory stops is then to select the locations with probability to be requested for service "close" to 1. As a consequence, the problem is easily solved by deducing such probabilities from the data. Finally, observe that this approach can easily integrate compulsory stops established with other methods, e.g., the ones selected as transfer points at the strategic level, or the ones needed because of other possible managerial or political reasons. Addressing the remaining core problem is more involved. We detail them in the next two sections.

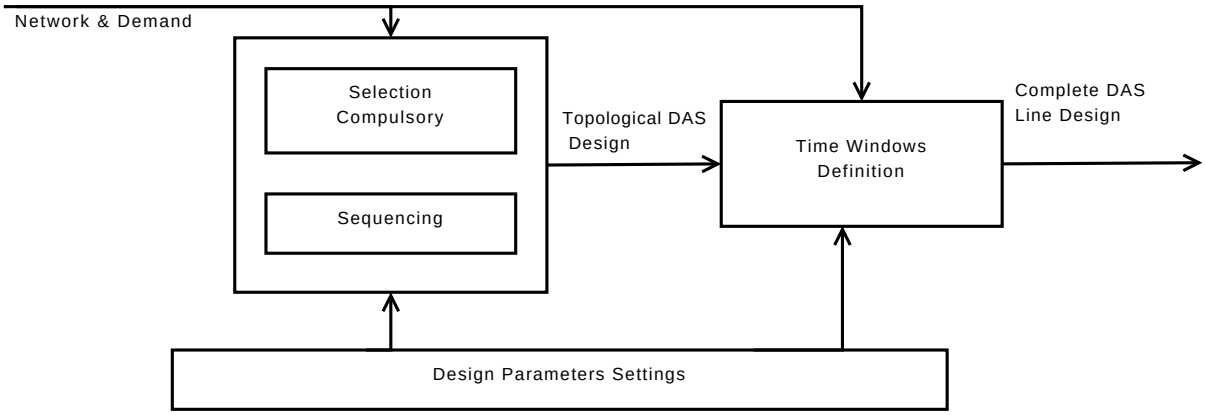


Figure 3: Structure of the Design Process

4.2 Sequencing: The General Minimum Latency Problem

Both hierarchical decompositions present a sequencing step. In the case of SA, the items that must be sequenced are all the potential demand points in the service area while, in the case of SC, the compulsory stops that have been chosen at the previous step of the decomposition. The sequences have the objective of inducing a design of the line that is at the same time economically efficient and suitable to offer a good level of service to the users (as mentioned before, in the sense of small latency values). Observe that, except for the items to sequence, the underlying problem is the same for both SA and SC. We modeled it as the *General Minimum Latency Problem* (GMLP) and addressed it in Errico et al. (2011a), where details on definition, formulation, polyhedral properties, and solution methods can be found.

Formally, the GMLP can be described as follows. Consider a complete and undirected graph with non negative costs. The nodes of the graph represent the compulsory stops or all possible demand points, according to the context. The cost associated to an edge corresponds to the estimated operational cost of establishing service between the demand

points represented by the nodes connected by the edge. Demand for transportation is represented by an O/D matrix where each entry specifies the number of people to be transported between the corresponding origin and destination points. Similarly to the Traveling Salesman Problem (TSP), the scope of the GMLP is to find a Hamiltonian circuit among all the demand points. Differently from the TSP, the objective function of the GMLP has two components: the costs of establishing the service between two stops on the one hand, the average time the users spend in the vehicle on the other hand.

The main design parameter related to the GMLP is the relative weight of the two terms of the objective function. We express the objective function in the following form:

$$\min \quad \{ \ (1-\alpha) \text{ Operational-Costs} + \alpha \text{ Latency} \ \} \tag{1}$$

Consequently, when $\alpha < 0.5$, the optimization process emphasizes efficiency aspects and the resulting sequences are expected to be less expensive. On the contrary, when $\alpha > 0.5$, the optimization process emphasizes the user level of service and the resulting sequences are expected to be longer, but the average user travel times to be shorter.

Observe that, in actual practice, most transit lines are operated in the two opposite directions. To represent this situation, once the service is established between two stops, the demand flows can move in either direction. The assumption we make is that the demand will always distribute along the shortest portion of the cycle (representing the fact that users will always choose the fastest way to reach their destination). The two components of the objective function make the GMLP much more difficult to address then the TSP. We developed a Branch and Cut approach for the GMLP based on the use of Benders reformulation and a set of valid inequalities inspired by the TSP literature.

Notice that it is quite straightforward to represent the case where lines are operated in one direction only. Instead of an undirected graph, it is sufficient to consider a directed graph. Once the service from one node to the other is established, demand movements must follow the established direction. From the solution point of view, this fact has no impact except for a few minor changes in the separation of valid inequalities. We therefore do not address this case in any more detail.

4.3 Time Windows: The Master Schedule Problem

In order to introduce the *Master Schedule Problem* (MSP) and related design parameter, we need to define two main related entities: the *width* of a time window and the *distance* between two time windows. Recalling the terminology introduced in Section 2.2, the width is defined as the difference between the LDT and EDT. The distance between two time windows is defined as the difference in time between their *centers* (defined as EDT+ (LDT-EDT)/2). The distance between two consecutive time windows is proportional to

the flexibility available to service the segment included between them, while the time-window width is related to how much of this flexibility can be transferred from one segment to the following one.

When designing the master schedule, one should take into account the following conflicting factors. On the one hand, allocating long distances between consecutive time windows, allows to service more optional stops if needed, but it also increases the possibility to experience idle times. On the other hand, designing wider time windows allows to transfer the flexibility between adjacent segments, but the passive times might considerably increase.

The MSP has been addressed in Crainic et al. (2010), where details on definition, formulation and solution methods can be found. Summarizing, the MSP considers two kind of inputs. On the one hand, the *topological* design, i.e., the location and sequence of the compulsory stops, and the partition of the service area into segments. On the other hand, for each demand point i in the service area, a probability π_i to be requested for service is supposed to be known. The purpose of the MSP is to define a time window for each compulsory stop such that the LDT of the last compulsory stop is minimized, and the probability to be able to serve *all* the requests at operations time is given by $\epsilon \approx 1$. In order to account for a form of fairness in the service provided to passive users, we consider time windows having the same width δ. Consequently, the parameters regulating the MSP are ϵ and δ.

From the solution method point of view, the most challenging part is the estimation of the probability distribution of the time needed to travel a given segment. We proposed (Crainic et al. 2010) a very efficient sampling mechanism and also showed how to compute the probability distribution of the arrival time at the last compulsory stop and, then, suitable time windows.

5 Computational experience

The main purpose of the experimental study is to 1) compare the design approaches described in Section 4.1 (SC and SA), and 2) evaluate the impact of parameter changes on the design process.

The experimentation was carried out according to the following methodology. We generated several scenarios by varying the demand in a given service area. We applied the SC and SA solution approaches for each input scenario to obtain two single-line DAS designs. Different designs were then produced for several design parameter settings. For each parameter setting, design type (SC or SA), and input scenario, we simulated the operations of the system by generating a set of requests for transportation according to

the input scenario. Scenarios and designs are compared according to several performance measures such as idle time, passive times, maximum riding time, latency, etc.

In Section 5.1 we give the details and settings of the experimental study, while in Section 5.2 we present and analyze the results.

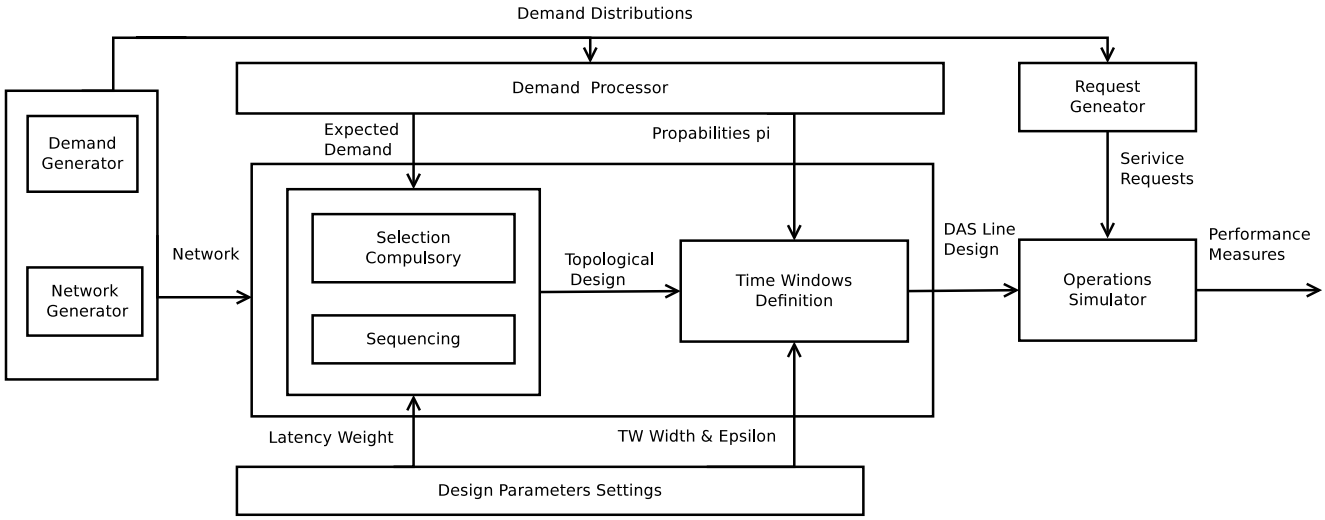


Figure 4: Structure of the experimental method

5.1 Experimentation Setting

Figure 4 represents the general scheme of the experimental method. The leftmost block represents the scenario generator. It consists of two components, one generating the service area and the locations of the demand points, the other generating demand probability distribution matrices representing several possible demand scenarios.

The service area we consider is a square where 40 demand points are located according to the uniform distribution. In order to set up realistic costs and distances, we scaled the edge of the square and speed values such that the duration of the tour along the whole set of points is around 6500 seconds. Such a value is important because it represents the time needed by a hypothetic traditional transit line to service the whole set of demand points (we indicate such a value as Traditional Service Time, TST). Consequently it represents an upper bound of the *service* time of the DAS line (i.e., the time the vehicle needs to service the line, including running times, idle times, time to serve customers) and can be used to measure how much it is gained in service time when operating a DAS line instead of a traditional line.

Demand probability distributions are constant in the considered time slice. As previously mentioned, in real-life contexts a small percentage of the demand points is more

attractive (more requested) than others. To represent this fact, demand distributions are such that exactly 4 attractive demand points are generated. We then considered two possible demand scenarios, one with relatively low demand where the average value of π_i (probability at least one requests is issued for the optional demand points i, see Section 4.3) is around 30%, and another scenario where the line is almost saturated with probabilities π_i around 70%.

The block in the center of Figure 4 represents the design process as described earlier. The block on the top of the design meta-block, named *Demand Processor* is in charge of transforming the initial demand distribution in expected demand matrices for the sequencing problem and in probabilities π_i for the selection of compulsory stops and the MSP. The block under the design block represents the parameter settings used in the experimentation, namely the relative weight α of the latency vs. operational costs in the GMLP, the time-window width δ, and the probability the system is able to serve every request ϵ in the MSP. We considered the following values: $\alpha = \{0.1, 0.9\}$ to evaluate the effects of opposite emphasis in the objective function of the sequence problem, $\delta = \{300, 500\}$ seconds to evaluate the effects of the time-window width on idle and waiting times, and $\epsilon = \{0.85, 0.95\}$ to evaluate different policies regarding the reliability of the system (the higher the ϵ, the more reliable the system, as the probability to serve all the requests becomes higher).

The rightmost block in Figure 4 represents the simulator of the operations. This simulator operates according to the DAS1 policy (Malucelli et al. 1999) which can be described as follows: Requests for transportation might be *rejected* if their acceptance causes infeasibility with respect to the master schedule. If a request is accepted, users are picked up and dropped off exactly at the location they asked for. The solution methods used to solve the operational problem are reported in Malucelli et al. (2001) and Crainic et al. (2005). The request generator, depicted on the top of this block, is in charge of generating transportation requests between origins and destinations according to the considered demand distribution. For each topological design type, parameter setting, and input scenario, we simulated operations over 100 instances with about 25 requests each in average.

We considered several measures when evaluating the performance of a particular line design: The LDT at the last compulsory stop of the line, the per-user latency, average per-segment idle times, average per-segment waiting times, percentage of rejected requests, maximum occupancy of the vehicle, defined as the maximum number of people on the vehicle at the same time. The LDT at the last compulsory stop is important because it represents the Upper Bound on the Service Time of the line (UST). The UST is not only important by itself, providing an indirect measure of the operating costs of the line, but also because it quantifies, when compared with the TST, how much it is gained in terms of service time compared with a traditional service. Consequently, in our tables, we usually report the UST and two related values, the Lower bound on the Service Time

Improvement (LSTI = TST - UST), and the Actual Service Time Improvement (ASTI = TST-AST), where AST is the Actual Service Time (the time the vehicle actually arrives at the last compulsory stop). Regarding the maximum occupancy of the vehicle, we observe that the tactical design, and so the SDDP, does not take into account capacity explicitly because this issue, related with the frequency and quality level of the service, belongs to the strategic planning. It can be interesting, however, to monitor the maximum occupancy as a way to verify that the considered time-slice intervals are reasonable and, consequently, the demand aggregation is sound.

For each input scenario and design parameter change, we analyze the resulting effects in terms of the performance measures of the system and normally we report results averaging on the number of instances. To avoid redundancies and for ease of presentation, when the variation of a certain input scenario or design parameter value has different effects on the two design approaches, we underline this fact and we quantify it in tables and figures. On the contrary, when the effects are similar for the two approaches, we only report results for one of them. The experimentation is done by first defining a *basic* parameter setting and then comparing it to several alternative parameter settings differing from the basic one in the value of exactly one parameter. The parameter values in the basic setting are $\alpha = 0.1$, $\epsilon = 0.95$, $\delta = 500$.

5.2 Results

In the present section we present and analyze the results. In particular, Section 5.2.1 analyzes the effects of α, Section 5.2.2 the effects of ϵ and δ, Section 5.2.3 the effects of demand fluctuation, and in Section 5.2.4 we propose some concluding remarks.

5.2.1 Effects of the sequencing parameter, α

By varying parameter α, we expect to obtain different sequences and segments, since small α values (i.e, $\alpha = 0.1$) give more emphasis to the efficiency of operations, resulting in short itineraries, while values of α close to 1 (i.e., $\alpha = 0.9$) give more emphasis to latency, thus producing possibly longer itineraries but with overall shorter travel times for passengers. However, due to the fact stops are considered differently, SA and SC react quite differently to changes to α. Figure 5 represent the DAS lines output by SA when $\alpha = 0.1$ (on the left), and when $\alpha = 0.9$ (on the right). Numbers close to compulsory stops give the sequencing and dots with the same greyscale define the stops belonging to the same segment.

The compulsory stops are the same in both cases since their selection is independent from the value of α. We observe, however, a change in the sequence of compulsory stops

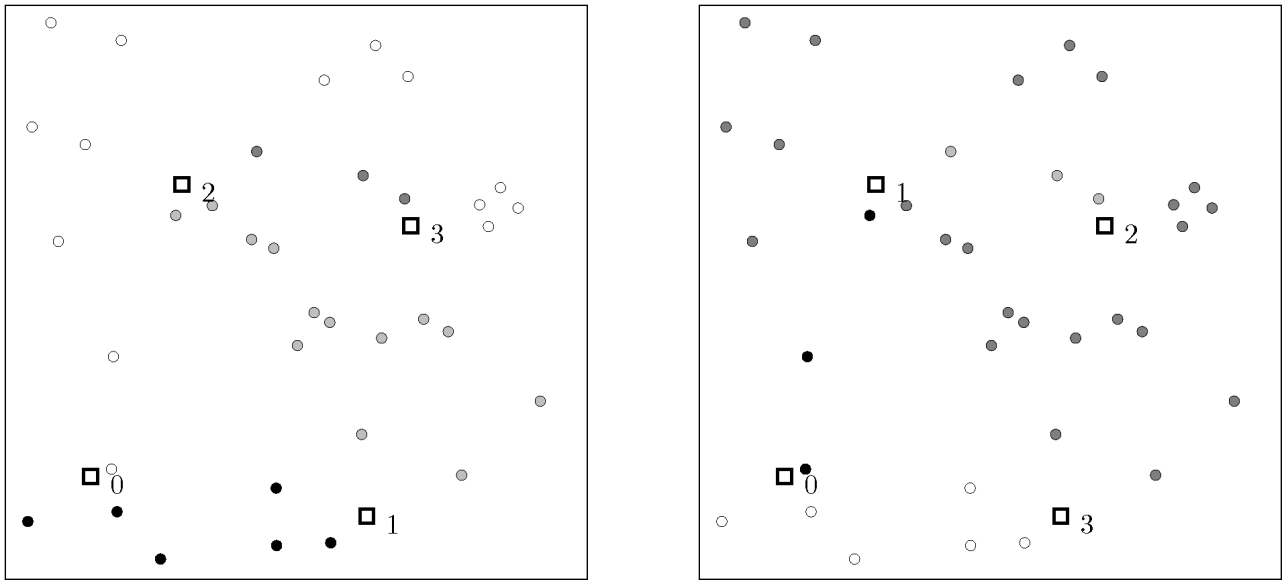


Figure 5: SA Design approach. Effects of α

and the partition of demand points into segments. As expected, $\alpha = 0.1$ with emphasis on the efficiency of operations, results in a more linear sequence on the entire set of stops, while $\alpha = 0.9$ with emphasis on the latency, yields a more convoluted sequence. This has a direct consequence on the segment shape which, for high α, have a more complex configuration than for low α and this is particularly evident for the segment whose stops are colored in dark gray. Our intuition about the shape of the sequence is confirmed as the line output by SA when α is large is more convoluted. However notice that, considering compulsory stops only, when α is large the sequence is not necessarily more involved, as in our example, though the shape of the segments is clearly more tortuous.

A similar comparison is illustrated in Figure 6 for the SC approach where the sequencing decision involve only compulsory stops. In this case the relative small number of compulsory stops does not allow the α to have an influence, indeed in our example the two settings of α yield the same solution.

We used the previous topological design for a numerical study and, by considering demand distributions with average $\pi_i \approx 30\%$, built the set of time windows by solving the MLP with design parameters $\delta = 500$ and $\epsilon = 0.95$. We simulated the operations over 100 scenarios of requests for each value of α and for each direction. The values of the performance measures, averaged on the set of request scenarios, are reported in Table 1. The first column reports the design approach used, the second the value of α. The third, and forth, and fifth columns report the UST, the percentage of the relative LSTI

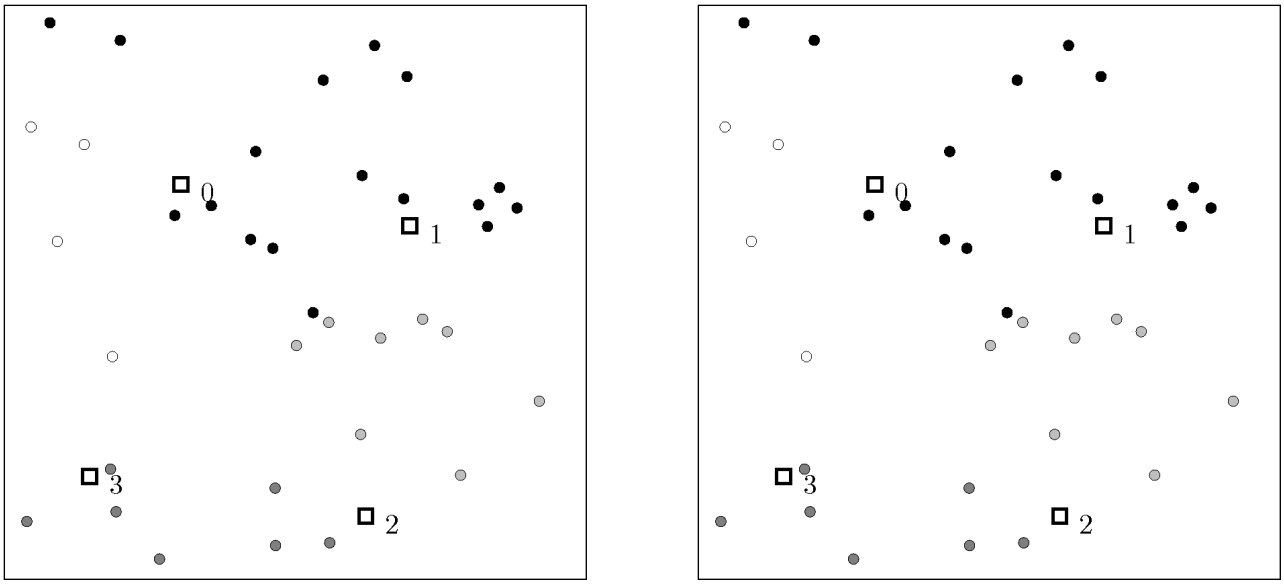


Figure 6: SC Design approach. Effects of α

Approach	α	UST	%LSTI	%ASTI	Latency	IdleT	PassiveT	%Rej/Total	maxCap
SA	0.1	5095.0	21.6	27.2	1396.86	49.03	158.98	0.40	11.23
SA	0.9	5310.0	18.3	23.7	1295.20	69.68	229.27	0.54	12.27
SC	0.1	5230.0	19.5	26.0	1361.63	164.24	103.46	0.42	11.23
SC	0.9	5230.0	19.5	26.0	1360.79	165.30	100.78	0.44	11.23

Table 1: Different design approaches. Values: $\pi_i \approx 30\%$, $\epsilon = 0.95$, TW=500

and ASTI, respectively. The sixth column reports the value of the average user time on the vehicle (latency). The seventh and eighth columns report the per-segment idle time and waiting time, respectively, while the last two columns report the average ratio of rejected requests over total number of requests and the average occupancy of the vehicle, respectively.

Focusing first on SA, we observe that, as expected, the time needed on average to operate the line for $\alpha = 0.1$ is lower then for $\alpha = 0.9$, and this is confirmed by the variations of UST, LSTI, and ASTI values. A worse UST when $\alpha = 0.9$ is balanced in terms of latency, which is better than for the $\alpha = 0.1$ case. In fact, results confirm that UST is better for low α even when the resulting sequence of compulsory stops is more convoluted then for high α. A similar, but reversed, consideration also holds for latency values. We finally notice that for $\alpha = 0.9$ the idle and passive times are slightly worse and we observe that this is related to the higher value of the UST. For the SC case, the simulation confirms that parameter α does not have much influence in this case.

ϵ	UST	%LSTI	%ASTI	Latency	IdleT	PassiveT	%Rej/Total	maxCap
0.95	5230.0	19.5	26.0	1361.63	164.24	103.46	0.42	11.23
0.85	4812.5	26.0	31.7	1269.29	94.99	150.22	1.27	11.11

Table 2: Master Schedule (ϵ). Settings: $\pi_i \approx 30\%$, $\alpha = 0.1$, $\delta = 500$, DesignType=SC

It is interesting to compare the two methods. The first evident difference is that SA is much more sensitive to variations of α then SC. Another difference is that the configuration of the segments for the SA case is more complex than for the SC case. Notice that, despite this complexity, the best value of the latency is obtained with SA and $\alpha = 0.9$ and that the best value for UST is obtained again with SA but with $\alpha = 0.1$. However, to the best value of the latency also corresponds the worse value of UST and, analogously, to the best values of UST corresponds also the worst value of latency. This suggests that, according to the particular applications, for SA, the values of α can be modulated to obtain different efficiency/latency tradeoffs.

For general comments on DAS behavior, the values of LSTI are quite interesting as they prove, in our opinion, the potential for cost reductions of DAS in the presence of relatively low demand ($\pi \approx 30\%$ in this case). Observe that, to a consistent reduction of the running times with respect to a traditional line (LSTI around 20% for both approaches and both values of α), corresponds a very low percentage of rejected requests (less than 0.6% of the total demand in all cases). This is even more evident when we compare LSTI with ASTI. Actually, the running time is considerably shorter than the upper bound and this might imply advantages both on the vehicle availability and the operation costs (e.g., fuel consumption).

5.2.2 Effects of the Master Schedule Parameters, ϵ and δ

Parameters ϵ and δ concern the definition of the time windows in the MSP. To test the effects of ϵ on the system behavior, we performed simulations on two designs of a DAS line obtained from two parameter settings differing only in the value of ϵ. The rest of the parameter setting was $\alpha = 0.1$ and $\delta = 500$, and the demand scenario considered was such that $\pi_i = 30\%$. As in the previous case, the simulation was performed on 100 request scenarios for each direction and generated accordingly to the demand scenario. In Table 2 we report the results averaged over the set of requests scenarios and the meaning of the column is the same of the previous table. We did not observe significant differences in how SA or SC lines responded to the parameter variation, we report only the values for the SC approach. We observe that, as expected, for $\epsilon = 0.95$ the percentage of rejected requests is less than for $\epsilon = 0.85$, but this is paid in terms of UST, LSTI, ASTI and per user latency. It is interesting to observe that while the idle times increase, the passive times decrease for higher setting of ϵ. To understand this, remember that the closer to 1 is the value of ϵ, the higher is the probability the system will be able to serve all the

δ	UST	%LSTI	%ASTI	Latency	IdleT	PassiveT	%Rej/Total	maxCap
500	5230.0	19.5	26.0	1361.63	164.24	103.46	0.42	11.23
300	5400.0	16.9	20.8	1448.73	247.84	43.14	0.50	11.20

Table 3: Master Schedule (δ). Settings: $\pi_i \approx 30\%$, $\alpha = 0.1$, $\epsilon = 0.95$, DesignType=SC

requests. To accomplish this, the design has to properly adjust the time windows. In particular, it seems that to an increased ϵ, the design process responds with increased *distances* between time windows. This explains the changes for idle and passive times. In fact, if the distance between time windows increases but the demand volumes remain the same, the probability the vehicle arrives earlier than the EDT increases and consequently longer idle times are expected. Conversely, because the increased probability that the vehicle departure time is exactly at or close to the EDT, the expected passive waiting time decreases.

We performed for parameter δ a study similar to the one done for ϵ and performed simulations on two designs of a DAS line obtained from two parameter settings differing in the only value of δ. The rest of the parameter setting was $\alpha = 0.1$ and $\epsilon = 0.95$, and the demand scenario considered was such that $\pi_i = 30$. As in the previous case, the simulation was performed on 100 requests scenarios for each direction of the line generated accordingly to the demand scenario considered. In Table 3 we report the results averaged over the set of requests scenarios. The meaning of the column is the same as for the previous experimentation. Because we did not observe significant variations in the way the design produced by SA or SC responded to the parameter variation, we report only the values for the SC approach.

We recall from the discussion in Section 4.3, that time windows width is the tool to commute flexibility among adjacent segments and that the main obstacle in increasing δ is that passive users in the worst case have to wait for the whole length of the time window. By inspecting the values corresponding to the passive times in Table 3, we observe, as expected, a higher value for $\delta = 500$. Notice, however, that the absolute value is still quite low. We also observe that wider time windows allow for a better use of flexibility. In fact, even though the probability the system is able to serve all the stops is the same, both efficiency and latency improve when higher δ are considered, as is it possible to deduce by inspecting the values of UST, ASTI, and latency. The idle times also improved because, for the mechanism described earlier, the probability the vehicle arrives at compulsory stops before the earliest departure time is lower and consequently lower are the expected idle times. Generally speaking, it seems that, when increasing δ, to a small deterioration of the passive times, corresponds a much better performance underlined by the improvements all the other performance measures.

$avg(\pi_i)$	UST	%LSTI	%ASTI	Latency	IdleT	PassiveT	TotReq	%Rej/Total	maxCap
0.3	5230.0	19.5	26.0	1361.63	164.24	103.46	25.8	0.42	11.23
0.7	6097.6	6.2	12.1	1585.39	111.79	145.26	49.7	0.22	20.57

Table 4: Effects of demand volumes. Settings: $\alpha = 0.1$, $\epsilon = 0.95$, $\delta = 500$, DesignType= SC

5.2.3 Effects of demand fluctuations

The last study addresses changes in the design due to different volumes of demand. This issue is relevant for the method described in Section 6 to build the SDDP on a given time horizon starting from the solution of S-SDDP. To this purpose, we considered two demand scenarios, differing in the demand volumes. We generated the scenarios in such a way the average resulting probabilities for demand points to be requested are $\pi_i = 30\%$ and $\pi_i = 70\%$. Observe that, $\pi_i = 70\%$ implies a very high demand volume and it is close to volume levels typical of traditional transit. For each of the two demand volume values, we generated a DAS line design with identical design parameters, namely $\alpha = 0.1$ and $\delta = 500$, $\epsilon = 0.95$. As in the previous case, we then simulated the operations on 100 request scenarios for each direction of the line generated according to the considered demand scenario. We report in Table 4 the results averaged over the set of requests scenarios. The meaning of the columns is as previously but we inserted a new column in position eight to report also the average total amount of requests. Since no significant variation in the way SA or SC responded to the parameter variation was observed, we only report the values for the SC approach. We observe that, in order to keep fixed the probability the system is able to serve all the requests (ϵ), when $avg(\pi_i) = 70\%$, UST, LSTI, ASTI and latency considerably worsen, in particular the UST. This is due to the fact that the design increased the distance between consecutive time windows in order to allocate more time to service requests. It is possible to make observations about idle and passive times similar to those made for the previous two cases. In fact, idle times slightly decrease while passive times increase for $avg(\pi_i) = 70\%$ and this is because the number of requests are likely higher and the resulting probability the vehicle arrives earlier than the earliest departure time at compulsory stop decreases. We finally observe that even if deteriorated, the performances of the DAS are still interesting with respect to a traditional line even when $avg(\pi_i) = 70\%$. In fact, by inspecting the LSTI, we observe that to a reduction of the service time close to 6% in the worse case, corresponds a very low percentage of unserviced requests (less than 0.3%).

5.2.4 Concluding remarks

Summarizing the results of the experimentation, we observe that the two decomposition methods SA and SC mostly differ in the provided topological design, while responding quite similarly to variations of parameters affecting the time windows. In particular, we

observed that SA usually produced segments with more complex configuration than SC but that those "counter-intuitive" configurations were able, under specific settings of parameter α (the relative weight in the objective function between latency and operating costs), to provide the design with lower UST or latency. We also observed that SC, when the number of compulsory stops is low, is weakly sensitive to variation of α. Regarding the effects of the MSP parameters, we observed that to higher values of ϵ (the probability to be able to serve all the customers) usually correspond longer distances among time windows, with the effect of worsening the performances of the system. The effects of variation of the time windows width δ are very interesting: While to higher values of δ statically correspond increasingly worse values of passive times, the experimentation shows that the actual increment of passive time is quite small. The advantages derived from the possibility of commuting flexibility between consecutive segments actually allow for shorter distances between time windows and this considerably improves the performance. Finally, we studied the effects of demand volume variations and we observed that, even in case of very high demand, DAS presents interesting characteristics because to very small percentages of rejected demand still correspond interesting efficiency improvements.

A last observation is related to the computational efficiency. Both SA and SC in the first phase have to solve the GMLP, in the second the MLP. The running times required by those algorithms are reported in Errico et al. (2011a) and Crainic et al. (2010). The main difference between the two decomposition strategies is that SA must run the GMLP on the whole set of demand points while SC only on compulsory stops and this makes SA more challenging from the computational point of view than SC.

6 The S-SDDP as a tool for the solution of the SDDP

The question of how the design of transit systems should accommodate demand fluctuations is actually a general issue as it does not regard the design of DAS lines only. At the present, most methods for the design of traditional transit first build the structure of the system based on aggregated data. Once the service network structure and a first approximation of frequencies is determined and fixed, timetables and (sometimes) frequencies are modified according to the expected demand in a given time period. Presumably, the reason behind fixing the service network, is to keep the management of the system relatively easy and to facilitate customers' understanding of the transit system. Nowadays, however, new technologies make it possible to establish real-time communications between transit system and operators, and between transit system and customers. This allows for more complex configuration of the transit system.

In order to model the general SDDP, where demand may vary during the given time horizon, we need to know how the system is allowed to accommodate demand variations.

The range of possible strategies is wide, the choice depending on several factors among which planning objectives, type of users, available technologies, transit agency policies, etc. We describe here three representative settings. One possibility is to consider the design responding to every demand variation. At the other extreme of the spectrum, one can consider the design of the system fixed and not responding to any demand variations. In between, there is a range of solutions where part of the design remains fixed and part changes according to demand fluctuations. Let us now see how the solution of the S-SDDP can be used to obtain the solution of the SDDP for the three case considered.

In the first scenario considered, the design responds to every demand fluctuation. This means that the design might change at every homogeneous time period present in the time horizon. To see how the solution of the S-SDDP can be used, it is sufficient to consider the homogeneous time periods in the time horizon and partition it in time slices according to the frequency requirements established at strategic planning level. Then consider, for each homogeneous period, demand data related to a single slice and solve the S-SDDP. The solution of the SDDP simply consists of the repetition of the S-SDDP solutions for each homogeneous period.

The second scenario considers the design never responding to demand fluctuation. This means that the design needs to account for different demand volumes at the same time. In this case it is possible to aggregate data over the whole time horizon and deduce demand data relative to a single slice interval. Solve the S-SDDP and repeat the same solution for every slice in the considered time horizon.

The last scenario considers the design partly fixed and partly responding to demand variations. There are several possible ways to choose what part of the design has to be fixed. We consider here a case inspired from a common practice in traditional transit. In such a context, the stops and the sequence of a bus line are usually predefined and fixed, independently of demand volumes. However, schedules might vary in correspondence to demand variations (for examples, schedules differ for rush and night hours). A similar approach for the SDDP would consider a design where the topological part is fixed while the master schedule is adjusted according to demand variations. To obtain the topological design, similarly to the second scenario, one should aggregate the data over the whole time horizon, deduce demand data relative to a single slice, find the topological design and keep it fixed along the time horizon. To obtain the master schedules, similarly to the first scenario, one should consider, for each homogeneous period, demand data related to a single slice, and solve the MSP. Then repeat, for every time slice in a given homogeneous period, the master schedule relative to it.

The brief analysis above explains how the methods developed in the present paper to build the S-SDDP can be easily used to build the solution of the SDDP for general time horizons.

7 Conclusions

The present paper addressed tactical planning issues for the Demand Adaptive Transit System, which is a general model for a wide class of transit systems usually called *semi-flexible*.

We introduced and formally defined the Single-line DAS Design Problem and a specialization of it where the time horizon corresponds to a single time slice (S-SDDP). The S-SDDP requires to establish the location of the compulsory stops, their sequence, the partition of the service area into segments, the time windows of passages at compulsory stops. We proposed two alternative hierarchical decomposition strategies for this problem, the Sequence All and the Sequence Compulsory (SA and SC, respectively). SA and SC differ in the decision-taking order, but they share the resulting core sub-problems. One of the most important features of the proposed methodology is that it does not rely on particularly restrictive assumptions and consequently it is suitable to be applied to real-life cases.

We performed an extensive computational study where the effects of the main design parameters and of variations in demand volumes where analyzed in terms of system performances.

The experimental results showed that the performance of DAS is very interesting with respect to that of traditional transit, as to a significant reduction in terms of service times corresponded very small percentages of requests that could not be serviced within the computed time windows. This turned out to be true even when DAS was tested under high demand volumes. The comparison between SA and SC revealed that, although SA usually produces more complex configuration of the segments, is more sensitive to variations of the relative weight of the latency with respect to operating costs in the objective function. In fact, under specific settings of this weight, SA produced the best designs from the latency or operational-cost points of view. On the other hand, SA is computationally more demanding than SC. Results also showed that SA and SC respond similarly to variations in the values of design parameters related to the master schedule. In particular, to higher values of the probability to be able to serve all the customers corresponded a general deterioration of the system performances. Results also showed that, in most cases, it is advantageous to set relatively high values of the time window width. On the one hand, even if higher width values correspond to higher upper bounds on passive waiting times, the actual observed passive times only slightly increased. On the other hand, wider time windows allow for a better use of flexibility and this considerably improves performances.

We finally showed how the methodology developed for the S-SDDP can be applied to solve the general SDDP, illustrating for three representative particular cases.

Acknowledgments

While working on this project, T.G. Crainic was the NSERC Industrial Research Chair in Logistics Management, ESG UQAM, and Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal, and the Department of Economics and Business Administration, Molde University College, Norway, and F. Errico was postdoctoral researcher with the Chair.

Funding for this project was provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Industrial Research Chair and Discovery Grant programs, by our partners CN, Rona, Alimentation Couche-Tard, the Ministry of Transportation of Québec, and by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT) through its Team Research Project program.

References

Ceder, A. and H. M. Wilson (1997). Public Transport Operations Planning. In C. ReVelle and A. E. McGarity (Eds.), *Design and Operation of Civil and Environmental Engineering systems*, pp. 395–434. John Wiley & Sons, Inc., New York.

Cordeau, J.-F. and G. Laporte (2003). The Dial-a-Ride Problem (DARP): Variants, Modeling Issues and Algorithms. *4OR 1*(2), 89–101.

Crainic, T., F. Errico, F. Malucelli, and M. Nonato (2010). Designing the master schedule for demand-adaptive transit systems. *Annals of Operations Research*, 1–16. 10.1007/s10479-010-0710-5.

Crainic, T. G., F. Malucelli, and M. Nonato (2001). Flexible Many-to-few + Few-to-many = An Almost Personalized Transit System. In *Preprints TRISTAN IV - Triennial Symposium on Transportation Analysis*, Volume 2, pp. 435–440. Faculdade de Ciências da Universidade de Lisboa and Universidade dos Açores, São Miguel, Açores, Portugal.

Crainic, T. G., F. Malucelli, M. Nonato, and F. Guertin (2005). Meta-Heuristics for a Class of Demand-Responsive Transit Systems. *INFORMS Journal on Computing 17*(1), 10–24.

Daganzo, C. F. (1984). The length of tours in zones of different shapes. *Transportation Research B 18*(2), 135–145.

Diana, M., M. M. Dessouky, and N. Xia (2006). A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research Part B: Methodological 40*(8), 651 – 666.

Durvasula, P., B. Smith, R. Turochy, S. Birch, and M. Demetsky (1998). Peninsula transportation district commission route deviation feasibility study, final report. Technical Report VTRC 99-R11, Transportation Research Council, Virginia, USA.

Errico, F. (2008). *The design of flexible transit systems: models and solution methods*. Ph. D. thesis, Politecnico di Milano, Italy.

Errico, F., T. G. Crainic, F. Malucelli, and M. Nonato (2011a). A Benders Decomposition approach for the Minimum Symmetric Hamiltonian Circuit with Generalized Latency. CIRRELT publication, Centre interuniversitaire de recherche sur les réseaux dentreprise, la logistique et le transport, Université de Montréal.

Errico, F., T. G. Crainic, F. Malucelli, and M. Nonato (2011b). An unifying framework and review of Semi-Flexible Transit Systems. CIRRELT-2011-64, Centre interuniversitaire de recherche sur les réseaux dentreprise, la logistique et le transport, Université de Montréal.

Fu, L. (2001). Simulation Model for Evaluating Intelligent Paratransit Systems. *Transportation Research Record 1760*, 93–99.

Fu, L. (2002). Planning and Design of Flex-Route Transit Services. *Transportation Research Record 1791*, 59–66.

Koffman, D. (2004). Operational Experiences with Flexible Transit Services. Volume 53 of *TCRP Synthesis Report*. Transportation Research Board.

Malucelli, F., M. Nonato, T. G. Crainic, and F. Guertin (2001). Adaptive Memory Programming for a Class of Demand-Responsive Transit Systems. In Voß, S. and Daduna, J.R. (Eds.), *Computer-Aided Scheduling of Public Transport*, Volume 505 of *Lecture Notes in Economics and Mathematical Systems*, pp. 253–273. Springer, Berlin.

Malucelli, F., M. Nonato, and S. Pallottino (1999). Some Proposals on Flexible Transit. In Ciriani, T.A., Johnson, E.L., and Tadei, R. (Eds.), *Operations Research in Industry*, pp. 157–182. McMillian.

Potts, J. F., M. A. Marshall, E. C. Crockett, and J. Washington (2010). A Guide for Planning and Operating Flexible Public Transportation Services. Volume 140 of *TCRP Synthesis Report*. Transportation Research Board.

Quadrifoglio, L., M. M. Dessouky, and F. Ordóñez (2008). A simulation study of demand responsive transit system design. *Transportation Research Part A: Policy and Practice 42*(4), 718 – 737.

Quadrifoglio, L., M. M. Dessouky, and K. Palmer (2007). An insertion heuristic for scheduling Mobility Allowance Shuttle Transit (MAST) services. *J. of Scheduling 10*(1), 25–40.

Quadrifoglio, L., R. W. Hall, and M. M. Dessouky (2006). Performance and design of mobility allowance shuttle transit services: Bounds on the maximum longitudinal velocity. *Transportation Science 40*(3), 351–363.

Smith, B. L., M. J. Demetsky, and P. K. Durvasula (2003). A Multiobjective Optimization Model for Flexroute Transit Service Design. *Journal of Public Transportation 6*(1), 89–100.

Welch, W., R. Chisholm, D. Schumacher, and S. R. Mundle (1991). Methodology for evaluating out-of-direction bus route segments. *Transportation Research Record* (1308), 43–50.

Wilson, H., I. Sussman, H. Wang, and B. Higonnet (1971). Scheduling algorithms for dial-a-ride systems. Technical Report USL TR-70-13, Urban Systems Laboratory, MIT, Cambridge, MA.

Zhao, J. and M. Dessouky (2008). Service capacity design problems for mobility allowance shuttle transit systems. *Transportation Research Part B: Methodological 42*(2), 135 – 146.