Flexible many-to-few + few-to-many = an almost personalized transit system

Teodor Gabriel Crainic
D.S.A.-U.Q.A.M. and Centre de Recherche sur le Transports - Université de Montréal
theo@crt.umontreal.ca
Federico Malucelli
Dipartimento di Elettronica e Informazione - Politecnico di Milano
malucell@elet.polimi.it
Maddalena Nonato
Dipartimento di Ingegneria Elettronica e Informazione - Università di Perugia
nonato@diei.unipg.it

1 Introduction

In this paper we analyze a new transportation system where several transit lines are operated in an integrated environment with conventional lines, (i.e., with predetermined timetable and itinerary), interact with lines of innovative kind, (i.e., with flexible itinerary and timetable) as those proposed in [3]. The peculiarity of the proposed system is to provide an almost personalized transportation service at the cost of a traditional, fixed route, transportation system. This system is particularly suited for medium/low demand transportation settings, and goes into the direction of sustainable mobility in large city suburbs or in an inter urban setting by enhancing the attractiveness of public transportation with respect to the private car. The system exploits the idea that the seamless integration of flexible "many to few" and "few to many" flexible systems, yields a "many to many" transportation system able to meet the personal needs of transportation of a large set of customers. If the system is operated efficently it is possible to achieve a service level almost comparable with a "personalized" door to door transportation without the need of the overhead usually required by the management of real door to door services.

In this paper several variants of the service are analyzed, and the optimization aspects involved by efficient management are enlighted. Mathematical models and the algorithmic approaches are also discussed.

2 General context of the service

The system we contemplate is designed as follows. Let us consider a set of bearing swift lines (such as underground, light surface rail, or express bus lines) and a set of flexible lines (possibly acting as the feeders of the first ones). The traversal of a line is described as usually in terms of a set of timetabled trips. Indeed, flexible lines are described in terms of *compulsory stops*, that is, a set of stops where the vehicle must transit each within a specified time window and where typically passengers may transfer to other lines (both fixed or flexible). Besides compulsory stops, a set of stops on-demand (*optional stops* hereafter) is available to the users; each optional stop is located in the area between a pair of compulsory stops which are consecutive along the line. In this framework, a user may issue a transportation request concerning a boarding and an alighting stop. At least one of these two may be optional: in this case we have a so called *active user* because he/she must explicitly make a reservation for the service. On the contrary, if both stops are compulsory the service is comparable to the one provided by a traditional fixed route line, so that there is no need to make a reservation. More generally, the user itinerary may be given by a chain of portions of trips on flexible and fixed lines. The vehicle itineraries depend on the transportation requests currently issued. In the absence of requests involving optional stops, the vehicle travels along the shortest path on the network between each pair of consecutive compulsory stops. A request involving optional stops induces a partial rerouting of the vehicle in order to serve the involved stops; due to time windows at compulsory stops, requests may happen to be refused as otherwise the route would be infeasible. In fact the detour may cause a delay of the transit time at the later stops.

In an integrated environment, where all lines are part of a seamless network, users may travel from one of the many optional stops to any other one, transfering at one of the few compulsory stops, and possibly taking advantage of swift line transportation. Depending on the actual design of the service network, the optimal itinerary of each user is not determined in a unique way, but it depends on the current schedule and on the availability of the transportation system, that is on the itineraries induced by the current set of requests.

In this sense the vehicle route adapts itself to catch the current demand, and this service model is called Demand Adaptive System.

We assume that, in case a request cannot be served by the regular service, a collective taxi or other forms of personalized transportation services are provided to the user without extra fees so that the minimum guaranteed service level is maintained. Moreover, we disregard any capacity constraint of the vehicles assuming that the vehicles capacity is sufficient and that in a low demand setting the capacity is not a tight constraint. This assumption holds in particular if in the system may coexists active and passive users, for which it is not possible to predict the request of capacity in the vehicles.

Let us highlight the main issues to be considered in the proposed transportation system:

1. *Synchronization*
   At compulsory stops, where transfers occur, vehicles have to be synchronized so that connections between lines are guaranteed. Note that also synchronizations may be dealt in an on-demand base as for optional stops. We consider two possible ways to deal with the synchronization issue:

   a Synchronization is solved at design level. That is a set of possible connections is determined a priori and are enforced regardless of current demand for tranfer. For example, each time a vehicle may meet other vehicles at the same compulsory stop, or in two compulsory stops within walking distance, (i.e., when their scheduled time windows intersect), the vehicle is allowed to leave only after the other vehicle has arrived. Thus at design level, time windows must be opportunely designed to enforce the synchronization. This idea is particularly suitable in case of flexible lines connecting with a fixed one, that is when flexible lines act as *feeders*.

   b Synchronization is determined dynamically depending on the actual requests of transportation. This kind of approach implies that a decision must be taken each time requests are considered and vehicle itineraries are determined. In fact allowing a synchronization may induce a delay in a vehicle itinerary so that possible subsequent detours are no more possible. Note that, due to the dynamic setting, also users going from a compulsory stop to another compulsory stop must make an explicit reservation if their trip involves connections.

2. *Request management*
   From the service management point of view, we consider two possible ways of collecting requests, the first one goes towards a real-time service, the second one achieves better service levels and a more profitable management:

   a Requests are collected and managed *on-line* and dynamically in real time during the service operation. This means that the optimization of the itineraries (both from the user and form the service management viewpoints) is done by considering one request at a time. Requests are accepted and the vehicle itineraries of the flexible lines are possibly modified starting from the next compulsory stop they will pass by. Note that in this way, although the position of the vehicle must be kept updated, a sophisticated vehicle positioning system is not really needed, in fact it suffices to know the last compulsory stop served by each vehicle to have enough information about vehicles positions. In any case, the optimization problems to be solved each time a request is issued are rather trivial, though the attained result could be far from the global optimality that could have been achieved if all requests had been known in advance.

b Requests are collected *off-line*, and some time in advance with respect to the requested departure time (e.g., 30') they are reconfirmed to the users providing them with their complete routes (boarding and alighting stops, departure and arrival times, boarded vehicles, transfers). Obviously, in this case the optimization of the vehicle and users itineraries can be made on a broader basis, thus better results can be obtained with respect to the on-line approach. This means also that more difficult problems must be solved each time a bunch of requests is processed.

3. *Request classification*

   Each active user specifies either the earliest departure time from the origin stop (*target time at the origin*), or the latest arrival time at destination (*target time at the destination*). The *actual travel time* of the user is thus computed as the difference between the arrival time and the target time at the origin (in the first case), or the difference between the target time at the destination and the actual departure time (in the second case). We may distinguish requests on the basis of the of minimum service level guaranteed (which is usually a function of the actual travel time). Suppose that for each pair of origin/destination stops of the network (or more broadly for pairs of origin/destination areas) an *ideal travel time* is known. Such knowledge can be assumed to define the minimum service level that the transit company assures to its users: for example the company may guarantee that the maximum travel time is no longer than a multiple of the ideal travel time. This maximum acceptable travel time, combined with the target time at the origin ore at the destination, defines the so called *customer time window.* We have two cases to consider:

   a *Single class of users*: the transportation company assures the same service level to all users.

   b *Multiple classes of users*: several classes of users can be considered each having its own level of service and its own service fare; for example we may think of an express service with a small customer time window, and a regular service with larger time window and lower fare.

4. *Additional flexibility*

   In order to satisfy the maximum number of requests, it may be useful to introduce additional flexibility in the services. For example it may occur that some request cannot be satisfied because it is not possible to let the vehicle pass by the requested stops at the requested time. In this case the service operator may negotiate with the user a *displacement in space* (proposing him/her to board or alight in alternative stops in the neighborhood of the desired ones) or a *displacement in time*, (proposing alternative pick up and/or drop off times in a broader time window with respect to the expected one). These displacements may involve a discount in the service fare to compensate the inconvenience.

3 Mathematical models

In this section we present some mathematical models for the integrated transportation system described above. As mentioned before, the on-line case does not really involve an optimization process as the requests are processed sequentially. Here we devise models only for the more challenging off line case. Let us make some assumptions, just to have a well defined transportation system to discuss as a starting base. In particular we assume that the synchronization is made at the design level (1a), a single class of users is considered (3a), no additional flexibility is allowed (neither in time nor in space), and non overlapping optimization periods are considered. We will generalize the mathematical models to all the other possible cases of the framework. We are given the following data:

- a set of lines L (which can be either flexible or conventional) and for each line $l \in L$ a set of occurrences of the trip $\{1, \ldots, K_l\}$;
- each line l is defined by a sequence of compulsory stops f_h^l , with $h = 1, \ldots, n^l$, and $f_1^l = f_{n^l}^l$ is the terminal of the line, assuming to have a circular line, and that the line does not change with the trip occurrence.
- for each line l, compulsory stop f_h^l, line and trip occurrence k a time window $[a_h^{lk}, b_h^{lk}]$ defines when the vehicle operating that trip occurrence is allowed to leave from the stop; note that in case of fixed lines all stops are compulsory and each time window reduces to a single value (i.e., the scheduled departure time);
- for each flexible line l and trip occurrence k, a set of optional stops N_h^l is defined for each pair of consecutive compulsory stops f_h^l and f_{h+1}^l; the union of all N_h^l and of all compulsory stops f_h^l defines the set of stops served by the system; let us call N this set. Moreover let us denote by N^l the set of stops involved by line l.
- R is the set of requests currently available; each request r is specified by the desired departure stop $s(r) \in N$, the desired arrival node $d(r) \in N$, the customer time window $[a(r), b(r)]$, that is the interval of time in which customer of request r can travel, the benefit of serving request $u(r)$ (for example the ticket fare). Let $\bar{u}(r)$ be the cost of the taxi ride from $s(r)$ to $d(r)$, in case the request cannot be served by the regular service.
- the travel time and the travel cost between any two nodes i and j of the physical network are known and are given by τ_{ij} and c_{ij} respectively;
- the planning horizon is T, where a suitable value for T can be twice the minimum common multiple of the lines headways.

The observations on which we base the mathematical formulation is that vehicles itineraries (called *tours*) can be decomposed into sequences of paths between consecutive compulsory stops. Also passenger itineraries (called *routes*) can be decomposed similarly, except for the first and the last portions which may involve optional stops as starting or ending points. Paths between consecutive compulsory stops can be taken as basic components of the formal description of the problem. We call *segment* h of tour lk (also referred as segment $\sigma = (l, k, h)$) the subgraph defined by node set $N_h^l \cup \{f_h^l, f_{h+1}^l\}$. Let S denote the set of all segments $\sigma = (l, k, h)$ for $l \in L, k = 1, \ldots, K_l$, and $h = 1, \ldots, n^l$. Let P_σ be the set of feasible paths for vehicle lk between f_h^l and f_{h+1}^l, that is the set of elementary paths in segment $\sigma = (l, k, h)$ starting in f_h^l and ending in f_{h+1}^l whose travel time does not exceed $b_{h+1}^{lk} - a_h^{lk}$. We assume that the simplest path given by the single arc (f_h^l, f_{h+1}^l), (*basic path* hereafter) always belongs to P_σ. In case of a fixed line, P_σ contains only the basic path, while in the case of flexible lines P_σ can be exponentially large depending on the cardinality of the segment node set and on the width of the time windows at the compulsory stops. Assuming that the sets of feasible paths for each segment are given, we can construct a suitable *path graph* used to formulate the problem. In the graph there is a node for each path in P_σ, for all $\sigma \in S$. The arcs of the path graph are classified as follows:

1. *Compatibility arcs*: These arcs connect two nodes representing paths which belong to consecutive segments, or more generally, that must be run consecutively by the same vehicle. There is an arc between two nodes if there exist feasible departure times for the corresponding paths such that the itinerary defined by the two paths can be operated by the same vehicle. In general, any two paths in P_σ and $P_{\sigma'}$ with $\sigma = (l, k, h)$ and $\sigma' = (l, k, h + 1)$ are not necessarily compatible especially if they have both a long travel time and the time windows are narrow. Note that there are compatibility arcs between nodes corresponding to paths of the last segment of one line and of the first one of the next occurrence, if they are operated by the same vehicle.

2. *Connection arcs*: These arcs connect two nodes representing paths belonging to segments of different lines. In particular there will be a connection arc if the arrival time of the first path is smaller than the departure time of the second one (and maybe also not greater than a given maximum waiting time) and the ending compulsory stop of the first segment is the same or within walking distance from the starting compulsory stop of the second segment. Note that by the assumption of fixed synchronization, if two segments σ and σ' are synchronized, then there are connection arcs between any two nodes representing paths in P_σ and $P_{\sigma'}$ respectively.

3. *Boarding and alighting arcs*: The boarding arcs connect node r' with the nodes corresponding to paths passing by optional stop $s(r)$ within a feasible departure time for r, while the alighting arcs connect nodes corresponding to paths passing by optional stop $d(r)$ within a feasible arriving time for request r and node r''.

4. *Taxi arcs*: They connect directly r' with r'' for each request r.

The problem consists in selecting exactly one path per segment so that paths of consecutive segments are compatible, and to determine the optimal routes of all requests by using the selected paths. Let us denote by π a feasible route and let $\Pi(r)$ be the set of all feasible routes for request r; obviously the taxi arc (r', r'') always belongs to $\Pi(r)$: let us call $\bar{\pi}_r$ the path given by the taxi arc (r', r''). We introduce the following variables: z_p which is equal to one if and only if path $p \in P_\sigma$ is selected, and zero otherwise, t_σ which gives the departure time of the vehicle from the starting compulsory stop of segment $\sigma = (l, k, h)$, θ_π which is equal to one if and only if request r is routed through $\pi \in \Pi(r)$, and zero otherwise. For notational purposes we introduce coefficients $\delta_{p\pi}$ that are equal to one if path p appears in route π, and zero otherwise. Let also $c(p)$ and $\tau(p)$ denote the cost and the travel time of path p. The problem of optimally managing the service is thus formulated as follows:

$$P2 : \min \sum_{\sigma \in S} \sum_{p \in P_\sigma} c(p) z_p + \sum_r \bar{u}(r) \theta_{\bar{\pi}_r} \tag{1}$$

$$\sum_{p \in P_\sigma} z_p = 1 \qquad \forall \sigma \in S \tag{2}$$

$$t_\sigma + \sum_{p \in P_\sigma} \tau(p) z_p \leq t_\sigma \qquad \forall \text{ consecutive } \sigma, \sigma' \in S \tag{3}$$

$$\sum_{\pi \in \Pi(r)} \theta_\pi = 1 \qquad \forall r \in R \tag{4}$$

$$\sum_{\pi \in \Pi(r)} \delta_{p\pi} \theta_\pi \leq z_{\sigma p} \qquad \forall r, \forall \sigma \in S, \forall p \in P_\sigma \tag{5}$$

$$a_h^{lk} \leq t_h^{lk} \leq b_h^{lk} \qquad \forall \sigma \in S \tag{6}$$

$$z_p, \theta_\pi \in \{0, 1\} \qquad \forall r, \forall \pi \in \Pi(r), \forall \sigma \in S. \tag{7}$$

In the objective function (1) we should maximize the difference between the benefits of collecting requests which are constant because we assume to satisfy all requests (i.e., $\sum_r u(r)$), and the traveling costs plus the costs of taxi rides. Thus the objective function reduces to minimizing the global cost. Constraints (2) guarantee the service coverage by imposing the selection of one path for each segment. In constraints (3) we consider pairs of segments run consecutively by the same vehicle; segments $\sigma = (l, k, h)$ and $\sigma' = (l', k', h')$ are consecutive for example when $l' = l$, $k' = k$ and $h' = h + 1$ for $h = 1, \ldots, n^l - 1$, or when $h = n^l$ and $h' = 1$ and the two segments have been assigned to the same vehicle. The constraints state that one vehicle cannot leave from one compulsory stop before it has arrived from its duty in the preceding segment. One route must be selected for each request (constraints (4)), and if a route is selected each portion of the route must be served by a vehicle (constraints 5). The usual time window constraints (6) complete the model. Note the time windows constraints are not present in model $P1$ because they are implicitly stated by the space-time graph structure.

The above formulation can be very large: the number of variables related to path selection (z_p) and especially to route selection (θ_π) can be huge, also the number of constraints can be very high in particular for (5). This is sufficient to make the model unbearable by any commercial integer linear programming solver. However the feasible paths within the segments and the request routes can be dealt quite efficiently from an algorithmic point of view; we will exploit the proposed mathematical model as a suitable representation used to devise efficient heuristic algorithms.

In the full paper we show how to generallize the above formulation in order to include all the other aspects of the problem such as different class of users, dynamic synchronization and additional flexibility. Moreover we specialize the model for a particular case and we introduce decomposition techniques that make possible to approach its solution with the algorithms developed for the single line case in [1].

References

[1] Crainic, T. G., Malucelli, M. Nonato, F. Guertin. *Adaptive memory programming for a class of demand responsive transit systems* Technical Report CRT-2001-08 (2001).

[2] Gendreau, M., J. Y. Potvin, *Dynamic vehicle routing and dispatching*, in Fleet Management and Logistics, T. G. Crainic and G. Laporte Editors. 1998, Kluwer Academic Publishers: Boston 115-126.

[3] Malucelli F., M. Nonato, S. Pallottino, *Demand Adaptive Systems: some proposals on flexible transit*, in Operations Research in Industry, T. A. Ciriani, S. Gliozzi, E. L. Johnson and R. Tadei Editors. 1998, McMillan Press: London. to appear (1999).