A new methodology for the public transport network design

Francesco Ciaffi, Ernesto Cipriani, Marco Petrelli, Rasa Ušpalytė-Vitkūnienė

Abstract

The present paper deals with the bus network design problem. Such problem is formulated as an optimization problem involving the minimization of all resources and costs related to the public transport system. The optimization problem is subject to user equilibrium on public transport network as well as to the bus capacity constraints and a set of feasibility constraints on route length and line frequency. The objective function is defined as the sum of operator’s costs and users’ costs. The input data are the public transport demand matrix, the characteristics of road network, the operating and users unit costs. Outputs are routes and frequencies for the lines of the public transport network. The performances of the network are estimated by a hyperpath transit assignment model, which reproduces the choice behaviour of transit users. The solving procedure consists of a set of heuristics, which includes a first routine for the definition of the roads and the zones to be served, a second step for the routes generation and then a genetic algorithm for finding a sub-optimal set of routes and associated frequencies. The GA is implemented in the C# language as a parallel genetic algorithm while the fitness evaluation requires computing, for each solution generated, the two terms of the objective function by simulating the public transport network with the EMME software. The proposed procedure will be implemented on a real large size network (two districts in the city of Rome), in order to compare its effectiveness with the performances of the existing transit network and to provide an extensive sensitivity analysis in bus frequency changes.

Keywords: Public transport network design; frequency setting; metaheuristics.

1. Introduction

In the last decades there was a huge developments of policies and strategies for managing and planning urban public transport networks, mainly due to the suburban sprawl that can be found all over the world and the resulting overuse of cars. Differently from past studies and experiences, nowadays most of researchers have to take into account main issues of sustainable mobility, introduced from 80’s just for facing increases of air and noise pollution, energy consumption, traffic congestion and car accidents. Designing a reliable and efficient public transport network represents one of the most effective tools to achieve final goals of sustainability.

Based on such issues, this paper presents a methodology for solving the “Bus Network Design Problem” (BNDP). The BNDP consists in determining the optimal network configuration in terms of bus routes and service frequencies in order to minimize an objective function representing total costs involved by the public transport system considering fixed origin-destination demand.

The design criteria implemented try to develop an intensive rather than extensive bus network in order to improve efficiency, integration among direct routes and effective transfer points that strongly affect service quality and ridership. The basic framework of the procedure is based on three sequential stages: 1) a first routine for the definition of the roads and the zones to be served; 2) an heuristic algorithm that generates a set of feasible routes; 3) a genetic algorithm (GA) that finds a sub-optimal set of routes with the associated frequencies. The performances of the network are estimated by a hyperpath transit assignment model, which reproduces the bus line choice behaviour of transit users. The procedure provides as outputs the set of bus routes and their associated frequencies solving simultaneously the two sub-problems.

The structure of this paper is organized in 6 sections including an introduction about the BNDP. The second section provide an extensive review of the state of the art about bus networks design problem. The third section deal with the problem formulation and the solving procedure framework. Finally are reported a detailed analysis of literature and a sensitivity analysis both focused on frequency setting problem.
2. BNDP in literature

The transit network design is a complex non convex problem [1, 19]. It is usually formulated as a non-linear optimization problem with both discrete and continuous variables and constraints. The best and most efficient solution methods are based on heuristic procedures but their applications are mainly limited to test cases or real-life networks of small size. A global review about route design, frequency setting, timetabling of transit lines and their combination is proposed by Desaulniers and Hickman (2007) [12], Guillaume and Hao (2008) [14], Kepaptsoglou and Karlaftis (2009) [15].

The best and most efficient solution methods are based on heuristic procedures due to the non-convexity and the combinatorial and multi-objective nature of the problem. Among the most remarkable works on this matter, it is possible to mention Mandl (1979) [17], Baaj and Mahmassani (1995) [2], Ceder and Israeli (1993) [5], Carrese and Gori (2002) [3] and Lee and Vuchic (2005) [16]. Mandl (1979) [17] developed an algorithm that, at each step, individuates a different route as the shortest path connecting a pair of terminals and serves the greatest number of origin-destination node pairs. Baaj and Mahmassani (1995) [2] used an Artificial-Intelligent heuristic algorithm for route generation, which selects a given number of high-demand node pairs and builds an initial network skeleton by connecting these node pairs with the shortest paths. The skeleton is then progressively expanded to routes according to a node selection strategy that reflects different trade-off between performance measures, users’ and operators’ costs. Ceder and Israeli (1993) [5] proposed a transit network design procedure based on the mathematical programming approach: the first step generates a very large set of feasible routes connecting every node to all others. Then, the system creates the minimal subsets of routes solving a Set Covering Problem and selects the most suitable subset by applying a multi-objective analysis. Carrese and Gori (2002) [3] proposed a bus network design model for the development of a hierarchical transit system. The design procedure is subdivided in 2 phases: in the 1st one, the skeleton of the bus network as resulting from a flow concentration process is identified. Then, in the 2nd phase, integrated bus network levels, hierarchically articulated in express, main and feeder lines, are defined. Lee and Vuchic (2005) [16] proposed an iterative procedure that, starting from an initial set of routes composed by the shortest paths for all the origin-destination pairs, tries to improve it by realigning the routes and eliminating the less efficient ones, explicitly taking into account the change of modal split.

In the last years, the evolution of operational research and computer technology produced great and renewed attention for the transit network design problem. New approaches based on metaheuristic techniques (Genetic Algorithms, Simulated Annealing or Tabu Search) have been frequently applied to solve optimization problems. Due to the discrete nature of several variables as well as the nonlinearity and the nonconvexity of the objective function, probabilistic optimization techniques as Genetic Algorithms (GA) seem to be suitable. Xiong and Schneider (1993) [25] showed that GA may efficiently solve transport network design problem and Chakroborty (2003) [6] analyzed and showed the effectiveness of GA in solving the urban transit network design problem. Other authors have applied GA to solve specific transit network design problems (routing, scheduling, etc.). Pattnaik et al. (1998) [21] implemented a two phases procedure for the transit network design. First, a set of feasible routes is generated and, then, a GA selects the optimum routes’ set. Much effort is done in the coding-decoding phase: any single line belonging to the candidate network is represented by a binary substring whose length depends on the precision needed in the decoding scheme. Such an approach is impractical when dealing with real size network because it would imply very long string. Krishna Rao et al. (2000) [22] used GA technique to solve sequentially the routing and scheduling problems by means of two different applications of GAs: the optimal network configuration is obtained in the first application; such network is the input for the second GA that searches for the optimal set of frequencies. Adopting a more accurate approach, the frequency setting problem should be solved within (simultaneously) the optimal network configuration searching. Ngamchai and Lovell (2003) [20] proposed new genetic operators specifically designed for the route generation phase problem.

3. Problem formulation and solution framework

3.1. Problem formulation

The optimization problem can be formally defined as follows:

\[
\left( \tilde{r}, \tilde{f} \right) = \arg \min \ z(r, f, q_t^*) \quad (0.1)
\]

subject to hyperpath assignment of the passenger demand \((D_{ij})\) on transit network, where is the user route choice model function

\[
q_t^* = \Lambda[C_t(r, f)] \quad (0.2)
\]

and a set of feasibility constraints that define both minimum and maximal values for route length and bus frequency:

\[
L_{\text{min}} \leq L_r \leq L_{\text{max}} \quad (0.3)
\]

\[
f_{\text{min}} \leq f_r \leq f_{\text{max}} \quad (0.4)
\]
The road network is represented with the definition of a unidirectional graph \( G = (N,E) \), where \( N \) is the set of nodes and \( E \) is the set of links representing connections between nodes. A route is a sequence of adjacent nodes in \( G \).

The public transport supply is represented as a frequency based service. Equation (1.2) represents the demand-supply consistency constraints (assignment constraints). Such constraint corresponds to a hyperpath approach for the simulation of user choice behaviour on transit (see Spiess and Florian 1989) \([23]\). Transit capacity constraints are not explicitly considered in the assignment phase. Vehicle capacity is taken into account when computing line frequency in order to prevent high crowding levels on buses.

The feasibility constraints for route length (1.3) and line frequency (1.4) have been introduced. The required frequency of service on the resulting route does not exceed the maximum operationally implementable value because it is impossible to maintain as well as the length does not exceed a maximum allowable value because schedules are too difficult to maintain. Analogously, both frequency and length are not lower than a minimum value because it is not possible to serve a very close origin-destination demand pair with a bus (walking would be a better option and no operator would maintain such a line) or to offer a very low frequency service in an urban context (it would be perceived by users as no service availability).

The objective function \( z \) is defined as the sum of operator’s costs \( z_1 \) and users’ costs \( z_2 \) plus an additional penalty related to the level of unsatisfied demand \( z_3 \):

\[
z = z_1 + z_2 + z_3
\]

This objective function formulation, the same proposed by Cipriani et al. (2012a) and Cipriani et al. (2012b) \([9–10]\), is developed to properly weight \( z_1 \), \( z_2 \) and \( z_3 \) terms in order to represent specific needs of bus networks. Public transport operator’s costs \( z_1 \) are computed as a combination of total bus travel distance and total bus travel time. The public transport users’ costs \( z_2 \) are a weighted sum of in-vehicle travel time, access time, waiting time and a transfer penalty. To provide transit services to as many public transport users as possible, another additional component is included in the objective function \( z \). This supplementary term \( z_3 \) represents a penalty that is proportional to the unsatisfied public transport demand of the design network. The third term reflects the need to reject the banal solution of minimum cost (“zero users and zero service”). Solutions characterized by large increase of the unsatisfied origin-destination demand are also discarded.

The input data are the public transport trip demand matrix, the characteristics of the road network and the rapid rail transit system, the operating and users’ unit costs. Outputs are bus routes and their frequency as well as the total costs and the vector of flows on the public transport network.

### 3.2. Solution framework

The proposed solution framework consists of three main stages:

1. a first routine for the definition of the roads and the zones to be served;
2. a heuristic route generation algorithm (HRGA) that generates a large and rational set of feasible routes, by applying different design criteria and practical rules;
3. a genetic algorithm (GA) that finds the optimal sub-set of routes and their frequencies.

Of course, given the well-known non-convexity of the problem and the heuristic nature of the method, there is no guarantee that the solution found will be optimal.

In the first stage of the solution procedure, a basin definition procedure is carried out aiming at identify areas and nodes to be served on the basis of the expected number of trips generated and/or attracted. This definition is completed by the selection of a subset of the road network in order to identify which road links are involved in the connection with the zones to be served.

In the second stage, a heuristic algorithm (HRGA) generates three different and complementary sets of feasible routes. A first set of routes (called A-type routes) is composed of direct routes connecting the highest demand node pairs not served by the rail system. The second feasible set (B-type routes) aims at developing a bus network composed of routes connecting the main transit centers (eventually railway stations) and of links carrying the highest passengers volume level. Finally, the third set (C-type routes) is simply made up of the routes of the existing bus network. The route generation phase is carried out by heuristic techniques in order to obtain routes built according to specific criteria of efficiency and rationality.

The set of feasible routes is then the input data for the successive stage where the GA is utilized. The design variables are public transport routes; some of them are selected by the GA thus generating a public transport network that has to be evaluated in terms of the objective function. For this evaluation, one key steps have to be performed: the public transport service frequency setting. As this step influences users route choices, the procedure is performed iteratively and in combination with public transport assignment.

Stage 1 and 2 has been implemented in the EMME scripting language; in stage 3, the main procedure (GA) has been implemented in C# language as a parallel genetic algorithm running ad hoc EMME macros to perform the public transport assignment and to compute the objective function components. Specific details of the solving procedure are reported in Cipriani et al. (2012a) \([9]\).
4. Frequency setting assessment

BNDF is, in general, composed of two sub-problems: the route design and the frequency setting. This last optimization problem aims to determine the time interval between subsequent runs for a set of public transport lines given by their routes (sequences of stops and links sections). The solution usually has to satisfy a given origin-destination demand and a constraint on the available fleet of buses. During the strategic planning of a public transport system (in particular, when designing the route of the lines), a preliminary setting of frequencies is needed. Also, during the tactical planning, it is necessary to adjust the frequencies to demand variations along different seasons of the year or time of day, or as response to changes in the route network design. The frequencies impact both on the users (waiting time, capacity of the lines) and also on the operators (operational cost determined strongly by the size of the required fleet).

The optimization problem has been approached in the literature as an optimization problem, where usually the objective function states the minimization of the overall travel time of the users (walking, on-board and waiting), under a fleet size constraint as well as other infrastructure and policy constraints. Since frequency optimization models should include measures relative to the performance of the systems from the viewpoint of the users, they should include a sub-model of the behaviour of the users with respect to a set of bus lines. For this reason, such optimization problem has to be considered as a bi-level problem: the lower level is the assignment model, usually with a complex formulation and solution method, while the upper level involves the frequency optimization.

The problem solution process can be made: separately from the design of routes or, simultaneously, with the route design. In the first case, the frequency optimization deals with the routes of the lines as predefined input. Procedures of this type, in the literature, differ for the optimization techniques applied: the methods based on the passenger load profile of a single line (Ceder 1984) [4]; the approach of Constantin and Florian (1995) [11] where the problem is stated as a nonlinear bilevel problem solved by an iterative algorithm based on a gradient descent which uses specific properties of the problem; the enumeration and calculation of all the possible solutions, only on very small networks (Chan 2005) [7]; metaheuristics techniques such as Hooke and Jeeves algorithm (Dell’Olio et al. 2011) [13], Tabu Search (Martinez et al. 2009) [18] and GA (Yu et al. 2010) [26]. In the second case, the approach requires the simultaneous solution of the route and frequency optimization problem. Also in this case, the solution approach is based on metaheuristics techniques as Simulated Annealing combined with greedy and Tabu Search (Zhao and Zeng 2008) [27], the Greedy algorithm (Szeto and Wu, 2011) [24] or the GA combined with an internal frequency setting procedure based on the maximum load sections (Cipriani et al. 2012 a, b) [9–10].

5. Numerical tests about the frequency setting assessment

The solving procedure has been applied on two north-eastern urban districts of the city Rome. The study area, about 145 km² wide and with a population of about 380,000 inhabitants is characterized by the presence of two subway lines (Metro B and B1) and one urban railway service (FL1) plus a bus network composed of 31 lines, with many overlapping routes and low frequencies (average lines headway is about 25 minutes). The bus service is managed by one operator with a fleet of about 190 vehicles in the morning peak hour.

Starting from the results of this procedure application, described extensively in Ciaffi et al. (2012) [8], this section focus the analysis on the frequency setting problem showing the results of a sensitivity analysis with respect to the lines frequency. Generally speaking, a sensitivity analysis is used to determine how “sensitive” a model is to changes in the value of the parameters of the model and to changes in the structure of the model. In this case, the attention is focused on performance sensitivity respect to the lines frequency. The analysis was performed as a series of tests using sets of different values of lines frequencies to see how a change in this parameter causes a change in the bus network performance.

The analysis is carried out using 8 different values of headway (4, 8, 10, 12, 15, 20, 30, 60 minutes) and their different combination so generating more than 200 different scenarios. A sub-set of 31 scenarios, containing the most significant ones, has been used to analyzing performances of bus network with 3 different levels of origin-destination demand (the actual one and two lower values).

All scenarios generated provide very useful results and a synthesis of these is reported in Figure 1 and Figure 2. The first figure shows the comparison of the changes, for the 31 different scenarios, of the average waiting times respect to the procedure solution computed for the 3 different demand levels. The second one compares the fleet size among the 31 different scenarios and the 3 procedure solution (straight lines) for the different demand levels.

The results obtained, as expected, are in line with the main contributions presented in the literature, showing modest difference among solutions in terms of performances by the viewpoint of users (average waiting times). Taking into consideration the high level of demand, the frequency is very close to the maximum allowable value for a large number of lines depending on capacity more than other issues. Reducing the level of demand, there is an increase in the changes of average waiting times but not so large in absolute values (few minutes) depending also by the assessment of these values (half of the headways for regular services). Results show low decrease of the average waiting time even in presence of large increases in the size of the bus fleet. By the other viewpoint, the operator one, the changes in fleet size are instead always very large passing from value larger than 225 to values lower to 50 buses. It is important to underline that fleet size, an effective indicator of operator services cost, has to be considered much more sensible to frequency changes respect to users performance indicator as waiting times.
Conclusion

In this paper, the authors propose a procedure for solving the bus network design problem in a urban area characterized by a multimodal transit system. The solving procedure consists of a set of heuristics, which includes a first routine for the definition of the roads and the zones to be served, a second step for the routes generation and then a genetic algorithm for finding a sub-optimal set of routes and associated frequencies.

From the point of view of the level of application of the procedure, the present version allows to properly deal with a different and more detailed level of street network representation, more appropriate for the bus network design problem, directly considering the whole road network including also local roads and other specific detail of transport supply. The application of various generation criteria in the HRGA has led to a consistent, diversified and exhaustive set of feasible routes. The implemented genetic algorithm has proved to be robust and effective in producing reasonable solutions. Specific numerical experiments about performance sensitivity to changes in lines frequency, carried out on the network of two districts if the city of Rome, highlighted that fleet size is much more sensible to supply changes in terms of runs respect to performance for users.

Based on such issues, further developments will be focused on supplementary analysis of the frequency setting phase testing a completely new and opposite approach to define bus lines frequencies. Instead of the classic definition of an objective function stating the minimization of the overall travel time of the users (walking, on-board and waiting), under a fleet size constraint, the frequency setting should be approached as an optimization problem aiming at minimizing the fleet size under specific constraint about appropriate values of frequency, preserving an effective level of services for users.
References


