A transit network design model for urban areas

M. Petrelli
Dip. di Scienze dell’Ingegneria Civile, Università di Roma “Tre”, Italy

Abstract

Transit network design is discussed in an extensive amount of literature, with several optimization procedures to solve the design problem by determining transit routes and associated frequencies. In recent years, following the evolution of operational research and the increasing power of computational machines, studies have had a new impulse. The design model must become an important tool to develop an intensive, and not extensive, transit network of a new generation characterized by efficiency, integration between routes, and structured to improve service quality and riderships.

The purpose of this paper is to develop a new heuristic bus design model for urban areas, formulated as an optimization problem (minimization of the overall system costs). The basic framework of the model is established on three phases: 1) a heuristic algorithm to generate a set of feasible routes; 2) a genetic algorithm to find the optimal sub-set of routes; 3) final improvement of the network configuration. The model output is the bus routes, the frequencies and the vehicle sizes.

This paper presents a synthetic state-of-the-art proposal about the bus network design problem, describes the developed model and presents some final considerations about the solution approach proposed, and possible further developments.

Keywords: transit network design, genetic algorithms, route development.

1 Introduction

Transit network design is the object of extensive literature. In the past, different approaches were adopted. Optimization procedures resulted in the best and most useful approach to solve the network design problem by determining transit routes and associated frequencies. This approach, which will be followed in this paper, is concerned with the minimization of an objective function (FO), which
generally includes all different possible impacts of transport (but is usually limited to a combination of operator and user’s costs).

The transit network design is a non-convex problem (Newell [11]) and solution methods are based on heuristic procedures. It is also important to underline that design procedure should explicitly involve relationships between different elements of the transit planning process and, in particular, route development and frequencies setting should be solved simultaneously (Ceder and Wilson [4]). Among the most remarkable works on this matter, it is possible to mention Mandl [10], who proposed an algorithm that at each step individuates a route on the shortest path that connects a pair of terminals and serves the greatest number of O/D pairs. Filippi and Gori [7] developed an integrated model system for solving the transit network design problem, where routes generated are joined at main nodes or transit centers. The most innovative feature of this model concerns the route design procedure, which is carried out by means of a system of integrated models, facing problems ranging from low demand to high-density areas specific problems. Ceder and Israeli [5] proposed a public transport design system based on the mathematical programming approaches. The first step generates a very large set of feasible routes connecting at every node to all others. Then, the system creates the minimal subsets of routes solving a Set Covering Problem and, using a multi-objective analysis, it is possible to select the most suitable subset. Baaj and Mahmassani [2] used an Artificial-Intelligent heuristic algorithm for route generation, which searches the transit demand matrix for high-demand node pairs and selects them as seeds of skeletons. The skeletons are expanded to routes according to a node selection strategy reflecting different trade-off between performance measures, users’ and operators’ costs.

In the last few years, researchers have tested new approaches based on artificial neural networks or genetic algorithms (GAs) to solve optimization problems. Xiong and Schneider [13] show that GA may efficiently solve transport network design problems. This model uses an improved version of GA in which good solutions are always saved and used together with new population for the reproduction. Other authors developed GA applications to solve specific transit network design problems (routing, scheduling, etc.). Pattnaik et al [12] implemented a two phases procedure. First, a set of feasible routes is generated and then a GA selects the optimum routes set. Different typologies of GA are implemented by adopting a fixed or variable string length coding scheme that represents the number of routes of the network. Also Dhingra et al [6] proposed an application of GAs. The solution approach provides a double application of GA to solve separately, the routing and the scheduling problem.

2 Problem formulation

The main objective of the model is to identify a bus transit network configuration that achieves the minimization of the overall system costs (operator and user’s costs). The model is developed for medium-large urban areas, characterized by complex network topology, not obviously describing as simple radial or grid network, where the public transport system is usually multimodal and the main component of the service is carried out by bus. The demand structure results in
the “many-to-many” type, with origin-to-destination (O/D) trips of elevated entity. The heuristic model proposed assumes as given, the O/D matrix demand, the road network characteristics, as well as the operating and users costs.

The problem may be formulated as an optimization problem, introducing a FO where are measured resources and impacts related to different interested subjects (operator and users). The proposed formulation considers total bus travel distance and total bus travel time, to find the variable operator costs and the users disutility, determined by in-vehicle travel time, waiting time and the transfer penalty.

The FO can be expressed as follows:

\[ FO = C_1 \cdot VKM + C_2 \cdot VH + C_3 \cdot (TV + w_1 \cdot TW + w_2 \cdot NT) \]

where:

- \( VKM = \sum_{i \in I} f_i \) is the total bus travel distance; 
- \( VH = \sum_{i \in I} \sum_{hk \in I_{kj}} t_{hk,i} \cdot f_i \) is the total bus travel time; 
- \( TV = \sum_{i \in I} \sum_{hk \in I_{kj}} t_{p_{hk,i}} \cdot p_{hk,i} \) is the user’s total in-vehicle travel time; 
- \( TW = \sum_{i \in I} \sum_{hk \in I_{kj}} tw_{hk,i} \cdot p_{whk,i} \) is the user’s total waiting time; 
- \( NT = \sum_{n \in n} n_{nt} \) is the user’s total number of transfers.

The notations used above are so defined:

- \( L_i \) shows the length of line \( i \) (km); \( f_i \) is the frequency of line \( i \) (bus/h); \( I_i, I_{w,j}, I_n \) represent respectively the set of the network lines, the links of line \( i \) and the nodes of the transit network; \( t_{p_{hk,i}}, tw_{hk,i} \) indicate respectively the travel time and the waiting time for the link \( hk \) of line \( i \); \( p_{hk,i}, p_{whk,i} \) show respectively the in-vehicle and the boarding passengers on link \( hk \) of line \( i \); \( n_{nt} \) is the number of transfers carried out in node \( n \); \( C_1 \) = factor to convert bus kilometres to cost equivalent; \( C_2 \) = factor to convert vehicle-hours to cost equivalent; \( C_3 \) = factor to convert user travel time to user travel cost; \( w_1 \) = weight for the waiting time; \( w_2 \) = weight for transfer penalties.

The model outputs are bus routes, associated frequencies and vehicle sizes. The solution must satisfy the following constrains: minimum percentage of demand satisfied; maximum and minimum allowable frequency; maximum load factor. Route terminals are not defined as input data but are determined by the model.

3 Solution approach

The basic framework of the model is established on the following three phases:

1) a heuristic algorithm to generate, using different criteria and rules, a large and rational set of feasible routes;
2) a GA to find the optimal sub-set of routes with associated frequencies and vehicle sizes;
3) final improvement of the network configuration.

It is important to underline that heuristic procedures, in which are used deterministic rules and strategies to find the optimal solution, have a limited application field in a very small network because these procedures explore a very modest number of possible solutions. Therefore, the utilization of powerful and probabilistic optimization techniques as GAs become particularly convenient for the large urban areas bus network design, are necessary. The particular transport system characteristics ("many-to-many" trips demand and complex network topology) affect complexity problem in an extremely important way from the computational point of view. Moreover, the use of GAs is absolutely necessary for the discrete nature of several variables and for the non linearity of the objective function.

3.1 Heuristic route generation algorithm

In the first phase, the model generates three different and complementary sets of feasible routes. The aim is to develop a consistent, diversified and exhaustive set of routes built, with efficiency according to different criteria. First set (A-type routes) is composed by the shortest paths between the node pairs with demand for trips larger than a given minimum value. The second feasible set (B-type routes) is generated with the aim to develop routes of a hierarchical network composed by main and feeder lines. The third set (C-type route) is made up the routes of the existing transit network. The feasible routes set obtained should be the most useful and large as input for the GA, composed by good and rational routes, really suitable paths.

3.1.1 A-type routes generation

In this case, the route generation criteria are addressed to reduce transfers and to develop direct lines between terminals. The resulting routes should reproduce strictly the demand matrix and represents the best network for the users. The generation algorithm steps are illustrated in fig. 1. This set is composed by direct routes between O/D pairs along the connecting shortest path. Routes are generated for all O/D pairs in which trips exceeds a given threshold (Tmin).

![Figure 1: A-type routes generation.](image-url)
3.1.2 B-type routes generation

The second set of feasible routes is based on a different logic of construction and is composed by routes characterized as pertaining to a hierarchical network with main and feeder lines. For this set, the design strategies are addressed to reduce unused capacity on vehicles and to provide convenient one transfer service. The resulting routes should develop a more efficient network than A-type routes in respect to operator’s costs. From the user’s point of view, the routes should build a network of simple use and easy learning for the occasional travellers. The construction, starting by main routes, is founded on trips demand matrix and link flows analysis. Feeder routes are identified by relaxing the set of constraints and applying the same rules used for the main routes.

The steps to generate routes are reported as follows:

Step 1: identification of main routes destination terminals. Alternative strategies are possible: the selection of the main node attractors or the designer’s identification guided by existing network and real feasibility of terminal;
Step 2: All-or-Nothing (AoN) demand assignment;
Step 3: Construction of the main routes;
Step 4: Construction of the feeder routes.

The generation of the main and feeder routes (Step 3 and step 4) is illustrated by fig. 2. The construction of the routes starts, for each identified destination node, with the selection of the entering links that present a link volume greater than a given threshold (Vmin). The selected link represents the first part of the path. Then, the algorithm determines the tentative frequency of the route in progress. The frequency (fi) is calculated by following equation:

\[ f_i = \frac{V_{\text{max}}}{f_{\text{c max}} \cdot C_V} \]  

where: \( V_{\text{max}} \) represents the maximum volume segment recorded on the line i, \( f_{\text{c max}} \) is the maximum load factor and \( C_V \) is the vehicle capacity. The expansion of the route, inserting other links, is developed with the selection of the maximum volume entering link in the node just reached by the path. If constraints for the insertion link are complied, the link is added in the route and the process is repeated from reached node.

A link is inserted in the building route only if 1) the ratio between unused capacity and capacity is smaller than a given threshold; 2) every node of the route is strictly closer to the destination terminal. To increase the number of generated routes, link selection from reached node is not limited at the maximum volume link but is extended to all the entering links that comply the constraints for link insertion. If other links exist, the building procedure records these new routes and restarts from the selected links with the route expansion.

At the end of the generation of the main routes, the algorithm using the same rules and criteria, but with different and relaxed values for parameters and thresholds, starts to produce feeder routes. The feasible terminals nodes for the feeder routes is composed by all the pertaining nodes to the main routes.
3.1.3 Feasible routes set generation
At the end of A-type and B-type routes generation, all the routes generated and the existing network routes (C-type routes) are checked to determine if the constraints for routes are satisfied. In particular, the constraints are related to a maximum and minimum allowable route length and the presence of significant overlapping routes. The overlapping is verified by comparing each route with the others, starting from the longest one. The shortest routes are discarded if the not overlapping part of the path differs less than a minimum length value. If the constraints are complied, the routes are stored in the set of feasible routes as input data for the second phase.

3.2 Application of Genetic Algorithm

3.2.1 Overview of Genetic Algorithms
GA’s are search algorithms based on concepts of natural selection and natural genetics, and are used as general purpose optimization algorithms. They differ from other search methods because they search among a population of points simultaneously and work with a coding of parameters rather than the parameter values themselves. The transition scheme of genetic algorithm is probabilistic, whereas traditional methods use gradient information (Goldberg [8]). GA search the solution space of a function through the use of simulated evolution. In general, the “best” individuals of any population of solutions tend to reproduce and survive to the next generation, thus improving successive generations. GA explores all regions of the solution space through genetic operators (mutation,
crossover and selection) applied to individuals in the population. Genetic operators provide the search mechanism of the GA. Crossover takes two individuals and produces two new individuals while mutation alters one single individual. These particular aspects described above make this technique applicable in a very general way without the limitations imposed by other local search methods (i.e. continuity, derivative existence, presence of discrete variables etc.).

### 3.2.2 Implementation of GA for the bus network design

The selection of the sub-set of "optimal" routes is done using a GA. The set of feasible routes represents the input data for the GA. The GA selects the sub-set of routes supplying as output the network lines, relative vehicle sizes and frequencies. The variables of the optimization problem are the feasible routes and are identified by the route code. The representation scheme adopted is a float representation. In fact, recent studies have found that this coding scheme is superior to a binary representation scheme, working at an order of magnitude more efficient in terms of CPU time. This choice permits also a more consistent representation and a higher precision across reproductions (Houck et al. [9]). GA processes simultaneously a number of solutions (individuals) that forms a population; each solution is composed by a fixed number of variables (feasible routes). The GA used in this work, implemented in MATLAB and interfaced with EMME2 for the transportation analysis, is shown in the form of a pseudo code as follows:

- a) randomly initialize the population;
- b) evaluate each individual of the population by computing the FO value;
- c) evaluate fitness function value for each individual;
- d) generate new population using genetic operators;
- e) return to step b); the algorithm ends when the fixed iteration number is reached.

The GA is applied several times by using a different size of route set (number of variables) ranging between a minimum and a maximum route size for the network. The computation of FO value is done by estimating the different measures and performance reported in FO, at the end of an assignment and bus vehicle size and frequency setting procedure. In this case, the trip demand matrix is assigned using the hyperpaths technique, the more significant in a large urban area characterized by high frequency services and overlapping routes. The procedure adopted is proposed by Baaj and Mahmassani [1] and, sequentially, produce the demand assignment and, then the setting of vehicle size, and frequencies (fig. 3).

At the start of the procedure, demand assignment is done. Then, for each route, according to the maximum segment volume recorded, the vehicle size is defined and also the frequency is settled up according to the eqn (7). At this point an assignment is replaced. The process stops when the values of route frequency do not record relevant changes between two consecutive iterations. Before the first demand assignment, an initial set of vehicle size and frequency, for each route, is defined. The convergence of the procedure is not
guaranteed, but computational tests have always assured convergence in a reduced number of iterations.

At the start of the procedure, demand assignment is done. Then, for each route, according to the maximum segment volume recorded, the vehicle size is defined and also the frequency is settled up according to the eqn (7). At this point an assignment is replaced. The process stops when the values of route frequency do not record relevant changes between two consecutive iterations. Before the first demand assignment, an initial set of vehicle size and frequency, for each route, is defined. The convergence of the procedure is not guaranteed, but computational tests have always assured convergence in a reduced number of iterations.

The fitness function is derived by FO with the insertion of another single term representing a penalty associated to demand matrix unsatisfied. By means of this added term, the algorithm permits to directly consider the constraints related to a minimum percentage of demand to satisfy. It is so possible to select the optimum network configuration, automatically complying this important requirement. Weight to the penalty are given in terms of FO units.

### 3.3 Network final improvements

The final step foresees a systematic analysis of the obtained network. In order to improve the network configuration, suitable routes modifications are examined in terms of total demand served, network effectiveness and efficiency. For each line selected by GA, the impact is verified of the following actions: 1) route extension; 2) route shortening; 3) route expansion including other bordering nodes to the original path. The benefits evaluation of route changes is conducted at a network level by analyzing the variation of FO value. These final network
improvements, by means of simple operations before described, is placed, unlike other studies, at the end of the model after the selection of a sub-optimal network configuration. It is therefore preferred to set up, first, the selection process on a large number of rational routes characterized as pertaining to one network skeleton, and to postpone, in a successive moment, the search of route modifications of local sphere. In this way, it is possible to remarkably reduce the elaboration time required for these route changes because the search is only among a small set of selected lines and not among all the very large set of generated routes. It is important to underline that local modifications should not produce relevant modifications to network skeleton configuration.

4 Conclusion and further developments

In this paper, a solution methodology to design a bus transit network in the medium-large urban areas, in terms of route, vehicle size and frequency, is described. With respect to other models, the paper proposes an improvement to the design procedures by introducing different criteria for route generation simultaneously. This approach permits to take into account the urban areas structure and to generate a large number of feasible routes, providing a very good and useful input data for GA. This choice should also produce positive impacts on network performances for medium-large urban areas. The model output, a network composed by A-type and B-type routes, should provide not only a reasonable level of bus service, but should better satisfy different typologies of users and behaviours. By this point of view, it is important to remark that the bus network is generally designed according to the peak hour system characteristics and considering average traveller characteristics. But in a real network, the services have to satisfy several distinct family of users (for instance, young or old people, systematic or occasional travellers), different trip purposes and distinct trip time starting. The model provides a quite general and flexible framework for transit network design, which allows the designer’s exploiting his/her knowledge of the network and specify different particular transit network hypothesis, including the existing one, which will be included into the set of feasible routes and then processed by the algorithm in order to produce the best solution.

First applications of the model on two test networks, one sample and one real medium large network, show interesting and promising results. The set of generated routes are very large and complete and is composed by both route types. The final selected set is always a combination of the different route type, with a predominance of A-type routes. The implemented GA have proved to be robust and efficient and have produced reasonable, rational network configuration in terms of paths and number of lines. In the real network test, the model have proposed a network configuration better than existing one with a decrease of transfers and waiting users times.

The application of the model on a real large network represents an important step to further validate the procedure and to better investigate model flexibility and ability of working. Further developments will focus on the analysis of FO
components and parameters used by model, and on the improvement of GA efficiency.

In addition, there is also the need to enrich the model inserting other features. In particular, with the development of Intelligent Transportation Systems and telematics applications in public transport systems, effective design of transit routes and frequencies would take in account and, consequently, would exploit these new opportunities (in particular, the user information systems and the higher transit service reliability).

References


