Computers and Operations Research 80 (2017) 128–146

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Integrated approach to network design and frequency setting problem in railway rapid transit systems

Francisco López-Ramos[a,d,*], Esteve Codina[b], Ángel Marín[c], Armando Guarnaschelli[d]

[a] Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey Campus, Eugenio Garza Sada 2501, 64849 Monterrey, Mexico
[b] Universitat Politècnica de Catalunya, FIB, Campus Nord, Building C5, 08034 Barcelona, Spain
[c] Universidad Politécnica de Madrid, ETSI Aeronáuticos, Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain
[d] Pontificia Universidad Católica de Valparaíso, ITRA, 2147 Brasil Avenue, 2362804 Valparaíso, Chile

ARTICLE INFO

Article history:
Received 16 December 2015
Revised 6 October 2016
Accepted 7 December 2016
Available online 8 December 2016

Keywords:
Railway rapid transit
Network design
Working network
Frequency setting
Express service design
Lexicographic goal programming

ABSTRACT

This work presents an optimization-based approach to simultaneously solve the Network Design and the Frequency Setting phases on the context of railway rapid transit networks. The Network Design phase allows expanding existing networks as well as building new ones from scratch, considering infrastructure costs. In the Frequency Setting phase, local and/or express services are established considering transportation resources capacities and operation costs. Integrated approaches to these phases improve the transit planning process. Nevertheless, this integration is challenging both at modeling and computational effort to obtain solutions. In this work, a Lexicographic Goal Programming problem modeling this integration is introduced, together with a solving strategy. A solution to the problem is obtained by first applying a Corridor Generation Algorithm and then a Line Splitting Algorithm to deal with multiple line construction. Two case studies are used for validation, including the Seville and Santiago de Chile rapid transit networks. Detailed solution reports are shown and discussed. Conclusions and future research directions are given.

1. Introduction

Population growth in cities has led to traffic congestion. To alleviate this effect, transport agencies have designed Rapid Transit Systems. These systems are continuously revised to account for changes in passenger demand.

The overall transit planning process can be divided into the following phases [1,2]: Transit Network Design (TND), Transit Network Frequency Setting (TNFS), transit network timetabling, vehicle scheduling, and crew scheduling and rostering. Strategic and tactical decisions are taken in the first two phases. Because of the high costs of construction and exploitation of railway transit networks, it is important to optimize every strategic and tactical decision. The TND phase may build from scratch or expand the infrastructure of a rapid transit network (i.e., stations and stretches), considering budgetary restrictions and coverage demand satisfaction. Having determined the new infrastructure of the network, the TNFS sets the line frequencies and the number of vehicles needed to satisfy the passenger trip requirements at reasonable operative costs and not exceeding the capacities of the planning resources (i.e. the number of passengers per vehicle, the number of available vehicles, the maximum stretch frequency, and among others).

In current practice, TND and TNFS phases are solved separately as the infrastructure of the rapid transit network is considered as a stable component contrary to line frequencies which are treated as a flexible component [3]. However, it is admitted that to know how the infrastructure will be used by passengers, an assignment of passengers to lines is required. This assignment requires in turn the definition of lines and frequencies operating on them. Clearly, not considering lines and frequencies in the TND phase makes unrealistic estimates on passenger volumes at this early stage. Therefore, solving separately TND and TNFS phases is an approach that may lead the system to operate inefficiently because the new infrastructure of the rapid transit network is determined without considering the capacities of the planning resources. In other words, the expected amount of demand to be covered in the TND phase can be overestimated because when setting the line frequencies in the TNFS phase, it might not have enough vehicle capacity to fill that demand or the needed line frequencies exceed the stretch capacities. The need for new approaches integrating TND and TNFS phases is acknowledged in [2], and recently some

* Corresponding author.
E-mail address: francisco.lopez.r@itesm.mx (F. López-Ramos).

http://dx.doi.org/10.1016/j.cor.2016.12.006

works have explored this integration and showed promising results [4–6] improving substantially rapid transit systems.

Nevertheless, integrated models impose two great challenges. On the one hand, these models need to correctly represent the underlying structural properties of the TND and TNFS phases and allow a seamless integration. This means that models need to evaluate every trade-off regarding the decisions associated with each phase, and capture the impact of decisions in the TND phase on decisions in the TNFS phase, and vice-versa. On the other hand, the complexity of these models increases because of the integration. Consequently, customized solution algorithms and strategies are required.

In this work, the integration of TND and TNFS phases is achieved by means of an optimization model formulated as a mixed-integer linear goal-programming problem. Because of the size of real-world rapid transit networks, and the highly constrained set of design and frequency decisions, solving the model directly with an exact Branch & Bound optimizer, such as CPLEX, is impractical. Meta-heuristic approaches such as Genetic Algorithms (GA), Simulated Annealing (SA) and Tabu Search (TS), among others, have been used to provide a set of feasible solutions in reasonable time [5–9]. These methods are enough flexible to be adapted to a wide range of optimization models with different objective functions and constraints. However, this flexibility does not allow taking advantage of the mathematical structure of the model to explore the solution space in a way that complexity is minimized. In [7], the performances of GA, SA, TS, among other meta-heuristics, is examined showing that the solving times and the number of algorithmic steps are still strongly affected by the combinatorial nature of problems solving TND & TNFS phases, especially in the number of new lines to be constructed and the enumeration of all possible feasible line traces (corridors).

To deal with these complexity issues, a 2-stage methodology is proposed. The first stage employs a Corridor Generation Algorithm (CGA), whereas the second one uses a matheuristic (mathematical programming based heuristic) called the Line Splitting Algorithm (LSA). The CGA determines the pool of line traces that can be assigned to the new lines during the optimization process. The size of this pool is limited by infrastructure budgetary restrictions together with some user behavior rules which are verified in an ad hoc implementation of the Yen's k-shortest path algorithm [10].

The LSA solves as much instances of the optimization model as the number of lines under construction. Each instance builds only one line but determines the frequencies of all lines whose layouts (i.e. stretches and stations) are known (operating + already constructed + new constructed line). To this end, the constructed line in a given instance is transformed into a fictitious operating line for the next instance resolution. New lines are defined over the set of corridors built by the CGA. The instance optimization could be achieved with the aid of a solver based on either meta-heuristics or mathematical programming techniques. In this work, the latter approach is used under the premise that information about the quality of the solutions of the specific instances is available through the integrality gap metric of the Branch & Bound, and that the solving time is reasonable. In that manner, a quality standard for the instance solutions achieved by the LSA is guaranteed.

In each instance of the LSA, two goals are sought-after, minimize passenger riding time and minimize operator costs. To optimize these goals, goal programming [11] is used. This technique comes from multicriteria decision making and allows managing different and possibly conflicting objectives denominated goals in the context of an optimization problem. Current trends in urban planning emphasize the need for car-free cities [12,13], a concept that can only be realized by providing good level of service to passengers in public transport systems. Following this trend, this work assumes a strictly preference for minimizing passenger riding time over minimizing operator costs. In the context of well-defined preferences, the use of Lexicographic Goal Programming is recommended [14], where goals are optimized sequentially according to their priority levels defined by the decision maker preferences. This technique has recently been used in some public transport works [15,16].

Two case studies based on the rapid transit networks of Santiago de Chile and Seville (Spain) are used for validation. Detailed solution reports are shown and discussed. The paper is organized as follows. Section 2 surveys the literature on TND and TNFS, highlighting the contributions of the proposed approach. Section 3 states the problem. Section 4 describes the modelling strategy. Section 5 introduces the formulation of the optimization model. Section 6 presents the solving strategy. Section 7 shows the computational tests. The paper finishes with some conclusions and directions for further research in Section 8.

2. Literature review

Bruno et al. [17] is one of the first to tackle the TND phase. Their work aims to maximize demand coverage in a public network. Laporte et al. [18] incorporate demand using an origin-destination trip matrix. The papers of Laporte et al. [19] and Hamacher et al. [20] address the stations location problem on a given alignment. García and Marín [21,22] study the mode interchange and parking network design problem using Bilevel Programming. They consider multimodal traffic allocation problems using combined modes at the lower level of the bilevel program. Laporte et al. [23] extend the previous models by incorporating into the station location problem the possibility to include the construction of several lines. The resulting model also considers budgetary constraints; however, line terminal nodes are fixed. Marín [24] overcomes this limitation.

References related to TNFS may be classified into the ones that consider the point of view of the operator and the ones that account for the point of view of the user. In the first group, Claessens [25] consider the minimization of the service costs. They formulate a model accounting for the selection of services, frequencies and train lengths, considering vehicle types. Cordeau et al. [26] also consider different types of train compositions. From the point of view of the user, Bussieck et al. [27] maximize the number of passengers without considering transfers. Scholl [28] minimizes the number of transfers by using passenger routes in the model and heuristics to generate feasible solutions. Schöbel and Scholl [29] minimize travel time together with route selection from a predefined pool.

Works integrating TND and TNFS phases may be grouped according to the type of public transportation system to be planned. The vast majority of works deal with bus network design, whereas fewer works faces to the railway network planning. In the first group, Newell [30] studies the optimal geometries of bus routes depending on the demand trip distribution. The author proposes close formulas for computing desired frequencies depending on the choice of route geometries, and shows that the passenger cost function needed for route evaluation is not convex, therefore only local optimum can be found. Ceder & Wilson [1] formulate two sequential mathematical models for solving the bus network design problem. The first one minimizes the excess passenger travel time upon boarding, expressed as the deviation time from the shortest travel path, plus the transfer time (if any). The second model adds the passenger waiting time and vehicle operating and capital costs to the objective function. Baaj & Mahmassani [31] decompose the bus network design problem into three phases: a route generation step, in which routes and frequencies are constructed; a network analysis procedure, defining measures of effectiveness at the network-, route-, and stop-level; and a route improvement