An integrated train scheduling and infrastructure development model in railway networks

M. Shakibayifar\textsuperscript{a}, E. Hassannayebi\textsuperscript{b},\textsuperscript{*}, H. Mirzahossein\textsuperscript{c}, Sh. Zohrubnia\textsuperscript{d} and A. Shahab\textsuperscript{e}

\textsuperscript{a}. Department of Transportation Engineering and Planning, Iran University of Science & Technology, Tehran, Iran.
\textsuperscript{b}. Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran.
\textsuperscript{c}. Department of Civil Engineering, Faculty of Engineering, Imam Khomeini International University (IKIU), 34149, Qazvin, Iran.
\textsuperscript{d}. Faculty of Management and Accounting, Allameh Tabataba’i University (ATU), Tehran, Iran.
\textsuperscript{e}. School of Industrial Engineering, Islamic Azad University, Tehran South Branch, Tehran, Iran.

Received 7 April 2016; received in revised form 28 August 2016; accepted 1 October 2016

Keywords
Train scheduling; Railway infrastructure development; Capacity expansion; Network decomposition; Conflict resolution.

Abstract. The evaluation of the railway infrastructure capacity is an important task of railway companies. The goal is to find the best infrastructure development plan for scheduling new train services. The question addressed by this present study is how the existing railway infrastructure can be upgraded to decrease the total delay of existing and new trains with minimum cost. To answer this question, a mixed-integer programming formulation is extended for the integrated train scheduling and infrastructure development problem. The train timetabling model deals with the optimum schedule of trains on a railway network and determines the best stop locations for both the technical and religious services. We developed two heuristics based on variable fixing strategies to reduce the complexity of the problem. To evaluate the effect of railway infrastructure development on the schedule of new trains, a sequential decomposition is adopted for Iranian railway network. The outcomes of the empirical analysis performed in this study allow for gaining beneficial insights of identifying the bottleneck corridors. The result of the proposed methodology shows that it can significantly decrease the total delay of new trains with the most emphasis on the bottleneck sections.

© 2017 Sharif University of Technology. All rights reserved.

1. Introduction

The passengers’ demand on railway transportation is expected to increase substantially in the future. Hence, the railway network capacity has to be improved to handle this future demand. Constructing or upgrading railway infrastructure is an option for increasing the capacity. The railway capacity varies according to different factors, e.g. train heterogeneity, train speed, stop patterns, infrastructure layout, and schedule robustness [1]. Train stop time defines the amount of time trains spend stopped at a location. The stopping pattern is a decisive factor to determine the operational capacity of the network. In Iranian railway network, trains stop at intermediate stations for load/unloading passenger, technical services as well as praying services. The last factor is a specific religious related constraint that significantly affects the train timetables. Based
on religious regulation, during predefined praying time windows, train must stop to perform praying services. Passengers get off the train and pray in the station mosque during the time window. However, the praying time windows change over time and location due to the local geographic position. Thus, the optimal pattern of stop for praying service has direct influence on the capacity of the Iranian railway system. The optimization of the railway lines’ capacity plays an important role in railway transportation industry. The effective utilization of a railway network results in the avoidance of resource conflicts, and, at the same time, finding an appropriate balance between capacity utilization and level of service [2]. This paper is concerned with Railway Network Infrastructure Development Problem (called RNIDP). The proposed approach combines the scheduling of new trains and rescheduling of existing trains by choosing the best railway infrastructure development scenario for decreasing train delays. The focus of this work is on the combinatorial analysis of infrastructure configuration optimization and timetable generation. In other words, our work addresses the railway infrastructure planning and train timetabling problem in one integrated model. To the best of our knowledge, this approach has not been extensively investigated in literature, and if so, not to the equivalent scope to which it is taken in this study. In this paper, RNIDP is formulated as a Mixed-Integer Linear Programming (MILP) model. The key contribution of the study is the integrated investigation of the train timetabling and infrastructure development problems. The previous studies in the literature present an independent study of train timetabling and capacity planning problems. To deal with such a complex optimization problem, a novel network decomposition approach is presented. Furthermore, in order to solve the resulting sub-problem efficiently, a fast heuristic based on variable fixing strategy is proposed.

The rest of the paper is as follows. In Section 2, a literature review of train timetabling problem and related issues is discussed. In Section 3, the problem is described in detail. In Section 4, the mathematical programming formulation for RNIDP is described. In Section 5, heuristic algorithms are presented in detail. In Section 6, numerical analysis and the results are described. Finally, the discussion and conclusion are stated in Sections 7 and 8, respectively.

2. Literature review

Train timetabling is one of the most interesting problems in transportation planning systems. Train timetabling problem basically consists of determining the train’s departure and arrival times in each station. The assessment of train timetable is a complex procedure due to the fact that it is subjected to the capacity and resource constraints [3]. A majority of the approaches to train timetabling are analytical methods. Moreover, the application of flexible simulation approaches for railway systems has been recognized by previous studies [4]. Huisman et al. [5] provided an excellent state-of-the-art review of railway optimization problems. In the literature, numerous mathematical formulations are presented for train timetabling problem. Train timetabling problem is known to be NP-hard with respect to the number of conflicts in the schedule [6, 7]. Thus, it is difficult to determine the optimum solutions to industry-sized problems in a reasonable time, and this raises the need for efficient heuristic and meta-heuristic algorithms. In the related literature, several sophisticated search procedures are introduced, such as look-ahead search [8], backtracking search [9], and meta-heuristics algorithms [10-12]. D’Ariano et al. [13] addressed the real-time train scheduling problem in a railway network using an efficient Branch and Bound (B&B) algorithm. The empirical test cases make evident that the B&B algorithm outperforms the commonly-used dispatching in terms of both average and maximum secondary delays. A decision support system, called ROMA, was designed by D’Ariano [14] based on Alternative Graph (AG) techniques to cope with real-time train rescheduling problem with multiple delays. The aim was to improve punctuality index through better utilization of the railway infrastructure. The system was extended by Corman et al. [15] to present an innovative distributed approach to manage train operations more effectively in multi-area dispatching areas. The performance of the distributed approach was compared with the existing models in terms of computation time and reducing total delay. Hassannayebi and Kianifar [16] proposed three meta-heuristic algorithms based on Greedy Randomized Adaptive Search Procedure (GRASP) to find the near-to-optimal train timetable in double-track railway lines. The output results show the effectiveness of the proposed meta-heuristic algorithm in solving large-sized instances of the train timetabling problem.

In a vast majority of the previous studies, the train schedule is an input for railway capacity planning model. Some recent and related capacity planning literature that is associated with aspects of train scheduling problems is as follows: proposing a scheduling model for inserting additional trains into the existing time table without any railway infrastructure expansion plan. Sajedinejad et al. [17] designed a simulation-based decision framework called SIMARAIL for scheduling trains in railway networks. The proposed framework was used as a tool to assess the different train dispatching policies. Lai and Shih [18] developed a capacity planning procedure for evaluation of possible expansion scenarios and determine the optimal network investment plan to meet the future demand.
Hasannayebi et al. [19] developed an event-driven simulation model for train timetabling problem with consideration of train stops for praying services. Path re-linking algorithm was implemented to obtain good quality solutions. Gómez et al. [20] introduced the new indicator for dynamic infrastructure occupation assessment and analyzed infrastructure capacity under disturbances. The capacity assessment study was performed on Dutch railway corridor with different signalling systems under both scheduled and perturbed traffic circumstances. For scheduled traffic, the UIC standard was used to compute infrastructure capacity, whereas Monte Carlo simulation technique was employed to evaluate rail capacity under disrupted conditions. Schöbel et al. [21] proposed an optimization model for integrated timetable and infrastructure design of a railway system. The problem was formulated as a combinatorial optimization problem and solved using mathematical programming techniques and metaheuristics. Hasannayebi et al. [22] proposed a simulation-based optimization approach for train timetabling in urban rail. The problem was solved using genetic algorithm and the outcomes verify the effectiveness of the solution method.

The previous studies on capacity planning in railway are mostly dedicated to the maximization of the throughput without making new infrastructure investments. In the literature, two aspects of railway capacity planning problem exist. The first is the short-term planning of scheduling additional trains and the second is the long-term planning of railway infrastructure upgrading. Railway infrastructure development is strongly related to a famous railway optimization problem called railway saturation problem. In the railway saturation problem, the main objective is to insert a maximum number of additional trains in a predefined train set. A notable point is that the railway saturation problem does not deal directly with train scheduling problem. Deconde et al. [23] developed two heuristic approaches to evaluate the railway infrastructure capacity. These models do not consider cost-based objective function and also service level constraints. Shih et al. [24] proposed an optimization model for the optimal siding locations for single-track railway lines subject to the infrastructure and traffic characteristics. The experimental results proved that the optimal plan can maximize the return on investment and achieve the desired service level. Vansteenkoven et al. [25] proposed a method for adjusting an initial in case of planned temporary infrastructure unavailability. A trade-off was made between the level of service and the capacity. The proposed model incorporates the timetable robustness to rescheduling procedure. The developed procedure also aims to minimize the number of cancellations. Pouryoussef et al. [2] proposed a multi-objective linear programming model together with rail simulation tools to improve capacity utilization. The developed system has the capability of automatic resource conflict resolution and train scheduling tools for multiple-track corridors. The methodology was validated through comparison with RailSys simulation package. Hasannayebi et al. [26] proposed a robust multi-objective stochastic programming approach for train scheduling at rapid rail transit lines. The mathematical models were developed to minimize the expected waiting times and the cost of overcrowding. The effectiveness of the proposed model was validated through the application to Tehran underground railway network. Pöhle and Feil [27] conducted a research based on integration of train scheduling and infrastructure planning.

Corman et al. [28] addressed the integrated train scheduling and delay management in real-time railway traffic condition. The delay management model focuses on the effect of rescheduling actions on the quality of service. The result of empirical test experiments based on Dutch railway network confirms that good quality solutions can be obtained within a reasonable computation time. Samà et al. [10] applied meta-heuristic algorithms to the real-time train timetabling problem in complex and congested railway networks. The proposed methodology is initiated with generating a good quality solution via a Branch and Bound (B&B) algorithm. Subsequently, Variable Neighborhood Search (VNS) and Tabu search algorithms were implemented to improve the solution by re-routing of trains. The result indicated that the proposed algorithm outperforms the solutions attained by a commercial optimization package in terms of reduced computation times and search performance. Qì et al. [29] designed an integrated multi-track station layout design and train scheduling model on railway corridors. The optimization model decides the number of siding tracks or platforms within the budget constraints. The total construction cost and total train travel time are assumed as performance metrics. A b-level programming model was developed to solve a train scheduling model along with track assignment problem. The optimization models were solved almost efficiently by commercial software GAMS with CPLEX solver and local searching-based heuristic. As a concluding remark, a wide range of papers are devoted to the train timetabling problem in the last decades, but there is still lack of integrated models which combine infrastructure upgrading issues with timetable generation.

3. Problem description

The train timetabling problem deals with the optimum dispatching of trains on a railway network and determines the best station to stop for technical and service-based purposes including religious requests. In
the following sections, we describe the assumptions, notation, introduction of special constraints, and railway infrastructure terms in detail.

3.1. Assumptions

The problem is concerned with a railway network with a set of linked corridors. Railway network consists of single, double, and multiple-track routes. A train service is defined as a trip of a train that travels from its origin station to its destination station. Each train is assumed to have a pre-specified traveling route in the network. Furthermore, the free running times of trains at segments are assumed to be constant. Train services are determined by a set of planned and unplanned stops. There are three categories of trains’ stops:

1. Scheduled stops with fixed stop time;
2. Unscheduled stops with fixed stop time;
3. Unscheduled stops with variable stop time.

Train technical services are planned during the operation cycle, but religious services occur during dynamic and time-dependent intervals. Scheduled stops are permitted at any intermediate stations for any train. But, unscheduled stops are permitted in the limited set of stations which have the required facilities, e.g. mosque. We consider a track section between two adjacent stations. To prevent this scheduled delay, a minimum headway time needs to be considered. A minimum headway time is the minimum time difference between dispatching of two following trains that should be kept in order to satisfy traffic safety regulations.

3.2. Notation

The notations of indices and sets used in the mathematical model, the notations of parameters used in the mathematical model, and the decision variables used in the mathematical formulation are presented as follows:

**Indices:**

- \(j\) Train index
- \(k\) Segment index
- \(i\) Station index
- \(s\) Index of technical stop types \((s = 1, 2, 3, \ldots, S)\)
- \(r\) Index of praying service \((r = 1, 2, 3, \ldots, R)\)
- \(J_1\) Set of existing trains \(J_1 \in J\)
- \(J_2\) Set of new unscheduled trains \(J_2 \in J\)
- \(I_1\) Set of one-lane stations \(I_1 \in I\)
- \(I_2\) Set of two-lane stations \(I_2 \in I\)
- \(K_1\) Set of single track segments
- \(K_2\) Set of double track segments
- \(CO\) Set of corridors in the railway network

**Parameters:**

- \(\delta(j, k)\) Index of the \(k\)th traveling segment in the route of train \(j\)
- \(K_j\) Total number of segments on route of train \(j\)
- \(\theta_{jk}\) 1 if train \(j\) is an inbound train in segment \(k\), 0 otherwise
- \(\varphi_{ji, j_t, k}\) 1 if segment \(k\) exist in route of train \(j_i\) and \(j_t\)
- \(\mu_{ij}\) 1 if train \(j\) needs to have access to platform for loading/unloading passengers in station \(i\), 0 otherwise
- \(p_{ij}\) Free running time of train \(j\) at segment \(k\)
- \(dw_{ij}\) Minimum dwell time of train \(j\) in station \(i\)
- \(H_{ji, j_t, i}\) The minimum headway between arrival/departure time of train \(j_i\) and \(j_t\) at station \(i\)
- \(s_t\) Stopping time required for technical stop type \(s\)
- \(pl_{ij}\) Passenger load and unloading time for train \(j\) in station \(i\)
- \(sp_r\) The required time for praying type \(r\)
- \(ms_{ij}\) Scheduled stop time for train \(j\) at station \(i\)
- \(ed_{ij}\) Earliest departure time of train \(j\) from its origin station \((j \in J_1)\)
- \(ld_{ij}\) Latest departure time of train \(j\) from its origin station \((j \in J_2)\)
- \(sd_{ij}\) Scheduled departure time of train \(j\) from its origin station \((j \in J_1)\)
- \(\lambda_i\) 1 if the location of mosque facility is in the inbound lane at station \(i\), 0 otherwise \((i \in I_1)\)
- \(\gamma_{v(j, k)}\) 1 if there is an facility at \(k\)th traveling segment of train \(j\) for technical stop, 0 otherwise
- \(\alpha_i\) 1 if there is an overpass facility in station \(i\), 0 otherwise
- \(\beta_i\) 1 if there is a mosque facility in station \(i\), 0 otherwise
- \(\text{NL}_{i}\) Number of tracks which is not connected with any platform in station \(i\)
- \(N_{p_{i}}\) Number of tracks which is connected with a platform in station \(i\)
- \(L_{r_{i}}\) Lower bound of time window of praying service type \(r\) in station \(i\)
- \(U_{r_{i}}\) Upper bound of time window of praying service type \(r\) in station \(i\)
- \(t_{p}\) Minimum time to prepare for praying and performing praying service
C \quad \text{Desirable value for summation of trains traveling time}

M \quad \text{A large positive constant}

LC_i \quad \text{Construction cost of a new track in station } i

PC_i \quad \text{Construction cost of a new platform in station } i

\textbf{Decision variables:}

\begin{align*}
xd_{jk} & \quad \text{Departure time of train } j \text{ from beginning of the segment } k \\
xa_{jk} & \quad \text{Arrival time of train } j \text{ at the end of the segment } k \\
\delta_{ji,jk} & \quad 1 \text{ if train } j_1 \text{ is entered before train } j_2 \\
\varepsilon_{ji,jk} & \quad 1 \text{ if train } j_1 \text{ is entered before train } j_2 \text{ in the same direction on segment } k, 0 \text{ otherwise} \\
\alpha_{ji} & \quad 1 \text{ if train } j_1 \text{ is entered before train } j_2 \text{ in the opposite direction on segment } k, 0 \text{ otherwise} \\
y_{ij} & \quad 1 \text{ if train } j \text{ stopped for } \theta_1 \text{ at station } i, 0 \text{ otherwise} \\
w_{ij,t} & \quad 1 \text{ if train } j \text{ is assigned to } t \text{th track at station } i, 0 \text{ otherwise} \\
\omega_{jt} & \quad 1 \text{ if train } j \text{ occupies } t \text{th platform at station } i, 0 \text{ otherwise} \\
g_{ri,j} & \quad 1 \text{ if train } j \text{ stopped for } \theta_r \text{ at station } i, 0 \text{ otherwise} \\
A_{li} & \quad \text{Number of additional tracks (do not connected with platform) need to be constructed in station } i \\
A_{pi} & \quad \text{Number of additional tracks (connected with platform) need to be constructed in station } i
\end{align*}

Construction of new platform and track in the station is considered as infrastructure development options in the mathematical model. The optimization model determines which section of the railway network needs to be upgraded with what type of capacity improvement scenario. Construction of new lines and platforms in each station has a predefined cost which depends on the station infrastructure characteristics.

\subsection*{3.3. Religious-related constraints}

We considered several religious constraints in our model that appear in Iranian railway network. Train movements on the railway network in Iran are regulated by the three daily prayers. All trains should stop during the praying time window for a period of 20-25 minutes. Each praying service has a dynamic time window to be performed. This time window depends on geographic location of station and canonical time horizon. These time horizons consist of morning, noon, and sunset. We assume that each station has a local canonical time horizon. So, the start time of praying depends on the location of site and daily time horizon changes. Praying times change daily according to the seasons and train exact location in the railway network. Performing religious services for trains also depends on the dispatching time of the trains from their origin stations and the arrival time to destination station. Finding the best station for stopping to perform religious service is a challenging optimization problem added to train timetabling problem. Eqs. (1) and (2) state the condition which forces the trains to stop for praying service in the intermediate stations. We denote the dispatching time of train \( j \) from the origin station by \( xd_{j,\delta(j,i)} \) and arrival time of this train to destination station by \( xa_{j,\delta(j,K_j)} \). Every praying service type needs to be checked before determining the best station to stop. If the following conditions are satisfied, then train \( j \) should stop in one station (which have mosque facility) on its route for praying service type \( r \). The two conditions of praying service activation can be expressed as the following two equations:

\begin{align*}
xd_{j,\delta(j,i)} & \leq L_{r,v(j,i)} + t_p \\
\forall j & \in J, \quad \forall r = 1, 2, 3, \\
xd_{j,\delta(j,K_j)} & \geq U_{r,v(j,K_j+1)} - t_p \\
\forall j & \in J, \quad \forall r = 1, 2, 3,
\end{align*}

where \( (L_{ri}, U_{ri}) \) is the praying time interval in station \( i \), when trains are allowed to stop for performing praying service with type \( r \). Note that, in Eq. (1), \( L_{r,v(j,i)} \) is defined as a lower bound of praying interval for praying service \( r \) in origin station of train \( j \) on its route. Trains may have to stop for all praying services, or none of them.

\section*{4. Mathematical formulations}

In this section, we formulate the train timetabling problem as a mixed-integer linear programming. In the following, we state the details of the integrated mathematical model for generating an extended timetable according to the best railway infrastructure upgrading plan. The objective function is to minimize Eq. (3) subject to constraints given by Eqs. (4)-(20). The objective function is the minimization of railway infrastructure development cost:

\begin{align*}
\min Z = \sum_{i=1}^{II} \left( A_{li} \cdot LC_{i} + A_{pi} \cdot PC_{i} \right), \\
xd_{j,\delta(j,i)} & = sd_{j} \quad \forall j \in J_1, \\
ed_{j} & \leq xd_{j,\delta(j,i)} - ld_{j}, \quad \forall j \in J_2.
\end{align*}
\[ x_{a,j,k,k+1} = x_{d,j,k,k+1} + p_{j,k} \cdot h_{j,k} \cdot k, \]
\[
\forall j, k = 1, 2, \ldots, k_j, \tag{6}
\]
\[ x_{d,j,k+1,k} \geq x_{a,j,k,k+1} + d_{w,j,k+1,k} + m_{s_{w,j,k+1,k}} \]
\[
\forall j, k = 1, 2, \ldots, k_j, \quad i \in I, \tag{7}
\]
\[ m_{s_{w,j,k+1,k}} = \max \left\{ \| pt_{w,j,k+1,k} \| \sum_{s=1}^{n} s_{x,s,j,k+1,k} \cdot \right. \]
\[ \left. \sum_{r=1}^{n} s_{p,r,j,k+1,k} \cdot q_{r,w,j,k+1,k} \right\}, \quad i \in I_2, \tag{8}
\]
\[
x_{d,j,k} + p_{j,k} + H_{j,j,k,k} \leq x_{d,j,k} + (1 - z_{j,j,k,k}) \cdot M \]
\[
\forall j, z_{j,j,k,k} = 1, \quad \theta_{j,k} + \theta_{j,k} = \{ 0.2 \} \tag{9}
\]
\[
x_{d,j,k} + p_{j,k} + H_{j,j,k,k} \leq x_{d,j,k} + (1 - z_{j,j,k,k}) \cdot M \]
\[
\forall j, z_{j,j,k,k} = 1, \quad \theta_{j,k} + \theta_{j,k} = \{ 0.2 \} \tag{10}
\]
\[
x_{d,j,k} + p_{j,k} + H_{j,j,k,k} \leq x_{d,j,k} + (1 - u_{j,j,k,k}) \cdot M \]
\[
\forall j, u_{j,j,k,k} = 1, \quad \theta_{j,k} + \theta_{j,k} = \{ 0.2 \} \tag{11}
\]
\[
x_{d,j,k} - x_{d,j,k} \leq C \]
\[
\forall j \in J, k \in K, \tag{12}
\]
\[
\sum_{j=1}^{n} (x_{a,j,k,k+1} - x_{d,j,k+1,k}) \leq C, \tag{13}
\]
\[
\sum_{j=1}^{n} w_{i,j,t} \leq 1 + \alpha_i \quad \forall i \in I, \quad \forall t \in T, \tag{14}
\]
\[
\sum_{k=1}^{n} \beta_{v,k,j} \cdot g_{v,j,k+1,k} = 1, \quad \forall j \in J, \quad \forall r \in R. \tag{15}
\]
\[
L_{r,v,j,k+1,k} - M(1 - g_{r,j}) \leq x_{a,j,k,k+1} \leq U_{r,v,j,k+1,k}
\]
\[
+ M(1 - g_{r,j}), \quad \forall j \in J, \quad \forall k \in K, \quad \forall r \in R. \tag{16}
\]

Constraints (4) and (5) are related to the departure time constraints. Constraints (4) define the departure time of the train at the first segment. Constraints (5) ensure that the departure time of the train from the origin is within the allowed range. Eq. (6) states the running time constraints on segments. Eq. (7) defines the minimum station dwell time constraint. The train technical, passengers load/unload and religious services are parallel tasks which are performed in the stations. So, the maximum stopping time of a train in a station is the maximum time of all tasks above. The calculated stopping time in each station is distinguished in Constraints (8) and (9) according to the station characteristics. According to Eq. (8), the scheduled stop time for train j (running in the inbound direction) in station is the maximum time of all possible technical operations or religious services. Likewise, the scheduled stop time of trains at outbound route is formulated in Eq. (9). The minimum headway Constraints (10)-(12) also describe the minimum safety headway requirements between the departure times and arrival times of successive trains at the same segment. More specifically, Constraints (10) and (11) state the overtaking conflict for trains. Constraint (12) states the crossing conflict for inbound and outbound trains. Constraint (13) states that the total traveling time of trains should be less than a desired value. Constraint (14) ensures that the number of trains occupying a station in the same time is less than maximum capacity of the station. Eqs. (15) and (16) are related to the situation that trains stop for praying services. Constraints (15) ensure that only one station is assigned to each train to perform praying services. Constraints (16) state that the departure time of trains is between the allowed praying time windows. Constraint (17) defines the stopping pattern of trains for technical services. Trains are permitted to stop at a set of eligible stations for technical services. Constraint (18) states that a free platform should be allocated to a train which stops in the station. Constraint (19) ensures that the number of trains which needs free platform, at any given time, is less than the total number of existing and additional platforms in the stations. Eq. (20) states the logical constraint about occupation of a platform in the stations.
5. Solution methodology

In this section, we present two heuristics and a decomposition approach to reduce the complexity of the problem. The overview of the proposed decomposition-based heuristic algorithm to solve train timetabling problem in network is illustrated in Figure 1. Heuristic algorithms contribute to a decrease in the number of decision variables and constraints via variable fixing strategies. The decomposition approach is also used for dividing the complete optimization problem into subproblems. Each sub-problem is associated with a train timetabling model on a separate route.

5.1. Variable fixing strategies

Given the mathematical model presented in the previous sections, the problem turns into a large-scale mixed integer program. As is well known, this type of model is far more difficult to solve optimally. Here, to handle the above mixed-integer programming model with an exponential number of variables and constraints, we present a preprocessing algorithm for decreasing the number of active praying services. By applying this procedure, we can determine the condition of praying service activation before optimizing the mathematical model. By changing the dispatching time of a train between its earliest and latest departure times, the praying services may become active or inactive. The goal is to assign the trains into a set of definite and indefinite trains according to their praying service activation. The notations used to describe the heuristic procedures are presented as follows:

\( D_{r,j} \) Set of trains which their conditions for praying service type \( r \) are definitely active

\( ND_{r,j} \) Set of trains which their conditions for praying service type \( r \) are not definitely active

\( ea_j \) Earliest possible arrival time of train \( j \) to its destination station

\( AP_{r,j} \) 1 if the praying service type \( r \) is active for train \( j \), 0 otherwise

\( \gamma_{r,j} \) 1 if the first condition of praying service is true, 0 otherwise

\( \Phi_{r,j} \) 1 if the second condition of praying service is true, 0 otherwise

\( ea_j \) can be calculated as follows:

\[
\begin{align*}
    ea_j &= ed_j + \sum_{k=1}^{K_j+1} (p_{v_{i,j,k,j}} + dw_{v_{i,j,k,j}}) + \sum_{k=1}^{K_j} p_{j,i,k,j} \\
\forall j \in J.
\end{align*}
\]

(21)

The praying service activation conditions can be stated by parameters \( ea_j \) and \( ld_j \) in the following equations:

\[
\begin{align*}
    ld_j &\leq L_{r,v,i,j,k} + t_p, \\
\forall j \in J, \forall r \in R. \\
\end{align*}
\]

(22)

\[
\begin{align*}
    ea_j &\geq U_{r,v,i,j,k} - t_p, \\
\forall j \in J, \forall r \in R. \\
\end{align*}
\]

(23)

Next, we describe the main components of our preprocessing in Algorithm 1.

We proposed a mathematical optimization model which determines the departure time of trains from the origin station to minimize the number of active praying services. This algorithm is a class of heuristics which is based on constraint satisfaction techniques. In the following, we propose the mathematical model for the MAPS algorithm:

\[
\begin{align*}
    \min Z = & \sum_{j=1}^{n} \sum_{r=1}^{m} AP_{r,j}, \\
2 AP_{r,j} &\leq \gamma_{r,j} + \phi_{r,j} \quad \forall j \in J, \forall r \in R, \\
\gamma_{r,j} + \phi_{r,j} &\leq 1 + AP_{r,j} \quad \forall j \in J, \forall r \in R, \\
xd_{j,r,i,j} - L_{r,v,i,j,k} &- t_p \leq M \left(1 - \gamma_{r,j}\right), \\
\forall j \in J, \forall r \in R, \\
xd_{j,r,i,j} - L_{r,v,i,j,k} - t_p &> M \gamma_{r,j}, \\
\forall j \in J, \forall r \in R, \\
\end{align*}
\]

(24)

(25)

(26)

(27)

(28)
Step 1: (Initialization)
Set \( D_{ij} = \phi \), \( ND_{ij} = \phi \). Let \( r = 1, j = 1 \).

Step 2: (Checking the praying service conditions)
Calculate the following praying service conditions for train \( j \) for the praying service type \( r \):
\[ \forall j \in J, \forall r \in R, \quad x_{a_{r_j}, b_{r_j}, j, k, j, k, j, k} \leq U_{j, v_{j, j} + 1} + t_p \geq M.(\phi_{r_j} - 1), \]
(29)
\[ x_{a_{r_j}, b_{r_j}, j, k, j, k, j, k} \leq U_{j, v_{j, j} + 1} + t_p \geq M.\phi_{r_j}, \]
(30)
\[ x_{a_{r_j}, b_{r_j}, j, k, j, k, j, k} = x_{d_{r_j}, b_{r_j}, j, k, j, k, j, k} + \sum_{k=1}^{K_j+1} \left( p_{j, b_{r_j}, j, k, j} + d_{j, v_{j, j} + 1} \right) \]
\[ + \sum_{k=1}^{K_j} p_{j, b_{r_j}, j, k, j} \quad \forall j \in J, \forall r \in R, \]
(31)
\[ \phi_{r_j} = (0, 1), \quad \gamma_{r_j} = (0, 1). \]
(32)

In the formulation, Objective (24) defines the number of total active praying services. Constraints (25) and (26) ensure that if the two conditions of praying services are true, then the praying service will be active. Constraints (27) and (28) define the relation between the first condition of praying service activation with auxiliary variables. Similarly, Constraints (29) and (30) define the relation between the second condition of praying service activation with auxiliary variables. Constraints (31) define the trains traveling time.

5.2. Physical railway network decomposition
It should be noted that large-scale train scheduling problems with hundreds of trains, which move in different routes along hundreds of stations, are far more difficult to solve exactly. Based on a physical decomposition of the railway network in sub-networks, the problem is divided into sub-problems corresponding to these sub-networks. These sub-problems are then solved and the whole process is coordinated at a higher level in order to generate a global feasible solution. This decomposition procedure is performed in a hierarchy framework. The proposed decomposition approach consists of corridor-based decomposition, train route-based decomposition, and double-track segments decomposition. In this framework, computational effort for searching the infrastructure upgrading options will decrease.

In order to account for the evaluation of different capacity expansion scenarios, we propose a novel heuristic algorithm, which decomposes the physical network into bottleneck and underutilized corridors. The definition of bottlenecks can be based on the analysis of infrastructure, technical definitions, operational systems, economical, spatial and social context, and travel demand forecasts. Normally, the traffic density through the bottleneck sections is expected to be very high. We define the bottleneck corridor as the demand for transport exceeding the available capacity of infrastructure. The underutilized corridors are the other railway sections with lower demand. Bottlenecks in railway transport system usually are concerned with: (1) the level of transport speed (maximum travel speed, train travel time, passenger waiting time); and (2) the capacity (maximum number of operating trains per railway section). The analysis to identify the existing bottlenecks corridors come up with the necessary measures. On the main stream of bottleneck analysis for the existing railway infrastructure is the evaluation of the train timetables to calculate the average delay time of trains. The main idea of this heuristic algorithm is decomposition of a railway network based on our definition of the level of railway capacity utilization. Different policies for generating train timetables are then applied to the two types of corridors according to their properties.

The scheduling strategy initiates by solving the timetabling problem of the bottleneck corridors and then using the result of the solved problem to generate timetable for other underutilized corridors. In bottleneck corridors, the track topology usually consists of single-track segments, hence there is a huge potential for railway infrastructure development. In the infrastructure development problem, we test different scenarios to decrease the traffic density of the bottleneck corridors. In order to clarify the definition of the bottleneck corridors, we used the terms bottleneck section and bottleneck station to quantify the traffic
density on the railway resources. The bottleneck section and bottleneck station are defined by a percentage of potential traffic density. A saturation criterion is used to quantify the traffic density of a railway resource. Delorme et al. [23] defined the capacity of a segment of a rail system as the maximum number of trains that can be scheduled on it within a certain unit of time (e.g., an hour). This concept is extended here to determine the traffic density ratio of the sections. The following notations are used for describing the saturation criteria:

\[ K_c \] Set of segments exist in corridor \( c \)
\[ I_c \] Set of stations exist in corridor \( c \)
\[ \psi_k \] The minimum traffic density ratio of track segment \( k \)
\[ \pi_i \] The minimum traffic density ratio of station \( i \)
\[ \Delta_c \] The minimum utilization ratio of a corridor \( c \)
\[ LD_{jk} \] Latest departure time of train \( j \) from the beginning of the segment \( k \)
\[ EA_{jk} \] Earliest arrival time of train \( j \) at the end of the segment \( k \)
\[ ea_{ij} \] Latest departure time of train \( j \) from station \( i \)
\[ ld_{ij} \] Earliest arrival time of train \( j \) at station \( i \)

The theoretical expressions of the capacity of a section and station noted \( \psi_k \) and \( \pi_i \), respectively, can be defined as:

\[
\psi_k = \frac{\sum_{j=1}^{||I||} (EA_{jk} - LD_{jk})}{T} \quad \forall k \in K_c, \tag{33}
\]

\[
\pi_i = \frac{\sum_{j=1}^{||I||} (ea_{ij} - ld_{ij})}{T} \quad \forall i \in I_c, \tag{34}
\]

where \( T \) is the planning horizon. Now, we can calculate the weighted average of the station and section saturation ratios as utilization ratio of a railway corridor as follows:

\[
\Delta_c = \frac{||I_c|| \sum_{i \in I_c} \pi_i + \sum_{k \in K_c} \psi_k}{||I_c|| + \sum_{k \in K_c} \psi_k} \quad \forall c \in CO. \tag{35}
\]

Therefore, we can divide the set of corridors into the bottleneck corridors with the highest ratio of utilization and underutilized corridors. This separation is according to a predefined threshold ratio of utilization (i.e., 30%).

6. Computational results

In this section, some examples taken from the Iranian railway network are used to illustrate the proposed methods. Transport system of the Iranian railway is growing fast now (Figure 2). The length of railway as well as expressway network have doubled during the last 15 years. The economic aspects of this expansion plan have not been yet investigated extensively by the transport experts; this is the main motivation of the preset study. By 2020, the Iranian railway expects to reach every regional capital by rail. Nevertheless, the demand is much higher than the available trains, as new rolling stock is always one step behind the opening of new lines and the tracks are shared with heavy freight traffic.

6.1. Case data

The decomposed mathematical formulation has been applied to the whole railway network which consists of long railway corridors. The example includes 132 trains and 56 stations and the network is composed of 30 double-tracked and 25 single-tracked segments (Figure 3). The Tabriz-Mashhad corridor is a double-tracked line with segments separated by stations, where a mean of 90 daily long distance trains run from Tabriz to Mashhad, or vice versa. Because of special infrastructure characteristics of the railway network in Iran, there is a significant potential for increasing the railway capacity by upgrading railway infrastructure. The information of the existing railway network is presented in Table 1.

6.2. Numerical analysis

To demonstrate the capabilities of the approaches developed in this paper, a real case study is solved. The results of preprocessing algorithm are summarized for different train services. The output of the algorithm is the categorization of trains according to their praying service activation status. In this case, there are three daily praying services and each praying service could be definitely active or not. The trains with definite praying service status can help reduce the number
of decision variables and constraints in mathematical model. The total number of decision variables decreases 13% after implementing these algorithms in a mathematical model.

The purpose of decomposition approach is to evaluate how much the average delay can be decreased by the suggested decomposition approach and compare the results to the outcome if the infrastructure stays unchanged. In this example, the number of the existing trains in circulation is 100 before adding the additional train services. We plan to schedule 32 new trains to the railway network in the planning horizon to satisfy the constraints given in Section 4 by choosing the minimum cost infrastructure upgrading options. The result of the decomposition algorithm is shown in Table 2.

Different capacity expansion scenarios have been generated in the decomposed optimization models to increase the utilization of the railway capacity. The result of each sub-problem and the average delay of the existing and new trains are summarized in Table 3. Each of the sub-problems generated from the decomposition approach was solved by using GAMS/CPLEX 12.0 on a PC equipped with a 3.3 GHz Pentium IV processor. Regarding the computation time, the decomposed sub-problems are terminated in less than 12 min for all instances. However, it should be noted that this was due mainly to the use of the proposed decomposition method. According to the obtained result given in Table 3, the existing number of tracks and platform has been expanded, on average, by about 32% and 46%, respectively. The average delay of train in the whole network after upgrading the infrastructure is about 6.9 minutes. Furthermore, the total cost of upgrading the existing infrastructure is about 511000 (§). The objective value also shows a strong relation between the construction of new platforms
and decreasing the average delay resulted from the combination of the infrastructure development options. Also, the obtained results indicate that adding a new platform in the station is more effective than adding a new track in terms of decreasing the average train delay.

The decomposition method obtains a better objective value for the instances which belong to the bottleneck corridors rather than to those in the underutilized corridors (Table 4). However, the average computation time of the decomposition method for bottleneck instances is mostly obtained from the decomposition method for underutilized instances.

There is a need for upgrading railway infrastructure when the number of train increases. The decomposition approach helps find the minimum cost scenario for developing infrastructure and decreasing delay systematically. As stated before, the proposed

### Table 2. Physical decomposition of the railway network.

<table>
<thead>
<tr>
<th>Corridor name</th>
<th>Minimum utilization ratio ( (\Delta_0) )</th>
<th>Saturation status</th>
<th>Number of generated sub-problems</th>
<th>Number of train sub-route</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tehran-Mashhad</td>
<td>14%</td>
<td>U(^{b})</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Tehran-Tabriz</td>
<td>40%</td>
<td>B(^{c})</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Tehran-Alvaz</td>
<td>35%</td>
<td>B</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Tehran-Sarri</td>
<td>11%</td>
<td>U</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

\(^{a}\)Threshold ratio of utilization = 30%; \(^{b}\)U: underutilize; \(^{c}\)B: bottleneck.

### Table 3. Size, specification, and optimal solution of the generated sub-problems.

<table>
<thead>
<tr>
<th>Sub-problem</th>
<th>Infrastructure before upgrading</th>
<th>Infrastructure after upgrading</th>
<th>%Increase of NCT</th>
<th>%Increase of NCP</th>
<th>Infrastructure development cost ($</th>
<th>Average delay (min)</th>
<th>Computation time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>42.86%</td>
<td>44.44%</td>
<td>54000</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>33.33%</td>
<td>30.00%</td>
<td>24000</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>22.22%</td>
<td>42.86%</td>
<td>12000</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>40.00%</td>
<td>50.00%</td>
<td>42000</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>20.83%</td>
<td>125.00%</td>
<td>10000</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>26.67%</td>
<td>66.67%</td>
<td>20000</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>57.14%</td>
<td>28.57%</td>
<td>48000</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>14.29%</td>
<td>9.09%</td>
<td>28000</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>37.50%</td>
<td>40.00%</td>
<td>58000</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>11.11%</td>
<td>25.00%</td>
<td>70000</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>40.00%</td>
<td>33.33%</td>
<td>13000</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>25.00%</td>
<td>30.00%</td>
<td>20000</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>60.00%</td>
<td>50.00%</td>
<td>45000</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>25.00%</td>
<td>44.44%</td>
<td>30000</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>33.33%</td>
<td>100.00%</td>
<td>13000</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>20.00%</td>
<td>25.00%</td>
<td>25000</td>
</tr>
</tbody>
</table>

\(^{a}\)NT: The number of scheduled train; \(^{b}\)NS: The number of existing stations; \(^{c}\)NCT: The number of new tracks; \(^{d}\)NCP: The number of new platforms.

### Table 4. Improvement of the total delay by development of the bottleneck corridors.

<table>
<thead>
<tr>
<th>Corridor name</th>
<th>Saturation status</th>
<th>Improvement in total delay (%)</th>
<th>Average computation time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tehran-Mashhad</td>
<td>U</td>
<td>13</td>
<td>234</td>
</tr>
<tr>
<td>Tehran-Tabriz</td>
<td>B</td>
<td>47</td>
<td>650</td>
</tr>
<tr>
<td>Tehran-Alvaz</td>
<td>B</td>
<td>54</td>
<td>925</td>
</tr>
<tr>
<td>Tehran-Sarri</td>
<td>U</td>
<td>11</td>
<td>353</td>
</tr>
</tbody>
</table>


decomposition method identifies the potential infrastructure alternatives and chooses the near-optimum capacity expansion plan. Figure 4 shows how the average delays increase exponentially when the number of trains exceeds the saturation level. The average delay is computed as the ratio of the total train delay at destination to the total number of trains. This value is computed for each test problem and the associated best found solution. In order to obtain each point in this graph, we solve the optimization model to find the best capacity upgrading plan. Also, the timetabling model is solved for the case when no capacity expansion decisions are allowed. Then, the best found solutions are considered and their performances are compared in terms of average delay of trains at destinations.

7. Discussion

The question addressed by the present study is how the railway infrastructure can be upgraded to decrease the total delay of the existing and new trains. The main finding of the study is that developing the infrastructure of the bottleneck corridors is more effective than upgrading underutilized corridors for reduction of total delay. In our experiments, the average delay in generated timetable relates strongly to the railway infrastructure. As expected, by increasing the number of platforms in the stations, the average delay becomes shorter. However, railway development problem makes congestion issues that affect the resulting timetable. The experiments showed that there is a potential improvement for decreasing the average delay by implementing the decomposition approach and construction of new platforms in the bottleneck corridors.

Evidence that our model helps the railway operators in infrastructure development decisions is that the evaluation of several capacity expansion scenarios takes a lot of time even in the simulation software products. It should be noted that optimizing the use of railway infrastructure is a complex and time-consuming task. Therefore, numerous capacity expansion scenarios should be evaluated in order to work out how many extra trains can be scheduled by the existing infrastructure and how much investment will be required for new infrastructure. In other words, the railway capacity is extremely dependent on infrastructure. It is clear that the proposed decomposition approach does not guarantee for the optimal solution to complete mathematical model. As a consequence, the results lend further credence to our earlier suggestion that decomposition approach (especially route-based decomposition part) has short computation time as an advantage for large-scale problems. Specially, we conclusively recommend establishing a reductionism procedure for eliminating the extra decision variables before optimization. The preprocessing helps the decomposition approach to explore the regions in a more promising way in less time than the exact approaches does. The proposed method allows us to analyze the performance of the railway networks (effect of infrastructure development) on delays as it has been demonstrated in the real case.

8. Conclusion

The design and optimization of new railway infrastructure is a complex long-term planning procedure that involves several operational constraints. The new upgraded rail infrastructure affects the timetabling decisions as well. Thus, it is worth analyzing both tactical and strategic aspects of this problem in an integrated approach. Up to now, there is no optimization-based decision support tool to determine a minimum cost infrastructure upgrading plan by taking into account all the constraints defined by the operation of train traffic in the railway system. In this paper, we presented a mathematical programming model for the train scheduling and railway infrastructure development problems. The Iranian railway network was chosen as the test benchmark. The religious constraints associated with the passenger praying services were addressed in the timetabling model. The mathematical model of the present paper was then extended to handle the infrastructure planning issues. In our framework, the complex mathematical model was simplified by two heuristic methods based on variable fixing techniques in order to be solved efficiently. We also presented a network-wide decomposition procedure with the aim of reducing the complexity of the complete mathematical model of the centralized instance. The result of implementing the proposed method shows the advantages of the methodology to design the railway networks with minimum cost so that it makes us capable of scheduling new trains efficiently. We considered the construction of new tracks and platforms as infrastructure development options to upgrade the infrastructure for decreasing the average delay. For the future studies,
the existing mathematical model can be combined with train routing model to cope with real assumptions. Our mathematical modeling approach is also capable to be incorporated with detailed evaluation of the robustness and reliability indexes.

References


Biographies

Masoud Shakibayifar received his PhD degree in Industrial Engineering from Iran University of Science & Technology in 2017. He received BSc (1997-2001) and MSc (2004-2006) degrees in Civil Engineering at Sharif University of Technology, Tehran, Iran. He has managed more than 15 transportation planning projects in the Iranian Railway Company and the Ministry of Roads and Urban Development so far. Recently, he has managed a research project related to the design of railway traffic simulation software for the Iranian Railway Company. His research interests are railway electrification, railway scheduling and rescheduling, and optimization of transportation systems under uncertainty.

Erfan Hassannayebi is an Assistant Professor of Industrial Engineering at Department of Industrial Engineering, Faculty of Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran. He received his PhD degree in Industrial Engineering from Tarbiat Modares University in 2017. His research interests are in the area of applied operational research, simulation-based optimization, and metaheuristic algorithms. He has several publications in international journals.

Hamid Mirzahossein received his BSc degree in 2009, MSc degree in 2011, both from Imam Khomeini International University (IKIU), and PhD degree in 2017 from Iran University of Science and Technology (IUST). At present he is Assistant Professor at Imam Khomeini International University (IKIU). His research interests are mainly in the areas of land-use and transportation interaction with special focus on accessibility and mathematical models and optimization techniques for congestion pricing and traffic assignment, and more recently, on mathematical optimization of railway systems.

Shaghayegh Zohrabnia received her MSc degree in Industrial Management from Allameh Tabatabae University of Tehran, Iran. Her professional interests focus on multi-objective optimization, meta-heuristic methods in solving NP-hard problems such as Tabu Search, Simulated Annealing and Genetic Algorithm—target costing and value engineering, balanced scorecard, management information system, queuing system optimization, nonparametric methods in operations research such as Data Envelopment Analysis (DEA). Her current project includes timetabling optimization of mixed double and single tracked railway network.

Ali Shahabi is a PhD candidate at the Islamic Azad University, Tehran Branch, in the field of Industrial Engineering. His research interests include simulation, reliability engineering, supply chain management and logistics, optimization methods, and advance quality control. He has also several publications in international journals.