Stochastic Optimization Model and Solution Algorithm for Robust Double-Track Train-Timetabling Problem

Muhammad Babar Khan and Xuesong Zhou

Abstract—**By considering various stochastic disturbances unfolding in a real-time dispatching environment, this paper develops a stochastic optimization formulation for incorporating segment travel-time uncertainty and dispatching policies into a medium-term train-timetabling process that aims to minimize the total trip time in a published timetable and reduce the expected schedule delay. Based on a heuristic sequential solution framework, this study decomposes the robust timetabling problem into a series of subproblems that optimize the slack-time allocation for individual trains. A number of illustrative examples are provided to demonstrate the proposed model and solution algorithms using data collected from a Beijing–Shanghai high-speed rail corridor in China.**

Index Terms—**Slack-time allocation, stochastic optimization, train scheduling, train timetabling.**

I. Introduction

As ONE of the fundamental technical documents in the railroad industry, the train timetable provides a basis for synchronizing most of the scheduling activities over physical rail networks. In real-time railroad operations, published timetables are often affected by various random unforeseen events (e.g., temporary signal failure, inclement weather conditions, and equipment breakdown and track maintenance). As a result, the segment travel times and/or arrival times of trains could significantly deviate from the planned timetable. At the same time, interdependencies between trains could lead to knock-on delays, i.e., a delay from one train propagates to the following train(s). In particular, these delays could have a snowball effect on the subsequent operations in a rail line with heavy traffic. To improve travel time reliability for train users (including both passengers and freight shippers), slack time needs to be programmed in the planning stage to make published timetables robust against stochastic disturbances.

In the last few decades, a wide range of scheduling models and efficient solution algorithms have been developed. For a broad overview on different aspects of train-scheduling problems in planning and dispatching stages, see [1].

Manuscript received June 20, 2008; revised February 18, 2009 and June 12, 2009. First published September 25, 2009; current version published March 3, 2010. The Associate Editor for this paper was H. Dia.

M. B. Khan is with the School of Civil and Environmental Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan (e-mail: babarkhan152@hotmail.com).

X. Zhou is with the Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112 USA (e-mail: zhou@eng.utah.edu).

Digital Object Identifier 10.1109/TITS.2009.2030588

Pioneering theoretical work on the robust train timetabling problem was carried out by Carey's research group [2]–[4]. They proposed analytical models and sequential solution procedures for estimating train delay distributions and delay propagation for various performance measures in scheduled transport systems. Along the direction of analytically calculating expected delays in a train timetable, Huisman and Boucherie [5] and Huisman *et al.* [6] developed a stochastic queueing model for delay estimation at a double-track section. Hallowell and Harker [7] presented a line delay model for the North American railroad to predict the expected delay caused by meet/pass conflicts in real-time dispatching. Focusing on cyclic timetables, Kroon *et al.* [8] proposed a stochastic optimization model to minimize the average weighted delay of all trains by allocating time supplements and buffer times. Vansteenwegen and Van Oudheusden [9] integrated a linear programming model in a simulation-based framework to evaluate and determine ideal running time allocation.

To integrate the planning and operational stages in a theoretically rigorous manner, Section II first formulates the train timetabling problem as a two-stage stochastic recourse model, which can be decomposed into a series of robust timetabling subproblems. In Section III, the subproblem is further reformulated and solved as a stochastic time-dependent shortest path problem. The comprehensive solution procedure, along with illustrative examples, is presented in Section IV, followed by numerical experiments using a high-speed rail corridor test data set from China.

II. Two-Stage Recourse Model

A. Conceptual Framework

Tables I and II present the parameters and variables that are used in the two-stage stochastic-recourse model. The unit of all time-related parameters and variables is 1 min.

As shown in Fig. 1, segments are numbered as $1, 2, \ldots, m$, and the stations are numbered as $0, 1, \ldots, m$. A medium-speed train can only yield to a high-speed train at the intermediate stations. Without loss of generality, the following discussion considers trains traveling from station 0 to m. For each train type, the departure time and minimum dwell time at stations are determined in the earlier line planning stage, and the segment running times are determined by the timetabling process.

The commonly used two-stage decision procedure in railroad timetabling can be described as follows: First, the planner

TABLE I
SUBSCRIPTS AND PARAMETERS USED IN MATHEMATICAL FORMULATIONS

Symbol	Description
I	= set of trains, $\lvert I \rvert = n$
J	= set of segments, $\lvert J \rvert = m$
U	= set of stations, $\lvert U \rvert = m+1$
T	= planning time horizon under consideration
i	= train index
j	= segment/station index
ω	= subscript for random scenario
t	= time index, $t = 1, \ldots, T$
$d_{i,j}$	= minimum required station dwell time before train i entering segment j
h_j	= minimum headway between arrival and departure times of two consecutive trains at segment j
r_i	= planned departure time (release time) for train i at its first station
$r_{i,\omega}$	= realized departure time for train i at its first station under scenario ω
$f_{i,j}$	= free-flow running time for train i at segment j
$f_{i,j,\omega}$	= fastest possible running time for train i at segment j under scenario ω
$\overline{w}_i$	= penalty on total completion time for train i
w_i^-, w_i^+	= earliness and lateness, respectively, deviation penalty from planned timetable for train i
Δ_i'	= total slack time of all the stations for train i
Δ_i	= total slack time of all the segments for train i

TABLE II
VARIABLES USED IN MATHEMATICAL FORMULATIONS

Symbol	Description
$b_{i,j}$	= departure time (beginning time) for train i at segment j
$e_{i,j}$	= arrival time (ending time) for train i at segment j
$s'_{i,j}$	= slack time for train i at station j
$s_{i,j}$	= slack time for train i on segment j
$B_{i,k,j}$	= 1 if train i is planned before train k on segment j, 0 otherwise
$B_{i,k,j,\omega}$	= 1 if train i is scheduled before train k on segment j in scenario ω, 0 otherwise
x	= set of decision variables containing $(b_{i,j}, e_{i,j})$ in the planning stage
y_ω	= real-time schedule $(b_{i,j,\omega}, e_{i,j,\omega})$ under random scenario ω in the dispatching stage

constructs a train timetable x. At that time, he/she knows the requirements about planned departure time r_i, free segment running time $f_{i,j}$, minimum dwell time $d_{i,j}$, and minimum headway h_j. In addition, the planner is assumed to have information or estimates about the actual departure time and the fastest possible segment running time before conducting train scheduling, which is denoted as $\tilde{\xi} = \text{vector}(r_{i,\omega}, f_{i,j,\omega})$.

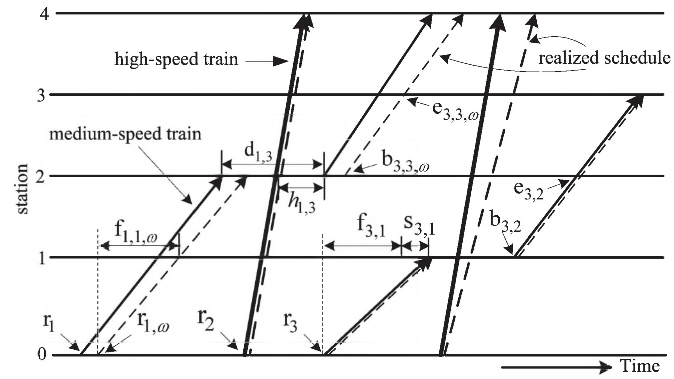

Fig. 1. Unidirectional double-track timetable at the planning and dispatching stages.

At the beginning of daily train dispatching, a realization ξ_ω of $\tilde{\xi}$ unfolds (i.e., scenario ω). The dispatcher adjusts the schedule to make a reactive decision (i.e., schedule) y_ω so that the deviation between schedule y_ω and timetable x is minimized under scenario ω. A scenario represents an instance in a set of real-time circumstances that change the departure time and free segment running times of trains. In reality, train trajectories are continuously realized, and the fastest possible segment running time could be changed even before a train reaches a segment. Accordingly, the rail line dispatcher can and should iteratively make corrective decisions using up-to-date estimates from station dispatchers or locomotive drivers. Thus, the real-time dispatching problem itself is essentially a multistage stochastic decision-making process. To construct a mathematically tractable model that can approximate the complex real-world situations, the real-time dispatching problem in our study is simplified to a one-stage decision, where we assume that the fastest possible running times on all segments at scenario ω are known at the *beginning* of daily scheduling.

In more detail, the conceptual two-stage recourse model is given as

$$\begin{aligned} &\text{Min } Z = c(x) + E_\omega q(x,\omega) \\ &\text{s.t.} \quad x \in \Omega \end{aligned} \tag{1}$$

where

$$\begin{aligned} &q(x,\omega) = \min g(y_\omega, x) \\ &\text{s.t.} \quad y_\omega \in \Phi_\omega(x). \end{aligned} \tag{2}$$

The function $c(x)$ in the first stage represents the total trip time of all trains. The second term in (1) is the weighted expected delay relative to the desired timetable. The recourse function $q(x,\omega)$ allows the model to capture the dispatcher's ability to take rescheduling actions, with $g(y_\omega, x)$ as the objective function. In addition, Ω is the set of feasible timetables, and $\Phi_\omega(x)$ is the set of feasible real-time schedules under scenario ω for a given x.

B. Medium-Range Planning Model

The medium-range planning model in this study is to minimize the total weighted trip time, where the weight $\bar{w}_i$ depends on the priority of train i. The following timetabling formulation is adopted from the optimization model proposed by Zhou and Zhong [10]:

$$\text{Min } c(x) = \sum_{i=1}^{n} \bar{w}_i(e_{i,m} - r_i). \tag{3}$$

Departure time constraints:

$$b_{i,1} = r_i \qquad \forall i \in I. \tag{4}$$

Segment running time constraints:

$$e_{i,j} = b_{i,j} + s_{i,j} + f_{i,j} \qquad \forall i \in I,\ j = 1, 2, \ldots, m. \tag{5}$$

Station dwell time constraints:

$$b_{i,j} \geq e_{i,j-1} + s'_{i,j} + d_{i,j} \qquad \forall i \in I,\ j = 1, 2, \ldots, m. \tag{6}$$

Safety headway constraints at segments:

$$e_{i,j} + h_{i,k,j} \leq e_{k,i} + B_{i,k,j} \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{7}$$

$$b_{i,j} + h_{i,k,j} \leq b_{k,j} + (1 - B_{i,k,j}) \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{8}$$

$$e_{k,j} + h_{k,i,j} \leq e_{i,j} + B_{i,k,j} \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{9}$$

$$b_{k,j} + h_{k,i,j} \leq b_{i,j} + (1 - B_{i,k,j}) \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{10}$$

where M is a sufficiently large number to model the preceding "either-or" type constraints.

Slack time constraints at segments:

$$\sum_j s_{i,j} \leq \Delta_i \qquad \forall i \in I,\ j = 1, \ldots, m. \tag{11}$$

Slack time constraints at stations:

$$\sum_j s'_{i,j} \leq \Delta'_i \qquad \forall i \in I,\ j = 1, \ldots, m. \tag{12}$$

Departure time constraint (4) ensures that the departure time of train i from the origin station should not be earlier than the planned departure time. Segment running time constraint (5) indicates that the planned running time should be greater than the free running time at a segment. Station dwell time constraint (6) connects the train activities of two consecutive segments through the stop time at a station. Constraints (7)–(10) impose the safety headway requirement between two consecutive trains running in the same direction at the same segment. The "either-or" relation in the segment headway constraints can be expressed as the two precedence constraints by introducing a binary variable $B_{i,k,j}$.

To absorb real-time stochastic disturbances, additional time supplements are added to the free segment running time and dwell times. Such time supplements help the dispatcher in two ways: 1) allowing trains to obey the planned segment running time or station departure time under minor disturbances and 2) making up for earlier delays by running faster than the planned segment running time. Constraints (11) and (12) depict the upper bounds for the total supplements on a trip. It is crucial to recognize the tradeoff of adding slack times in a train timetable. Additional slack times can improve the punctuality of a timetable, and a timetable with less or no slack time is prone to frequent delays and delay propagation among multiple trains. On the other hand, longer slack time could significantly increase the total train trip time, indirectly leading to a reduction of the total capacity of a rail line, particularly in a busy rail corridor with heavy freight and passenger demand. Thus, the essential goal of robust timetabling is to achieve the best tradeoffs between punctuality and efficiency of rail service.

C. *Daily Dispatching Model*

The second-stage formulation aims to minimize the real-time schedule deviation (under scenario ω) from the planned timetable, whereas trains arriving late or early at the final destination station are penalized. It should be noted that one could also consider the schedule deviations only at major stations/terminals and not just the final destination. For an outbound train traveling from station 0 to station m, the objective function in the second stage is expressed as

$$\text{Min } g(y_\omega, x) = \sum_{i=1}^{n} \left(w_i^+ (e_{i,m,\omega} - e_{i,m})^+ + w_i^- (e_{i,m,\omega} - e_{i,m})^- \right). \tag{13}$$

Given the planning timetable $(b_{i,j}, e_{i,j})$ and the realized fastest possible travel times $f_{i,j,\omega}$, the second-stage problem needs to determine train travel times and dwell times under scenario ω. The set of feasible real-time schedules $\Phi_\omega(x)$ can be defined by the following constraints:

Departure time constraints:

$$b_{i,1,\omega} = r_{i,\omega} \qquad \forall i \in I. \tag{14}$$

Segment running time constraints:

$$e_{i,j,\omega} \geq b_{i,j,\omega} + f_{i,j,\omega} \qquad \forall i \in I,\ j = 1, 2, \ldots, m. \tag{15}$$

Station dwell time constraints:

$$b_{i,j,\omega} \geq e_{i,j-1,\omega} + d_{i,j} \qquad \forall i \in I,\ j = 1, 2, \ldots, m. \tag{16}$$

Safety headway constraints at segments:

$$e_{i,j,\omega} + h_{i,k,j} \leq e_{k,i,\omega} + B_{i,k,j,\omega} \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{17}$$

$$b_{i,j,\omega} + h_{i,k,j} \leq b_{k,j,\omega} + (1 - B_{i,k,j,\omega}) \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{18}$$

$$e_{k,j,\omega} + h_{k,i,j} \leq e_{i,j,\omega} + B_{i,k,j,\omega} \times M \qquad \forall i \neq k, i, k \in I, j \in J \tag{19}$$

$$b_{k,j,\omega} + h_{k,i,j} \leq b_{i,j,\omega} + (1 - B_{i,k,j,\omega}) \times M \qquad \forall i \neq k, i, k \in I, j \in J. \tag{20}$$

For each scenario ω, it is assumed that the departure time of train i at its starting station $r_{i,\omega}$ is known *a priori*, and the actual segment running time should not be less than the fastest possible running time $f_{i,j,\omega}$. Similar to most of the previous studies on delay propagation, we impose the restriction that a train is not allowed to arrive at a station earlier than the planned arrival time.

As the dwell time is typically associated with passenger or freight loading/unloading activities, this study assumes that each train still needs to satisfy the (same) planned dwell times and may not leave a station earlier than the planned departure time under all possible scenarios, i.e.,

$$b_{i,j,\omega} \geq b_{i,j} \qquad \forall i \in I, \; j = 1, 2, \ldots, m \tag{21}$$

$$e_{i,j,\omega} \geq e_{i,j} \qquad \forall i \in I, \; j = 1, 2, \ldots, m. \tag{22}$$

As a result, no early arrivals can occur, and the objective function (13) reduces to

$$\text{Min } g(y_\omega, x) = \sum_{i=1}^{n} \left(w_i^+ (e_{i,m,\omega} - e_{i,m})^+ \right). \tag{23}$$

III. Solution Strategies

A. Sequential Decomposition

In the preceding model, safety headway constraints (7)–(10) and (17)–(20) become the coupling constraints in this two-stage problem. There are a number of strategies (e.g., Lagrangian relaxation) that could be applied to decompose the problem to a series of subproblems for individual trains. In this study, we use a sequential heuristic approach to iteratively determine the timetable for each individual train. A subproblem needs to handle a single train, and its inputs are the planning timetable and real-time schedule from the previously computed trains. In other words, a train needs to yield to the previously scheduled trains if a conflict exists in both planning and dispatching problems.

Regarding the sequence of trains to be scheduled, this study first computes high-speed trains and then medium-speed trains. For the same type of trains, trains departing earlier from the starting station are considered to have higher priority. It should be noted that the sequential solution strategy could particularly be ineffective in a single-track rail line setting, which requires additional future studies that are out of the scope of this paper.

B. Space-Time Network Representation

This section presents a space-time network representation to reformulate the optimal slack allocation problem as the shortest path problem, which is solvable by a wide range of computationally efficient algorithms. Let $G = (V, A)$ be a digraph, where V is the set of nodes, and A is the set of arcs. A node is jointly defined by a station number with a time index that represents the possible departure and arrival times to and from the stations. The time index ranges from 1 to T, where T represents the length of a day (i.e., 1440 min). For illustrative purposes, nodes are shown in Fig. 2 for a time interval of 10 min. The traveling activities of a train are represented with two types of arcs in set A: 1) segment arcs between two consecutive stations, which are associated with a cost as the segment running time, and 2) station waiting arcs, which indicate the stop time at a station.

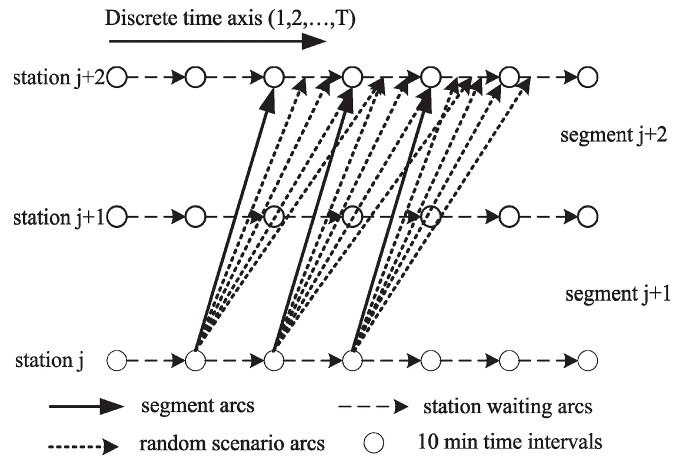

Fig. 2. Space-time network representation of a train timetable.

In Fig. 2, a train path trajectory with solid lines in an expanded network represents a slack time allocation (timetable) alternative for a train over different segments. The (schedule) scenarios with random travel times are shown by dashed lines. In the dispatching stage, the segment running times are given as random variables with certain delay probability distributions. The actual train trajectory between its origin and destination could be one of many different paths.

C. Stochastic Shortest Path Reformulation

Essentially, the subproblem under consideration seeks to determine a "reference" path that can minimize the total trip time (in the planning timetable) and the expected schedule deviations from the planned timetable among all possible random scenarios.

The train timetabling problem with stochastic segment running times is characterized as a special case of the stochastic shortest path problem. More precisely, this is the *a priori* stochastic least expected time path problem with the cost function as schedule delay, because the recourse decisions are taken once the values of one or more random variables are realized. Note that, to realistically model the multistage dispatching problem previously mentioned, the *adaptive* stochastic least-expected cost-path algorithm is needed to determine optimal switching strategies at each node for en-route decisions in response to experienced travel times, with updated travel time estimates. For congested road traffic networks, Miller-Hooks and Mahmassani [11] proposed efficient algorithms and procedures to determine a path that has the *least possible time* from all origins to a single destination for each departure time.

D. Sample-Based Scenario Representation for Random Segment Running Time

The proposed solution methodology for the single-train subproblem needs to find an optimal slack-allocation plan that minimizes a utility function of the total trip time and expected delay, as shown in (1), in a space-time network. Without loss of generality, the following description assumes no slack time at stations. As mentioned earlier, in daily railroad dispatching, the number of possible segment-running-time scenarios exponentially increases due to unforeseen events. An efficient sample average approximation method is detailed here to choose representative samples from the huge number of randomly generated segment running times.

Before solving all the train timetabling subproblems, we generate a set of random segment running times to be used in the second stage. Let F_ω be a vector of nonnegative integer segment running times $[f_{i,j,\omega}]$ for all trains and all segments, where scenario $\omega = 1, \ldots, W$, and W is the total number of randomly generated samples. Scenario ω occurs with given probability ρ^ω, where $\sum_{\omega=1}^{W} \rho^\omega = 1$. In other words, $f_{i,j,\omega}$ represents one realization of segment running times for train i on segment j under scenario ω.

In many stochastic shortest path algorithms (e.g., Miller-Hooks and Mahmassani [11]), a discrete probability distribution function (pdf) of segment (i.e., link) travel times are constructed from historical data samples and further used as the input, whereas the segment travel times are typically assumed to be independently distributed. For instance, the pdf-based representation could assume that there are k distinctive values of segment travel time on a segment for a train. If we consider the pdf propagation of stochastic trajectories of a train from one segment $j+1$ to the next segment $j+2$ along a corridor, as shown in Fig. 2, the combinations of random segment running times for a train on segment m is k^m. If there are a total of n trains, to describe the interactions (defined by safety headway constraints) between trains, the total number of such possible combinations could be $(k^m)^n = k^{m \times n}$. Considering segment $j+2$ in Fig. 2 with three trains on segment $j+2$, there are k^{j+2} scenarios for each train, and therefore, we need $(k^{j+2})^3$ scenarios to fully capture the possible conflicts.

The proposed sample-based approach is to draw W samples from the entire population to reduce the computational efforts and also capture the possibly correlated segment travel times, where W is significantly less than $k^{m \times n}$. For a single train, segment travel times in different samples can independently be distributed or correlated between different segments. For the same segment, the travel time distributions of different trains could also be independent and correlated.

E. *Stochastic Dominance Rules*

The stochastic dominance rule is used to compare two partial timetables with different slack-time-allocation strategies and prune off dominated train timetables as early as possible. As the slack time allocation aims to reduce average train delays on final stations, the probability of distribution of train delays can be used to indicate which slack time allocation strategy could effectively reduce the average delays. Consider train i on segment j at different partial timetables v' and v''. To ensure that the two partial train timetables v' and v'' are comparable, we require that the planned arrival times of train i on segment j, i.e., $e'_{i,j}$ and $e''_{i,j}$, in both partial timetables be the same.

Let $F'(\delta)$ and $F''(\delta)$ be the cumulative distribution functions (cdfs) of train delay δ corresponding to the planned arrival time $e'_{i,j}$ at partial timetables v' and v'', respectively. $E'(\delta)$ and $E''(\delta)$ are the expected values of delay δ at partial timetables v' and v'', respectively.

We have two dominance rules.

1) Timetable v'' *first-order stochastically dominates* timetable v', if $F''(\delta) \geq F'(\delta)$. Graphically, the cdf of delay distribution for timetable v'' is above or overlapping with the counterpart in timetable v' for any δ.
2) Timetable v'' *second-order stochastically dominates* timetable v', if $E''(\delta) < E'(\delta)$, i.e., the expected delay in timetable v'' is less than its counterpart in timetable v'.

Generally, it is difficult to simultaneously construct and apply stochastic dominance rules for different types of trains (and multiple trains). To apply the proposed stochastic dominance rule for a single train, we use the aforementioned sequential decomposition scheme, where the order is determined according to the type of trains or a prespecified level of priority.

IV. Simultaneous Timetable and Schedule Optimization Algorithm

The following algorithm details a stepwise procedure to solve the proposed two-stage stochastic model.

Notations

$\lambda(i,j,t)$	expected delay label for train i arriving at station j at planned arrival time t in the planning timetable;
$\delta(i,j,t,\omega)$	schedule delay for train i arriving at station j w.r.t. planed arrival time t under scenario ω;
$\pi_{i,j,t}$	arrival time predecessor at station $j-1$ for train i arriving at station j at time t;
$\sigma_{i,j,t}$	departure time predecessor at station $j-1$ for train i arriving at station j at time t.

Note that $\pi_{i,j,t} < \sigma_{i,j,t}$ if train i yields to other trains at segment j.

A. *Algorithmic Description*

Step 0: Generate a random segment travel time sample vector $F = [f_{i,j,\omega}]$, $i \in I$, $j = 1, 2, \ldots, m$, $\omega = 1, 2, \ldots, W$, with

$$\rho^\omega = \frac{1}{W}.$$

For train $i = 1, 2, \ldots, n$

Step 1: Feasibility checking for safety headway constraints.

Step 1.1: Fetch optimal timetable (i.e., $b_{k,j}$ and $e_{k,j}$) and related schedules (i.e., $b_{k,j,\omega}$ and $e_{k,j,\omega}$) for previously optimized trains $k = 1, \ldots, i-1, \forall j \in J$.

Step 2: Optimize robust timetable for train i.

Step 2.1: Initialize expected delay label $\lambda(i,j,t) = \infty$ for station $j = 0, 1, \ldots, m$ and time $t \geq r_i$.
Initialize $\lambda(i,j,t) = 0$ and $\delta(i,j,t,\omega) = 0$ for starting station $j = 0$, $t = r_i$, and $\omega = 1, 2, \ldots, W$.

Step 2.2: **For station** $j = 1, 2, \ldots, m$

For each **arrival time** t at station $j-1$

For each **slack time allocation** $s_{i,j} = 0, 1, \ldots, \Delta_i$ that corresponds to constraint (11)

Step 2.2.1: Find the earliest feasible departure time $b_{i,j}$ at station $j-1$ that satisfies $b_{i,j} \geq t + d_{i,j}$ and safety headway constraints, and find arrival time $e_{i,j}$ at station j as

$$e_{i,j} = b_{i,j} + s_{i,j} + f_{i,j}. \tag{24}$$

Step 2.2.2: Determine the resulting daily schedules $(b_{i,j,\omega}, e_{i,j,\omega})$ for planned departure time $b_{i,j}$ at station $j-1$ and arrival time $e_{i,j}$ at station j.

For random scenario index $\omega = 1, 2, \ldots, W$

Step 2.2.2.1: To solve constraints (16) and (21), find the earliest feasible departure time $b_{i,j,\omega}$ at station $j-1$ that satisfies

$$b_{i,j,\omega} \geq \max\left(b_{i,j}, t + \delta(i, j-1, t, \omega) + d_{i,j}\right) \tag{25}$$

where $\delta(i, j-1, t, \omega)$ is the schedule delay for train i arriving at station $j-1$ under scenario ω w.r.t. planned arrival time t so that $t + \delta(i, j-1, t, \omega)$ is the actual arrival time at station $j-1$ under scenario ω, and $t + \delta(i, j-1, t, \omega) + d_{i,j}$ is the earliest possible departure time at station $j-1$ under scenario ω by considering dwell time requirement $d_{i,j}$.

Step 2.2.2.2: To solve constraints (15) and (22), find arrival time $e_{i,j,\omega}$ at station j as

$$e_{i,j,\omega} = \max(e_{i,j}, b_{i,j,\omega} + f_{i,j,\omega}). \tag{26}$$

Step 2.2.2.3: Calculate the schedule delay w.r.t. planned arrival time $e_{i,j}$ as

$$\delta(i, j, e_{i,j}, \omega) = e_{i,j,\omega} - e_{i,j}. \tag{27}$$

EndFor // random scenario ω

Step 2.2.3: Calculate the average schedule delay of all scenarios

$$\overline{\lambda} = \sum_{\omega=1}^{W} \left(\delta(i, j, e_{i,j}, \omega) \times \rho^{\omega}\right). \tag{28}$$

Step 2.2.4: If $\overline{\lambda} < \lambda(i, j, e_{i,j})$, apply the second-order dominance rule, update the expected delay label by $\lambda(i, j, e_{i,j}) = \overline{\lambda}$, and update predecessors $\pi_{i,j,t}$ and $\sigma_{i,j,t}$ in the planning timetable.

EndFor // slack time s

EndFor // arrival time t

EndFor // station j

Step 3: Finalize the optimal timetable and daily schedules for train i.

Step 3.1: Find final arrival time t at station m that minimizes $(t - r_i) + \beta \times \lambda(i, m, t)$, where $(t - r_i)$ is the total travel time, $\lambda(i, m, t)$ is the expected schedule delay for train i to arrive at the final station at time t in the planning timetable, and β is the predetermined weight of expected schedule delay.

Step 3.2: From segment m to 0, use predecessors $\pi_{i,j,t}$ and $\sigma_{i,j,t}$ to backtrack optimal timetable (i.e., $b_{i,j}$ and $e_{i,j}$), and find the resulting schedule (i.e., $b_{i,j,\omega}$ and $e_{i,j,\omega}$) for each scenario $\omega = 1, 2, \ldots, W$.

EndFor // train i

B. Multiple Loops for Finding Feasible Departure and Arrival Times for Timetable and Schedules

The outer loop of the preceding algorithm sequentially finds a robust timetable for each train. Inside this loop, step 1 fetches the timetable and schedules of previously optimized trains $k = 1, \ldots, i-1$ from step 3, and step 2 computes the optimal timetable and schedules of train i from station 1 to station m that satisfy safety headway constraints between train i and trains $k = 1, \ldots, i-1$.

Step 2 includes three nested loops for enumerating planning timetable options: For each station, for each feasible arrival time, and for each feasible slack time allocation $s_{i,j}$. The (departure time, slack time) pair can be viewed as a segment running time arc at a station in the space-time network representation for the planning timetable. Moreover, step 2.1 initializes node labels in the space-time network, and step 2.2.1 finds feasible planning departure time $b_{i,j}$ and arrival time $e_{i,j}$ associated with segment j.

Step 2.2.2 is the core step for solving the second-stage daily scheduling problem. Inside the corresponding inner loop for each scenario, steps 2.2.2.1 and 2.2.2.2 sequentially handle constraints (15)–(22). Specifically, the random segment running time sample $f_{i,j,\omega}$ generated from step 0 is used in step 2.2.2.2 to ensure that the station arrival time $e_{i,j,\omega}$ in daily schedules considers that $f_{i,j,\omega}$ and $e_{i,j,\omega}$ should be no earlier than the planned arrival time $e_{i,j}$.

With the average schedule delay statistics, steps 2.2.3 and 2.3.4 apply the second-order stochastic dominance rule to keep the best slack allocation combination up to segment j for different time points.

C. Scenario Representation and Problem Decomposition

In many existing sample-based stochastic optimization studies, the second-stage problem is considered to be an optimization problem as a whole, and a scenario represents realized values for all the random variables of the second-stage problem. In our problem, a scenario ω corresponds to a 2-D vector of segment running time $[f_{i,j}]$ for all trains and all segments, and each scenario can be viewed as a single day in daily scheduling. To simultaneously *optimize* timetables and related daily schedules for all the possible scenarios, our proposed algorithm focuses on a much smaller set of decomposed constraints (i.e., for a single train and a single segment under different random running time samples), and it implicitly enumerates feasible slack-time-allocation options (i.e., segment running time arc in the space-time network) at departure time at each station and calculates the corresponding average schedule delays. By doing so, we are able to apply the efficient space-time network representation, shortest path algorithm, and stochastic-dominance rule to systematically search for better timetable solutions.

D. Schedule Delay Representation

Fig. 3 aims to illustrate how to calculate schedule delay $\delta(i, j, t, \omega)$ for train i arriving at the downstream station of

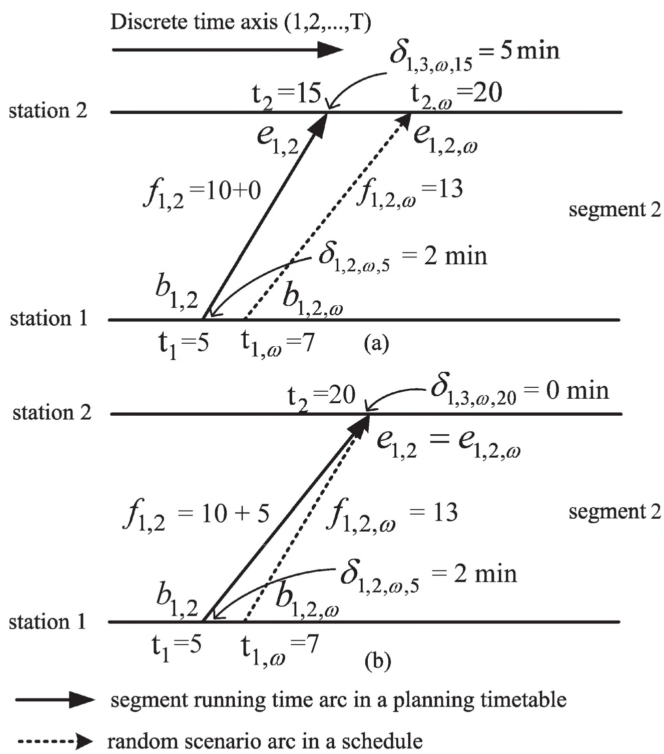

Fig. 3. Stochastic schedule-delay calculation for the single-train case.

segment j at scenario ω. For train 1, its planned departure time at station 1 is $b_{1,2} = 5$ min, whereas the free-flow running time at segment $j = 2$ is $f_{1,2} = 10$ min. Consider an unforeseen scenario ω unfolding in the second stage, which corresponds to a 2-min schedule deviation from the planned timetable at station 1 and segment running time $f_{1,2,\omega} = 13$. If no slack time is allocated in the planned timetable, as shown in Fig. 3(a), then the schedule delay is propagated from 3 min at station 1 to 5 min at station 2. In Fig. 3(b), 5 min of slack time is allocated to segment 2 in the planning timetable to reduce the delay propagation. By using such a timetable buffering strategy, the planned arrival time and the corresponding schedule arrival time are equal at $t_2 = 20$ min, leading to no schedule delay upon arriving at station 2.

In some cases, the delay of a train could be propagated to the following trains. Again, the term "knock-on delay" refers to the phenomenon of one train's delay that is caused by other trains in front of it. In Fig. 4(b), the effect of knock-on delay is illustrated between high- and medium-speed trains. It is more likely to have such delay propagation for high-density double-track corridor. If the trains are not allowed to change the planned order of trains, then the knock-on may cause the following trains to be greatly affected, as shown in Fig. 4(c).

In real-world dispatching situations, the arrival order of trains may deviate from the planned order due to many reasons, e.g., a train has experienced dramatic delay on the previous segments. In our study, we explicitly allow a possible change of order between high- and medium-speed trains traveling on the same segment. More precisely, in Fig. 4(d), if the high-speed train is significantly delayed and the medium-speed train is able to leave from station $j + 1$ at its planned departure time, then the medium-speed train could leave before the high-speed train. That is, the sequence of trains in this schedule is changed from the order specified in the planning timetable.

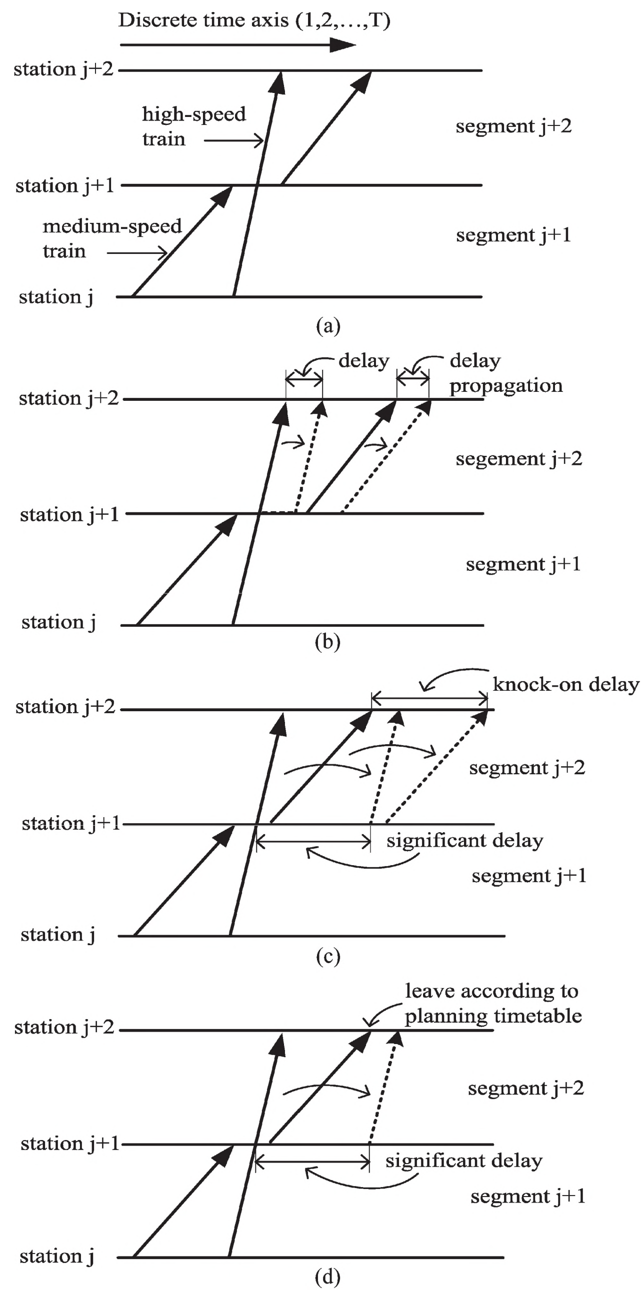

Fig. 4. Delay propagation and scheduling rules between trains. (a) Planning timetable. (b) Minor delay to high-speed train. (c) Significant delay to high-speed train. (d) Change of train order.

V. Numerical Experiments

In this section, we investigate the solution characteristics of the proposed algorithm through a series of numerical experiments. This study considers a train-scheduling problem for the double-track high-speed rail corridor between Beijing and Shanghai. The following experiments focus on a 615-km section, which consists of 17 sections between Shanghai and Xuzhou. Two types of passenger trains are planned to operate on this rail corridor, i.e., high-speed (250–300 km/h) and medium-speed (160 km/h) trains. For each train, it is assumed

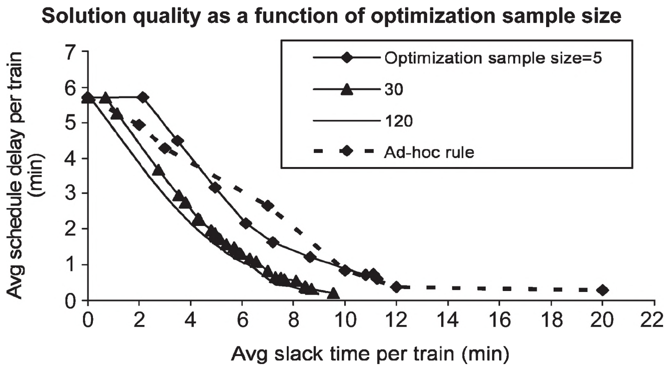

Fig. 5. Impacts of optimization sample size on solution quality.

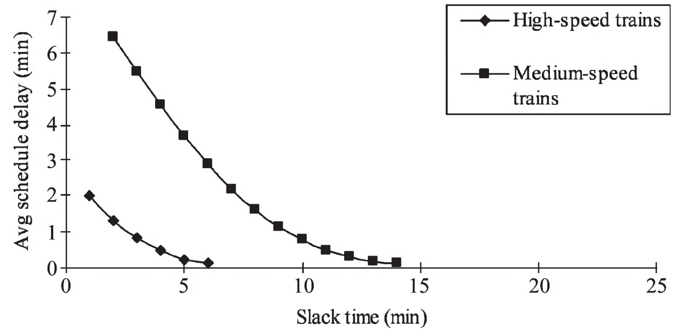

Fig. 6. Nondominated solutions for high- and medium-speed trains.

that the departure time at its origin station, segment running time, and minimum time for boarding and alighting at stations en route are prespecified by rail planners. In the test set, we consider ten high-speed trains and ten medium-speed trains running in the morning period (6:00 A.M.–11:00 A.M.), and we sequentially schedule high-speed trains according to departure times, followed by the medium-speed trains. The maximum slack time per segment is set to be 25% of the free-flow segment running time.

A. Impact of Optimization Sample Size on Solution Quality

We first investigate the effect of solution quality with different sampling sizes. It should be noted that two types of samples are used in our solution procedure: 1) optimization sample size and 2) evaluation sample size. The optimization sample size considers random segment running time samples taken from historical database (realized travel times) to represent the segment running time randomness in the daily train dispatching stage. A large evaluation sample size (e.g., 200 samples in the tests given here) is typically used to evaluate the solution quality of selected samples in optimization samples.

Fig. 5 shows the frontiers as results of the optimization sample size varying between 5 and 120. That is, the solution quality tradeoff curves are measured in terms of average schedule delay versus average slack time allocation per train. As expected, a large optimization sample size produces better tradeoff curves than a smaller sample size. In particular, for a 2-min average slack time, the optimization sample size of 120 drastically reduces the average schedule delay per train by 38%, compared with the optimization sample size of 5. As expected, an increase in the slack time per train further reduces the average schedule delay per train. On the other hand, adding too much slack time, e.g., 10 min in our experiments, the marginal reduction in average schedule delay becomes insignificant. Among the tested sample optimization sizes, overall, the sample size of 30 achieves the balance between computational efficiency and solution quality, and this setting is used for the rest of the experiments.

In Fig. 5, we also compare the best solutions found by our algorithm with solutions generated by an ad-hoc rule that allocates slack time in proportion to the free-flow segment running times. The dashed line connects different ad-hoc solutions generated by varying the total slack time budget. Clearly, with a large sample size, the robust optimization algorithm provides solutions that can almost dominate the results from the ad hoc rule.

B. Frontier of Nondominated Solutions (Average Delay Versus Slack Time)

In Fig. 6, we generate the frontiers of nondominated solutions for high- and medium-speed trains with different slack time allocations using five randomly generated schedules. As expected, the high-speed train, having high priority, shows less average delay, compared with the medium-speed trains. It should be noticed that, if a conflict occurs between two types of trains, the medium-speed train has to yield to the high-speed train, which accordingly increases the total delay and possible segment running time of the medium-speed train.

VI. Concluding Remarks

This study has proposed a two-stage stochastic recourse model that incorporates random segment travel time disturbances into a medium-term planning timetable. The first stage aims to minimize the total planned trip times of all the trains, and the second stage intends to reduce the average schedule delay under random instances. Focusing on improving the computational efficiency for solving this complex problem, we have further adopted a sequential solution procedure to decompose the complex two-stage stochastic optimization model into a series of subproblems for individual trains.

Our future research will address two main issues: First, the current solution algorithm only aims to minimize the expected delay at the final destination. It is more desirable and practical to also consider the reduction of schedule delays at intermediate (important) passenger terminals. Second, the effectiveness of the proposed sequential decomposition technique and stochastic dominance rules needs to be evaluated in detail with a variety of real-world test data sets.

References

[1] J. F. Cordeau, P. Toth, and D. Vigo, "A survey of optimization models for train routing and scheduling," *Transp. Sci.*, vol. 32, no. 4, pp. 380–404, Apr. 1998.
[2] M. Carey and A. Kwiecinski, "Stochastic approximation to the effects of headways in knock-on delays of trains," *Transp. Res. Part B*, vol. 28, no. 4, pp. 251–267, Aug. 1994.
[3] M. Carey and A. Kwiecinski, "Properties of expected costs and performance measures in stochastic models of scheduled transport," *Eur. J. Oper. Res.*, vol. 83, no. 1, pp. 182–199, May 1995.

[4] M. Carey, "Reliability of interconnected scheduled services," *Eur. J. Oper. Res.*, vol. 79, no. 1, pp. 51–72, Nov. 1994.

[5] T. Huisman and R. J. Boucherie, "Running times on railway sections with heterogeneous train traffic," *Transp. Res. Part B*, vol. 35, no. 3, pp. 271–292, Mar. 2001.

[6] T. Huisman, R. J. Boucherie, and N. M. van Dijk, "A solvable queueing network model for railway networks and its validation and applications for the Netherlands," *Eur. J. Oper. Res.*, vol. 142, no. 1, pp. 30–51, Oct. 2002.

[7] S. Hallowell and P. Harker, "Predicting on-time performance in scheduled railroad operations: Methodology and application to train scheduling," *Transp. Res. A*, vol. 32, no. 4, pp. 279–295, May 1998.

[8] L. Kroon, R. Dekker, G. Maroti, M. R. Helmrich, and M. Vromans, "Stochastic improvement of cyclic railway timetables," *Transp. Res. Part B*, vol. 42, pp. 553–570, 2008.

[9] P. Vansteenwegen and D. Van Oudheusden, "Decreasing the passenger waiting time for an intercity rail network," *Transp. Res. Part B*, vol. 41, no. 4, pp. 478–492, May 2007.

[10] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds," *Transp. Res. Part B*, vol. 41, no. 3, pp. 320–341, Mar. 2007.

[11] E. D. Miller-Hooks and H. S. Mahmassani, "Least expected time paths in stochastic, time-varying transportation networks," *Transp. Sci.*, vol. 34, no. 2, pp. 198–215, May 2000.

Muhammad Babar Khan received the B.Sc. degree in civil engineering and the M.Sc. degree in geotechnical engineering from National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2002 and 2004, respectively, and the Ph.D. degree in railroad transportation management from Southwest Jiaotong University, Chengdu, China, in 2008.

He is currently with the School of Civil and Environmental Engineering, NUST. He has conducted research on geotechnical problems in civil engineering. His research interests include stochastic modeling of scheduled transportation systems and solution algorithms related to timetabling and scheduling.

Dr. Khan is a member of the International Association of Railway Operations Research.

Xuesong Zhou received the B.S. degree in railroad management from Beijing Jiaotong University, Beijing, China, in 1995 and the Ph.D. degree in civil engineering from the University of Maryland, College Park, in 2004.

He is currently an Assistant Professor with the Department of Civil and Environmental Engineering, University of Utah, Salt Lake City. His current research interests include analytical modeling of transportation systems and real-time dynamic traffic estimation and prediction.

Dr. Zhou is a member of the Institute for Operations Research and the Management Sciences.