

# A Schedule-Based Model for Passenger-Oriented Train Planning With Operating Cost and Capacity Constraints

Jiemin Xie<sup>1</sup>, Shuguang Zhan<sup>1</sup>, Sze Chun Wong<sup>2</sup>, and Siu Ming Lo

**Abstract**—In the planning stage, train operators design timetables to serve passenger trips and a train circulation plan to support these timetables. These designs consider not only operating costs but also passenger convenience. In this study, we developed an optimization model for a new problem that focuses on timetabling and train-unit scheduling while also considering passenger itinerary choices in a schedule-based train system. This optimization model minimizes passenger travel costs within the constraints of a limited budget available for operating costs. The model is solved by an iterative heuristic that simulates the interaction between train operations and passenger itinerary choices. The heuristic solves the timetabling and train-unit scheduling problem using a decomposition approach to increase computational efficiency, while passenger loading is solved by a user-equilibrium passenger assignment model. An example based on the high-speed railway network in southern China was used to demonstrate the effectiveness of the proposed model and method.

**Index Terms**—Schedule-based, train planning, user-equilibrium passenger assignment, decomposition approach.

## I. INTRODUCTION

A TRANSIT system provides transit services to transport passengers with consideration of passenger and operating costs [1]–[3]. This study mainly focuses on the rail transit planning problem. During the planning stage, train operators

Manuscript received September 26, 2021; revised March 10, 2022; accepted April 25, 2022. This work was supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T32-101/15-R; and in part by the Guangdong-Hong Kong-Macau Joint Laboratory Program of the 2020 Guangdong New Innovative Strategic Research Fund, Guangdong Science and Technology Department, under Project 2020B1212030009. The work of Shuguang Zhan was supported in part by the National Natural Science Foundation of China under Grant 72171198 and in part by the Fundamental Research Funds for the Central Universities under Grant 2682021ZTPY072. The work of Sze Chun Wong was supported by the Francis S. Y. Bong Endowed Professorship in Engineering. The Associate Editor for this article was X. Ma. (Corresponding author: Shuguang Zhan.)

Jiemin Xie is with the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: xiejm28@mail.sysu.edu.cn).

Shuguang Zhan is with the School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: shuguangzhan@hfut.edu.cn).

Sze Chun Wong is with the Department of Civil Engineering, The University of Hong Kong, Hong Kong, and also with the Guangdong-Hong Kong-Macau Joint Laboratory for Smart Cities, Hong Kong, China (e-mail: hhecwsc@hku.hk).

Siu Ming Lo is with the Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong (e-mail: bcsml@cityu.edu.hk).

Digital Object Identifier 10.1109/TITS.2022.3173925

determine the number of scheduled train services and list these in the timetable by specifying stopping patterns and departure/arrival times. A train unit (TU) circulation plan helps operators to assign TUs with sufficient seats for onboard passengers to support the implementation of the timetable while considering the capability of depots to manage various types of TU and the availability of adequate train resources in depots for the next day's operations [4].

Additionally, a train plan should consider passenger dynamics. When a plan is used, passengers may change their itineraries accordingly. For example, fewer passengers may use train service T1 if a new train service T2 provides better service. These changes in passenger behaviors may warrant adjustments to the revised timetable and train circulation plan. In the above example, a smaller TU than the TU used in the new plan may be sufficient for T1. A planning process that could simulate this interaction would reduce the cost incurred from the various adjustments required to reach the point where passengers do not change their itinerary choices and operators do not need to further adjust their plans. Therefore, the passenger flows should not be fixed but should vary during the timetabling. However, most current train scheduling methods only consider fixed passenger flows.

Overall, train planning should (a) manage different types of TU; (b) allocate adequate TU resources in depots; and (c) consider passenger dynamics. Existing models incorporate one or two of these three important aspects, but no existing schedule-based model incorporates all three. Our study proposes a schedule-based model to combine these aspects. The model is to minimize the generalized passenger cost under the constraint of the operating cost, because the focus of operators has gradually shifted to market-oriented designs which aim to provide services that maintain the feasibility of passengers' itineraries and reduce their journey times [5], and operators would like to devote a certain amount of effort to satisfy existing customers and attract potential customers.

The model is solved by a new heuristic which addresses both timetabling and TU scheduling issues based on the decomposition approach, and adopts a user-equilibrium passenger assignment model to predict passengers' reactions to the timetable. The model and method are examined by applying them to the South China high-speed railway (HSR) network to demonstrate their efficiency and applicability.

The remainder of this paper is organized as follows. The literature is reviewed in Section II. The details of our model are described in Section III. The solution method is presented in Section IV. We apply the proposed methodology to one example network in Section V. Finally, a summary of the study is provided in Section VI.

## II. LITERATURE REVIEW

Our problem includes the line planning problem (LPP), transit timetabling problem (TTP), vehicle scheduling problem (VSP), and passenger assignment problem. The LPP, TTP, and VSP are the three important planning problems faced by operators and have been the subject of many studies. Guihaire and Hao [6] provided a review of the LPP and TTP and proposed integrated models for these problems. Parbo *et al.* [7] and Bunte and Kliewer [8] reviewed passenger-oriented TTPs and VSPs, respectively. We refer interested readers to these reviews for more details. Here, only the studies that are most relevant to our integrated planning problem are reviewed.

Some researchers have explored the integration of two or more of the three planning problems because the solutions obtained from such integration could potentially reduce costs and improve service quality. For example, Wang *et al.* [2] integrated the TTP and VSP to improve energy efficiency. The model of Zhao *et al.* [9] combined the TTP and VSP to balance the service level and operating cost. Michaelis and Schöbel [10] suggested a model for the LPP, TTP, and VSP to maximize the number of passengers using buses. Laporte *et al.* [11] proposed a model that solved the TTP, VSP, and user routing problem simultaneously to reduce the line running cost, fleet size cost, and passenger inconvenience.

However, previous integration studies have generally not considered the following aspects that are important to practical application.

(1) Vehicle type: The vehicle-type problem involves the assignment of an appropriate vehicle type to a transit service according to operating cost [12], passenger demand [4], [13], depot capabilities [14], and energy consumption [15]. If the assigned vehicle type supplies an insufficient vehicle capacity, passenger demand cannot be met. Conversely, if the assigned vehicle capacity supplies an excess vehicle capacity, operational expenses are wasted.

(2) End-of-day balance: Some depots may not be able to manage all vehicle types because of equipment restrictions. The end-of-day balance problem requires each depot to house the desired number of vehicles of each type at the end of an operational period, to ensure that sufficient vehicles are available for the next operational period [4]. Accordingly, to maintain an end-of-day balance, extra vehicles may need to be dispatched from some depots to other depots before the next operational period.

(3) Station-skipping: Station-skipping can help to decrease operating times, reduce fleet sizes, and enhance passenger services [16], [17].

(4) Station-track assignment: Station-track assignment considers station capacity and assigns tracks to train services to maintain safe operation [18] and smooth passenger

interchanges by ensuring that various train services share the same platform [17].

The above studies [2], [9]–[11] have not considered vehicle type or end-of-day balance. Schöbel [19] addressed the end-of-day balance problem in a frequency-based integrated model but ignored the stopping pattern design. Moreover, few integrated models assign tracks with consideration of passenger interchanges. Therefore, we developed a novel schedule-based model that considers vehicle type, end-of-day balance, station-skipping, and station-track assignment. In our study, the timetable design and TU scheduling require an analysis of passenger flows for each transit service, so compared with a frequency-based model, a schedule-based model is more suitable [20]–[22].

Furthermore, previous studies have used the system-optimum (SO) condition in passenger loading to establish timetabling models when considering vehicle capacity [23]. However, the SO condition cannot be applied realistically to a congested scenario, because it requires some passengers to sacrifice their ideal choices if the total passenger cost is to be minimized. In contrast, the user-equilibrium (UE) condition describes a state of equilibrium in which passengers (who are self-serving and aim to lower their own travel cost) cannot decrease their costs by choosing another itinerary. Hence, we used the UE formulation, as it is more appropriate for transit assignment than the SO condition.

Generally, a real-life problem considers multiple issues, involves large numbers of transit services and passengers, and consequently is too complex to solve directly. Thus, heuristics, such as the hybrid artificial bee colony algorithm [24] and column-generation-based heuristic [25], have been suggested to solve such problems. Alternatively, the iterative method [19], [26] can be used to solve the LPP, TTP, and VSP by executing an iterative process. Compared with other methods, the iterative method has the obvious advantage of capturing the interaction between passengers' choices and transit planning via a simulation process [27]. Thus, we adopt an iterative process.

The adoption of an iterative process alone may not be sufficient because the TTP is a non-deterministic polynomial-time (NP)-hard problem for which an exact global optimum is hard to find [18]. Many methods have been proposed to solve this problem, including genetic algorithm [28]–[31], Lagrangian duality theory [32], alternating direction method of multipliers algorithm [33], and decomposition approach (DA) [34]–[36]. The DA has been widely adopted to solve TTPs because it can handle complex problems. Generally, the DA involves decomposing the whole network into several zones and subsequently generating timetables within these zones with consideration of coordinated operations across the zones [34], or performing the decomposition at the transit service level [35], [36], wherein a timetable is created for each transit service. Our proposed DA provides a new approach to decomposition at the TU level that combines timetabling and TU scheduling. In each decomposed subproblem, a TU selects unscheduled train services to form its circulation path, and the timetables of train services are generated simultaneously such

that the operating cost can be easily controlled to fulfill the operating cost constraint.

Train type, end-of-day balance, station-skipping, station-track assignment, and passenger dynamics are all important aspects to consider for efficient train planning. However, no models have been developed that consider all of these aspects; in particular, station-track assignment has rarely been used to smooth passenger interchanges. This paper is thus a useful contribution to the train planning literature as it describes the development of a schedule-based model that considers all of the above aspects, and thereby improves passenger service under the constraint of an operating budget. An iterative heuristic based on a new decomposition approach was devised to solve this model. This heuristic simulates the interaction between passengers' choices and train planning, thereby ensuring that solutions are always in line with the main constraint, i.e., the operating budget constraint.

### III. MODEL FORMULATION

#### A. Problem Description

In this study, a *train service* is one run of a TU from the original terminal to the destination terminal, and a *line* assembles the train services that pass through the same stations in the same operating time window. Train services on the same line are allowed to have different stopping patterns. A schedule plan includes (a) a line plan that records which train services should be scheduled and which stopping patterns of scheduled train services can be used, (b) a timetable that details the departure/arrival times and station-track assignments of the scheduled train services, and (c) a train circulation plan that consists of TU assignments and reallocations.

Operating budget ( $\varepsilon$ ) is generally controlled to generate a schedule plan in real-life situations. Operators want to provide the best service to passengers possible in line with budget constraints. Hence, we treat the generalized passenger cost ( $C^P$ ) as the objective and include the operating cost ( $C^O$ ) as a constraint (Inequality III-A.1). Before the introduction of the problem details, the elements and sets used in this study are listed in Table I.

$$C^O \leq \varepsilon. \quad (1)$$

1) *Supply Side*: Operators are assumed to provide the operating cost boundary, the train network information, and the general planning framework. To clarify the train network information, a small-scale example is given. The simple network has three stations. A station can contain more than one station track to simultaneously accommodate several train services. One platform serves one station track; however, a station track that is not equipped with platforms can only serve train services that skip the station. Additionally, a real station can serve more than one train corridor and is divided into several yards, each of which serves train services in one train corridor and is modeled as an abstracted station. Travelers can change between these abstracted stations by using walking links. Double tracks are assumed to be used to connect stations, as shown in Fig.1. Each track serves one direction.

TABLE I  
ELEMENTS AND SETS USED IN THE DESCRIPTION OF THE MODEL

Element	
$f$	Train type
$i$	Itinerary
$\bar{i}$	Residual itinerary
$l, l'$	Line
$n, n'$	Train service
$s, s'$	Station
$s_{in}^B$ or $s_{in}^A$	Boarding or alighting station, respectively, if passengers on $i$ use $n$
$s_n^{Des}$ or $s_n^{Ori}$	Destination or origin terminal of train service $n$
$s_{ns}^{DS}$	Downstream station of station $s$ for train service $n$
$p, p'$	Track
$v$	Train unit (TU)
$r$	Origin–destination (OD) pair, which is defined as a group of passengers who depart the same origin zone for the same destination zone at the same preferred start time
Set	
$F$	Set of all train types
$F_s$	Set of all train types that the depot serving $s$ can serve
$N$	Set of all train services
$N_l$	Set of train services running on line $l$
$N_s^{iti}$	Set of train services used by itinerary $i$
$N_s^{End}$ or $N_s^{Sta}$	Set of train services terminating at or originating from $s$
$N^{Sch}$ or $N^{Uns}$	Set of scheduled or unscheduled train services
$N^{New}$	Set of train services newly added to $N^{Sch}$
$N^{Fix}$	Set of train services whose timetables are fixed
$S_n$	Set of passing stations for train service $n$
$S_n^{Stop}$ or $S_n^{Skip}$	Set of stations at which train service $n$ stops or skips
$I_r$	Set of itineraries that can be used by OD pair $r$
$L$	Set of all lines
$P_{ns}$	Set of tracks that train service $n$ can use at station $s$
$\bar{P}_{ns}$	Set of tracks that are equipped with platforms at station $s$ and can be used by train service $n$
$R$	Set of OD pairs
$V_f$	Set of TUs that belong to train type $f$

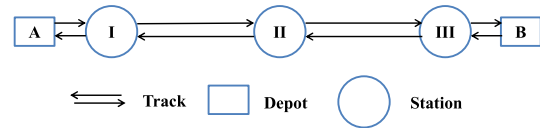


Fig. 1. Simple train corridor for the small-scale example.

A depot serves one adjacent terminal. For example, a TU is dispatched from depot A to the adjacent terminal, station I in Fig. 1, and then it provides the train service from station I to station III. When the TU arrives at station III and completes the train service, it may enter depot B or stop at station III for a certain duration, which is the layover time. During the layover time, the crew can prepare for the next trip from station III to station I. A depot may not be able to serve all train types because of equipment limitations.

A depot houses the TUs that it can handle at the start of operations for initial dispatching and the same number of the same types of TUs must be housed in the depot at the end of the day to ensure the availability of adequate TU resources for the next day's operations. If this requirement cannot be met, the end-of-day balance cannot be achieved and TUs must be reallocated.

Hence, the given train network information includes (a) the tracks without/with a platform at a station, (b) the train types that can be served by a depot, (c) the walking time between two platforms serving tracks  $p$  and  $p'$  ( $T_{pp'}^{Walk}$ ), (d) the safety train headway ( $T^H$ ), and (e) the minimum layover time ( $T^{LT}$ ).



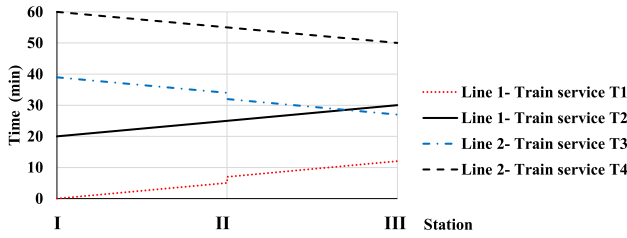


Fig. 2. Schedule plan for the small-scale example.

A general planning framework is derived using previous operation experience and provides information about line  $l$ , including (a) the passing stations, (b) the operating time window ( $[T_l^{\text{Start}}, T_l^{\text{End}}]$ ), (c) the minimum and maximum dwell times at station  $s$  ( $T_{sl}^{\text{Smin}}$  and  $T_{sl}^{\text{Smax}}$ ), (d) the minimum and maximum running times from  $s$  to  $s'$  ( $T_{lss'}^{\text{Rmin}}$  and  $T_{lss'}^{\text{Rmax}}$ ), (e) the fare, and (f) the number of train services.

Moreover, we assume that: (a) the number of available TUs is sufficient, and depots have sufficient space to accommodate parking TUs during the operating period; and (b) there are sufficient staff to form a feasible crew schedule.

The consideration of designing schedule plans can be explained using the network in Fig. 1. Two lines are planned, line 1 in the direction from station I to station III and line 2 in the opposite direction. Each line can contain a maximum of two train services in a one-hour period. A schedule plan for this example is presented in Fig. 2. If a train service stops at a station, its path in Fig. 2. extends along the time axis but not along the station axis. If a train service skips a station, its path moves along both the station and time axes.

If a train service is scheduled, the TU assignment problem is solved to allocate a suitable train type to the service depending on the onboard passenger flow and the facilities available at the terminals. Three TUs of the same type are assumed to be sufficient to support this schedule plan. As shown in Fig. 2, TU 1 serves T2 and T4 marked in black, while TU 2 serves T1 marked in red and TU 3 serves T3 marked in blue. T1 and T3 cannot be served by the same TU because the time difference between the arrival of T1 and the departure of T3 is less than the minimum layover time ( $T^{\text{LT}}$ ). In this case, the number of departures equals the number of arrivals at a terminal. The end-of-day balance is maintained. However, if these three TUs are not the same type or some train services are canceled because of a limited operating budget, TU reallocation is necessary to maintain the end-of-day balance. For example, if T4 is canceled, an extra dispatch from station III to station I is scheduled outside of the operation period.

As Eq. (2) shows, the operating cost includes the total variable cost ( $C^{\text{Var}}$ ), total fixed cost ( $C^{\text{Fixed}}$ ), and total cost of TU reallocation ( $C^{\text{TR}}$ ).  $C^{\text{Var}}$  includes the financial expenditures for energy, attrition, and human resources [8], and is assumed to be a linear function of the total train running time (Eq. (3)).  $C^{\text{Fixed}}$  is paid for investment in the planning period [8], and assumed to be a linear function of the number of TUs used (Eq. (4)).  $C^{\text{TR}}$  is introduced because of TU reallocation to maintain the end-of-day balance and calculated

based on Eq. (5).

$$C^{\text{O}} = C^{\text{Var}} + C^{\text{Fixed}} + C^{\text{TR}}, \quad (2)$$

$$C^{\text{Var}} = c^{\text{Var}} \times \sum_{n \in N, s=s_n^{\text{Ori}}, s'=s_n^{\text{Des}}} (t_{ns'}^{\text{Arr}} - t_{ns}^{\text{Dep}}); \quad (3)$$

$$C^{\text{Fixed}} = \sum_{f \in F} \sum_{v \in V_f} (c_f^{\text{Fixed}} \times (\sum_{n \in N} b_{vnn}^{\text{TU}})); \quad (4)$$

$$C^{\text{TR}} = \sum_{f \in F} \sum_{s \in S} \sum_{s' \in (S-s)} (c_{fss'}^{\text{TR}} \times \Pi_{fss'}^{\text{TR}}); \quad (5)$$

where  $c^{\text{Var}}$  is the average variable cost per minute;  $t_{ns}^{\text{Dep}}$  is the time at which train service  $n$  departs  $s$ ;  $t_{ns'}^{\text{Arr}}$  is the time at which  $n$  arrives at  $s'$ ;  $c_f^{\text{Fixed}}$  is the fixed cost for a TU of type  $f$ ; and  $b_{vnn}^{\text{TU}}$  equals 1 when  $n' \neq n$  and TU  $v$  serves  $n$  before serving  $n'$ , or when  $n' = n$  and  $n$  is the last train service served by  $v$ , and 0 otherwise. When  $\sum_{n \in N} b_{vnn}^{\text{TU}} = 1$ ,  $v$  is used; otherwise,  $v$  is not used.  $c_{fss'}^{\text{TR}}$  is the minimum variable cost between  $s$  and  $s'$  because the reallocated TU runs at maximum speed without stopping.  $\Pi_{fss'}^{\text{TR}}$  is the number of TUs of type  $f$  that are dispatched from  $s$  to  $s'$  to achieve the end-of-day balance.  $c^{\text{Var}}$ ,  $c_f^{\text{Fixed}}$ , and  $c_{fss'}^{\text{TR}}$  are given by the operators in advance.

2) *Demand Side*: Based on previous operations, train operators can predict the demand data and use them for planning. The given demand data for each OD pair include (a) the origin and destination zones, (b) the passenger flow, and (c) the departure time leaving the origin zone. In addition, access and egress times (i.e., the travel times between stations and their neighborhood zones) are given.

The generalized passenger cost is formulated as follows:

$$C^{\text{P}} = \sum_{r \in R} \sum_{i \in I_r} (A_i \times c_i^{\text{Iti}}); \quad (6)$$

$$c_i^{\text{Iti}} = c_i^{\text{Unc}} + c_i^{\text{Con}}, \forall r \in R, \forall i \in I_r; \quad (7)$$

where  $A_i$  is the number of passengers using itinerary  $i$ , and the generalized cost of  $i$  ( $c_i^{\text{Iti}}$ ) is the summation of the uncongested and congested costs (i.e.,  $c_i^{\text{Unc}}$  and  $c_i^{\text{Con}}$ ), as shown in Eq. (7).  $c_i^{\text{Con}}$  is related to the train capacity and calculated under the UE condition (the details are given in formulas (13)–(19) in the next section).  $c_i^{\text{Unc}}$  is the weighted sum of the in-vehicle time (IVT,  $c_{i1}$ ), waiting time ( $c_{i2}$ ), walking time ( $c_{i3}$ ), number of interchanges ( $c_{i4}$ ), access/egress times ( $c_{i5}$ ), and fare ( $c_{i6}$ ):

$$c_i^{\text{Unc}} = \sum_{j \in [1,6]} (w_{rj} \times c_{ij}). \quad (8)$$

The weights of these costs ( $w_{rj}$ ) can be calibrated in advance using passenger data. The last three cost components are fixed, whereas the first three vary as the timetable changes. The IVT of  $n$  in  $i$  from the boarding station  $s$  to the alighting station  $s'$  is given as follows:

$$t_{in}^{\text{IVT}} = t_{ns'}^{\text{Arr}} - t_{ns}^{\text{Dep}}. \quad (9)$$

Passengers of itinerary  $i$  wait for the first train service  $n$  at boarding station  $s$ , and the waiting time is given as follows:

$$t_{in}^{\text{Wait}} = t_{ns}^{\text{Dep}} - T_i^{\text{Ori}}; \quad (10)$$

where  $T_i^{\text{Ori}}$  is the arrival time of  $i$  at the origin station and is calculated by summing the departure time leaving the origin

zone and the access time needed for  $i$ . If passengers of  $i$  alight from  $n'$  at  $s'$  and interchange to  $n$  at  $s$ , the waiting and walking times can be computed as follows:

$$t_{in}^{\text{Wait}} = t_{ns}^{\text{Dep}} - (t_{n's'}^{\text{Arr}} + t_{in}^{\text{Walk}}); \quad (11)$$

$$t_{in}^{\text{Walk}} = \sum_{p' \in \bar{P}_{n's'}} \sum_{p \in \bar{P}_{ns}} (T_{p'p}^{\text{Walk}} \times b_{n'p'}^{\text{Track}} \times b_{np}^{\text{Track}}); \quad (12)$$

where  $b_{np}^{\text{Track}}$  equals 1 if  $p$  is used by  $n$ , and 0 otherwise. Eq. (12) shows that only when  $n$  stops at  $p$  and  $n'$  stops at  $p'$  ( $b_{n'p'}^{\text{Track}} = b_{np}^{\text{Track}} = 1$ ) does  $t_{in}^{\text{Walk}}$  equal  $T_{p'p}^{\text{Walk}}$ . The value of  $t_{in}^{\text{Walk}}$  depends on the value of  $b_{np}^{\text{Track}}$ , which is not determined in advance but is calculated during the station-track assignment. For example, consider a station with three tracks: track 1, track 2, and track 3. If  $n$  and  $n'$  are assigned to tracks 1 and 2, respectively, and the walking time from platform 1 to platform 2 ( $T_{2,1}^{\text{Walk}}$ ) is 5 min, then  $t_{in}^{\text{Walk}}$  is 5 min. However, if  $n$  and  $n'$  are assigned to tracks 1 and 3, respectively, and  $T_{3,1}^{\text{Walk}}$  is 10 min, then  $t_{in}^{\text{Walk}}$  is 10 min. This demonstrates how station-track assignment affects passenger walking time.

## B. Model Formulation

To build a schedule-based integrated model (SIM) for the proposed problem, the following constraints are needed:

(a) Constraint III-A.1, which ensures that the operating cost is within the budget;

(b) the constraint set for passenger loading  $CS^P$ , which assigns passengers according to the train capacity limit and UE condition;

(c) the constraint set for line planning  $CS^L$ , which sets rules to design the stopping pattern and determine which train services should be scheduled;

(d) the constraint set for timetabling  $CS^T$ , which ensures that the train timetable is safe and meets the operating requirements (such as the running and dwell time limit); and

(e) the constraint set for TU scheduling  $CS^S$ , which allocates the scheduled train services to TUs with consideration of train types and the end-of-day balance. Thus, the SIM model is as follows.

$$\min C^P,$$

s.t. Constraint III-A.1, constraint sets  $CS^P$ ,  $CS^L$ ,  $CS^T$ , and  $CS^S$ .

Details of constraint sets  $CS^P$ ,  $CS^L$ ,  $CS^T$ , and  $CS^S$  are introduced in the following sections.

### 1) Constraint Set for Passenger Loading $CS^P$ :

#### • Capacity constraints

In the proposed problem, we strictly control the passenger onboard flow and calculate the surcharge for using  $n$  from  $s$  to  $s'$  ( $c_{nss'}^{\text{Sur}}$ ) based on the passenger flow. The relative constraints are as follows ( $\forall n \in N$ ,  $\forall s, s' \in S_n$ :  $s' = \xi_{ns}^{\text{DS}}$ ):

$$A_{nss'}^{\text{Flow}} \leq \Pi_n^{\text{Cap-T}}; \quad (13)$$

$$c_{nss'}^{\text{Sur}} = \begin{cases} = 0, & \text{if } \Pi_n^{\text{Cap-T}} > A_{nss'}^{\text{Flow}} \\ \geq 0, & \text{if } \Pi_n^{\text{Cap-T}} = A_{nss'}^{\text{Flow}} \end{cases}; \quad (14)$$

where  $A_{nss'}^{\text{Flow}}$  is the number of passengers on  $n$  from  $s$  to  $s'$  ( $A_{nss'}^{\text{Flow}} = \sum_{r \in R} \sum_{i \in I_r} b_{inss'}^{\text{Iti}} \times A_i$ , where  $b_{inss'}^{\text{Iti}}$  equals 1 if itinerary  $i$  uses  $n$  from  $s$  to  $s'$ , and 0 otherwise); and  $\Pi_n^{\text{Cap-T}}$  is the capacity of  $n$  ( $\Pi_n^{\text{Cap-T}} = \sum_{f \in F} \sum_{v \in V_f} \sum_{n' \in N} (b_{vnn'}^{\text{TU}} \times \Pi_f^{\text{Cap}})$ , where  $\Pi_f^{\text{Cap}}$  is the capacity of a type  $f$  TU). Constraint (13) ensures that the passenger onboard flow from  $s$  to  $s'$  cannot be larger than the train capacity. Constraint (14) shows that if the train capacity exceeds the passenger flow on  $n$  from  $s$  to  $s'$ ,  $c_{nss'}^{\text{Sur}}$  equals 0, whereas if the train capacity equals the passenger flow on  $n$  from  $s$  to  $s'$ ,  $c_{nss'}^{\text{Sur}}$  is greater than or equal to 0. Then, the congested cost of itinerary  $i$  in constraint (6) can be calculated based on  $c_{nss'}^{\text{Sur}}$  ( $\forall i \in I_r$ ,  $\forall r \in R$ ), as follows:

$$c_i^{\text{Con}} = \sum_{n \in N} \sum_{s, s' \in S_n: s' = \xi_{ns}^{\text{DS}}} (b_{inss'}^{\text{Iti}} \times c_{nss'}^{\text{Sur}}). \quad (15)$$

$c_i^{\text{Con}}$  is incurred by passengers when they must exert a certain effort to board a train in a train system that does not allow seat reservations [37] or must buy tickets in a train system that does allow seat reservations [38]. Previous studies [37], [38] have proved that  $c_i^{\text{Con}}$  in both types of train systems can be calculated using the same equation, i.e., Eq. (15), which shows that  $c_i^{\text{Con}}$  is the sum of the surcharges for all sections along the itinerary.

#### • UE condition

The proposed problem considers assigning passengers based on the UE condition ( $\forall i \in I_r$ ,  $\forall r \in R$ ), which is represented as  $c_i^{\text{Iti}}$ , i.e., the generalized cost of  $i$ :

$$c_i^{\text{Iti}} \begin{cases} = \tilde{c}_r, & \text{if } A_i > 0 \\ \geq \tilde{c}_r, & \text{if } A_i = 0. \end{cases} \quad (16)$$

where  $\tilde{c}_r$  is the equilibrium cost over all of the itineraries of OD pair  $r$ . The UE condition states that the generalized costs of the chosen itineraries are equal to or not greater than those of any unchosen itinerary. Constraint (16) has been widely used to describe the equilibrium route flow pattern that satisfies the UE condition in transit networks (Szeto *et al.* 2013; Xu *et al.* 2018).

#### • Passenger flow constraints

The passenger assignment must be workable. For example, onboard passenger flows must be non-negative and if  $n$  is not scheduled to stop at  $s$  or  $s'$  ( $b_{sn}^{\text{Stop}} = 0$  or  $b_{s'n}^{\text{Stop}} = 0$ ), passengers cannot use  $n$  and the associated flow should be 0 ( $A_i = 0$ ). Thus, Constraint (17) is set as follows ( $\forall i \in I_r$ ,  $\forall n \in N_i^{\text{Iti}}$ ,  $\forall r \in R$ ):

$$0 \leq A_i \leq M \times b_{s_n n}^{\text{Stop}} \times b_{s'_n n}^{\text{Stop}} \quad (17)$$

where  $M$  is a sufficiently large number and  $b_{sn}^{\text{Stop}}$  equals 1 if  $n$  stops at  $s$ , and 0 otherwise. In addition, constraints (18) and (19) are set to seek a workable passenger assignment ( $\forall r \in R$ ):

$$0 \leq t_{in}^{\text{Wait}} \times A_i, \forall i \in I_r, \forall n \in N_i^{\text{Iti}} \quad (18)$$

$$\sum_{i \in I_r} A_i = \Pi_r^{\text{Pas}} \quad (19)$$

where  $\Pi_r^{\text{Pas}}$  is the number of passengers in OD pair  $r$ . As Constraint (18) states, if  $i$  is used ( $A_i > 0$ ), the waiting

time for the associated  $n$  must be non-negative to ensure that passengers do not miss  $n$ . Constraint (19) requires that the total passenger flows on the itineraries for OD pair  $r$  equal the total number of passengers. However, if no feasible itineraries can be found for an OD pair or if some passengers between an OD pair cannot use any feasible itineraries because of inadequate train capacity, then these passengers are accommodated by a residual itinerary ( $\tilde{i}$ ) with an unlimited capacity. The uncongested cost of  $\tilde{i}$  equals the loss caused by a passenger leaving the train system and opting to take an alternative mode of transport.

2) *Constraint Set for Line Planning  $CS^L$* : In this study, we did not consider terminal changes, and thus we set Constraint (20) as follows ( $\forall n \in N, \forall s \in S_n$ ):

$$\begin{cases} b_{sn}^{\text{Stop}} = b_n^{\text{Sch}} & , \text{ if } s = s_n^{\text{Ori}} \text{ or } s = s_n^{\text{Des}} \\ b_{sn}^{\text{Stop}} \leq b_n^{\text{Sch}} & , \text{ else;} \end{cases} \quad (20)$$

where  $b_n^{\text{Sch}}$  equals 1 if  $n$  is scheduled, and 0 otherwise. If  $n$  is scheduled, it should stop at the origin and destination terminals, and the stopping pattern of the intermediate stations is considered.

Moreover, we determined the stopping pattern and schedule train services based on onboard passenger flows. For determining the stopping pattern, we have ( $\forall n \in N, \forall s \in S_n$ ):

$$b_{sn}^{\text{Stop}} \times \Pi^{\text{Min}} \leq A_{ns}^{\text{BA}} \quad (21)$$

where  $A_{ns}^{\text{BA}}$  is the number of passengers on  $n$  who board or alight at  $s$  ( $A_{ns}^{\text{BA}} = \sum_{r \in R} \sum_{i \in I_r} (b_{ins}^{\text{BA}} \times A_i)$ , where  $b_{ins}^{\text{BA}}$  equals 1 if passengers on  $i$  board or alight from  $n$  at  $s$ , and 0 otherwise); and  $\Pi^{\text{Min}}$  is the minimum number of passengers. Constraint (21) requires that if the number of passengers on  $n$  who board or alight at  $s$  is less than  $\Pi^{\text{Min}}$ , this station must be skipped by  $n$  to reduce the variable cost.

For scheduling train services, we have ( $\forall n \in N, \forall s, s' \in S_n: s' = \zeta_{ns}^{\text{DS}}$ ):

$$b_n^{\text{Sch}} \times \Pi^{\text{Min}} \leq A_{nss'}^{\text{Flow}}, \text{ if } b_n^{\text{FL}} = 1; \quad (22)$$

where  $b_n^{\text{FL}}$  equals 1 if  $n$  is the first train service ( $\sum_{f \in F} \sum_{v \in V_f} \sum_{n' \in N, n' \neq n} b_{vn'n}^{\text{TU}} = 0$ ) or the last train service ( $\sum_{f \in F} \sum_{v \in V_f} b_{vnn}^{\text{TU}} = 1$ ) served by its assigned TU, and 0 otherwise. Constraint (22) requires that if  $n$  is the first train service or the last train service served by its assigned TU, the number of passengers on  $n$  must not be fewer than  $\Pi^{\text{Min}}$ . The model may continue to schedule a connecting train service, regardless of the number of passengers onboard. For example, consider a TU serving train services T1, T2, and T3 in the order T1–T2–T3, where T2 is a connecting train service. If T1 and T3 serve no fewer than  $\Pi^{\text{Min}}$  passengers, all three train services may be scheduled. This would be done if not scheduling T2 would result in the TU being no longer able to serve T1 and T3, as this would mean an additional TU would be required, the operating cost may increase. Hence, such connecting train services would be scheduled.

3) *Constraint Set for Timetabling  $CS^T$* : Timetabling should consider the safety and infrastructure restriction, and thus  $CS^T$  includes the following constraints ( $\forall l, l' \in L; \forall n \in N_l; \forall n' \in N_{l'}$ ):

- Operating time-window constraint

The proposed problem restricts the operating time window according to Constraint (23), as follows:

$$\begin{cases} t_{ns}^{\text{Dep}} \geq T_l^{\text{Start}} \times b_n^{\text{Sch}}, \text{ if } s = s_n^{\text{Ori}} \\ t_{ns}^{\text{Arr}} \leq T_l^{\text{End}} \times b_n^{\text{Sch}}, \text{ if } s = s_n^{\text{Des}}. \end{cases} \quad (23)$$

- Running time constraint

The proposed problem considers the train speed limitation in the form of Constraint (24), which sets the range of running times between two stations ( $\forall s, s' \in S_n: s' = \zeta_{ns}^{\text{DS}}$ ) as follows:

$$T_{lss'}^{\text{Rmin}} \times b_n^{\text{Sch}} \leq t_{ns'}^{\text{Arr}} - t_{ns}^{\text{Dep}} \leq T_{lss'}^{\text{Rmax}} \times b_n^{\text{Sch}}. \quad (24)$$

- Dwell time constraint

If  $n$  stops at  $s$  ( $\forall s \in S_n$ ) to allow passengers to alight and board, the proposed problem limits the range of the dwell time at  $s$  according to Constraint (25), as follows:

$$T_{sl}^{\text{Smin}} \times b_{sn}^{\text{Stop}} \leq t_{ns}^{\text{Dep}} - t_{ns}^{\text{Arr}} \leq T_{sl}^{\text{Smax}} \times b_{sn}^{\text{Stop}}. \quad (25)$$

- Constraints for headway

To ensure safety, we restrict the headway using Constraints (26)–(30), as follows ( $\forall s \in S_n \cap S_{n'}, \forall p \in P_{ns}$ ):

$$t_{ns}^{\text{Arr}} + T^{\text{H}} \leq t_{n's}^{\text{Arr}}, \text{ if } b_{nn's}^{\text{Arr}} = 1; \quad (26)$$

$$t_{ns}^{\text{Dep}} + T^{\text{H}} \leq t_{n's}^{\text{Dep}}, \text{ if } b_{nn's}^{\text{Dep}} = 1; \quad (27)$$

$$b_{nn's}^{\text{Arr}} + b_{nn's}^{\text{Arr}} = b_n^{\text{Sch}} \times b_{n'}^{\text{Sch}}; \quad (28)$$

$$b_{nn's}^{\text{Dep}} + b_{nn's}^{\text{Dep}} = b_n^{\text{Sch}} \times b_{n'}^{\text{Sch}}; \quad (29)$$

$$t_{ns}^{\text{Dep}} + T^{\text{H}} \leq t_{n's}^{\text{Arr}}, \text{ if } b_{np}^{\text{Track}} = b_{n'p}^{\text{Track}} = b_{nn's}^{\text{Arr}} = 1; \quad (30)$$

where  $b_{nn's}^{\text{Arr}}$  equals 1 if  $n$  arrives at  $s$  before  $n'$ , and 0 otherwise; and  $b_{nn's}^{\text{Dep}}$  equals 1 if  $n$  departs from  $s$  before  $n'$ , and 0 otherwise. Constraints (26) and (27) impose the necessary headway ( $T^{\text{H}}$ ) between the arrival and departure times of  $n$  and  $n'$  at  $s$ ; Constraints (28) and (29) apply a logical set of departure and arrival orders for  $n$  and  $n'$  at  $s$ ; and Constraint (30) maintains safety by ensuring that a sufficient time difference exists between  $n$  and  $n'$  if they are operating on the same station track.

- Constraint for station-track assignment

The proposed problem designs the station-track assignment, and thus the following constraint is set ( $\forall s \in S_n$ ):

$$\begin{cases} \sum_{p \in P_{ns}} b_{np}^{\text{Track}} = 0 & , \text{ if } b_n^{\text{Sch}} = 0 \\ \sum_{p \in P_{ns}} b_{np}^{\text{Track}} = 1, & \text{ if } b_n^{\text{Sch}} = 1 \text{ and } b_{sn}^{\text{Stop}} = 0 \\ \sum_{p \in P_{ns}} b_{np}^{\text{Track}} = 1, & \text{ if } b_n^{\text{Sch}} = 1 \text{ and } b_{sn}^{\text{Stop}} = 1. \end{cases} \quad (31)$$

Constraint (31) requires that if  $n$  is not scheduled, no track is assigned to  $n$ ; whereas if  $n$  is scheduled, one track is assigned to  $n$ . In addition, if  $n$  stops at  $s$ , a track equipped with a platform must be assigned to  $n$ .

- Overtaking constraint

Overtaking can only happen at a station which has more than one station track. Hence, we have  $(\forall s, s' \in S_n \cap S_{n'}: s' = \zeta_{ns}^{DS})$ :

$$b_{nn's}^{Dep} - b_{nn's}^{Arr} = 0. \quad (32)$$

Constraint (32) ensures that  $n$  and  $n'$  cannot overtake each other in a section  $(s, s')$ . If  $n$  stops at one station track, a later-arriving train service can use another station track to overtake  $n$  as this is allowed by Constraints (26)–(31). In addition,  $n$  cannot be overtaken at the station that it skips by any other train service, as this is disallowed by Constraints (25)–(31).

#### 4) Constraint Set for TU Scheduling $CS^S$ :

- Scheduled train service constraint

In the proposed problem, TUs are assigned to serve scheduled train services one by one. Hence, we have  $(\forall n \in N)$ :

$$\sum_{f \in F} \sum_{v \in V_f} \sum_{n' \in N} b_{vnn'}^{TU} = b_n^{Sch}; \quad (33)$$

$$\sum_{n' \in N, n' \neq n} b_{vnn'}^{TU} \leq \sum_{n' \in N} b_{vnn'}^{TU}, \quad \forall f \in F, \quad \forall v \in V_f. \quad (34)$$

Constraint (33) ensures that if  $n$  is scheduled ( $b_n^{Sch} = 1$ ), one TU is assigned to it. When the right side of Constraint (34) equals 0,  $v$  does not serve  $n$ ; therefore, the left side of Constraint (34) should also equal 0. When the right side of Constraint (34) equals 1 (i.e.,  $v$  serves  $n$ ), the left side equals 1 or 0. When the left side of Constraint (34) equals 0,  $n$  is the first train service served by  $v$ . When the left side of Constraint (34) equals 1,  $v$  serves one train service directly before serving  $n$ . When Constraint (33) applies, Constraint (34) requires a TU to serve train services one by one.

- TU connection constraint

An improper connection provided by TU  $v$  for scheduled  $n$  and  $n'$  must be prevented, and thus the following constraint is set  $(\forall f \in F, \forall v \in V_f, \forall n, n' \in N)$ :

$$b_{vnn'}^{TU} = 0, \text{ if } \begin{cases} A_{nss'}^{Flow} > \Pi_f^{Cap}, s, s' \in S_n: s' = \zeta_{n's}^{DS} \\ A_{nss'}^{Flow} > \Pi_f^{Cap}, s, s' \in S_{n'}: s' = \zeta_{n's}^{DS} \\ s_n^{Des} \neq s_{n'}^{Ori} \\ f \notin F_s, s \in \{s_n^{Ori}, s_n^{Des}, s_{n'}^{Ori}, s_{n'}^{Des}\} \\ t_{n's}^{Dep} < t_{ns}^{Arr} + T^{LT}, \text{ where } s = s_n^{Des} = s_{n'}^{Ori} \end{cases} \quad (35)$$

Constraint (35) prevents an improper connection being made if  $v$  cannot fulfill the capacity constraint, if the destination terminal of  $n$  and the origin terminal of  $n'$  are different, if the depots serving the terminals of  $n$  and  $n'$  cannot serve TUs of type  $f$ , or if the minimum layover time requirement cannot be met.

- TU reallocation constraints

The proposed problem considers the end-of-day balance, the extra dispatches for which are set based on the distribution of TUs. Thus, we have  $(\forall f \in F; \forall s \in S)$ :

$$\begin{aligned} \Pi_{fss'}^{TR} &\geq 0, \forall s' \in (S-s); \\ \sum_{v \in V_f} \left( \sum_{n \in N_s^{End}} \sum_{n' \in N} b_{vnn'}^{TU} - \sum_{n \in N_s^{Sta}} \sum_{n' \in N} b_{vnn'}^{TU} \right) \end{aligned} \quad (36)$$

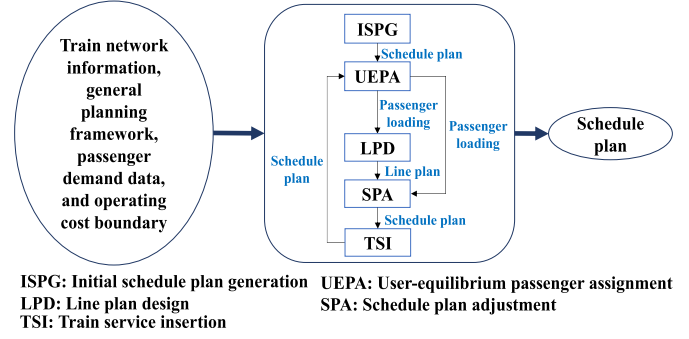


Fig. 3. Flowchart of the heuristic.

$$= \sum_{s' \in (S-s)} (\Pi_{fss'}^{TR} - \Pi_{f's's'}^{TR}). \quad (37)$$

Constraint (36) requires that the extra dispatches  $(\Pi_{fss'}^{TR})$  should be non-negative, and Constraint (37) determines the values of the extra dispatches.  $\sum_{n \in N_s^{End}} \sum_{n' \in N} b_{vnn'}^{TU}$  counts how many times  $v$  ends at  $s$ , and  $\sum_{n \in N_s^{Sta}} \sum_{n' \in N} b_{vnn'}^{TU}$  counts how many times  $v$  starts at  $s$ . If the left side of Constraint (37) equals 0, the number of TUs at the depot of  $s$  at the end-of-day equals that at the beginning-of-day and no extra dispatch is needed, whereas if the left side of (37) is positive, an extra dispatch from  $s$  is needed; otherwise, an extra dispatch to  $s$  is needed.

Note that mathematical transformations (like introducing a sufficiently large value of  $M$ ) can convert Constraints (26)–(31) to linear constraints. But not all constraints can be converted to linear ones (e.g., the UE condition). Thus, the SIM is a non-linear model.

## IV. METHODOLOGY

The proposed SIM, a non-linear model with many variables and complex non-linear constraints, is difficult to solve. Thus, we decompose the SIM into sub-models. In a sub-model, some variables of the SIM are fixed as parameters and set as the values offered by other sub-models, and thus the non-linear constraints can be linearized. For example, the sub-model for schedule plan generation provides timetables to calculate waiting times (i.e.,  $t_{in}^{Wait} = \bar{t}_{in}^{Wait}$ , where *variable* is a parameter which equals a fixed value set for a variable), and non-linear Constraint (18) in the sub-model for passenger assignment can then be a linear constraint as follows:

$$\bar{t}_{in}^{Wait} \times A_i \geq 0, \quad \forall n \in N_i^{Iti}, \quad \forall r \in R, \quad \forall i \in I_r. \quad (38)$$

Thus, the sub-models can be handled by the CPLEX that is commonly used for linear programming (LP) problems [26], [39]. Based on the above discussion, an iterative heuristic is suggested, as shown in Fig.3.

First, the initial schedule plan generation (ISPG) process generates a feasible schedule plan. Next, an iterative process is executed to improve the schedule plan. Each iteration comprises four steps: (a) the user-equilibrium passenger assignment (UEPA) process, which is used to identify the feasible itineraries and compute the passenger flows according to the



given schedule plan; (b) the line plan design (LPD), which determines the number of scheduled train services and designs a stopping plan based on the passenger loading outputted by the UEPA process; (c) the schedule plan adjustment (SPA) process, which uses the results of the UEPA process and the LPD to improve the schedule plan; and (d) the train service insertion (TSI) process, which inserts unscheduled train services into the schedule plan adjusted by the SPA because the adjustment may lower the operating cost and the budget may allow for additional train services to improve the passenger service.

The ISPG process and the LPD determine  $N^{\text{Sch}}$ ,  $N^{\text{Uns}}$ ,  $S_n^{\text{Stop}}$ , and  $S_n^{\text{Skip}}$ , and then  $b_n^{\text{Sch}}$  and  $b_{sn}^{\text{Stop}}$  will be determined by the constraint set  $CS^{\text{LH}}$  as follows:

$$b_n^{\text{Sch}} = \begin{cases} 1, n \in N^{\text{Sch}} \\ 0, n \in N^{\text{Uns}} \end{cases}; \quad (39)$$

$$b_{sn}^{\text{Stop}} = \begin{cases} b_n^{\text{Sch}}, s \in S_n^{\text{Stop}} \\ 0, s \in S_n^{\text{Skip}} \end{cases}; \quad \forall n \in N. \quad (40)$$

The constraint set  $CS^{\text{LH}}$  describes that if train service  $n$  belongs to the set of scheduled train services (i.e.,  $n \in N^{\text{Sch}}$ ),  $b_n^{\text{Sch}}$  should be 1, otherwise 0; and if station  $s$  belongs to the set of stations at which  $n$  stops (i.e.,  $s \in S_n^{\text{Stop}}$ ), the value of  $b_{sn}^{\text{Stop}}$  is determined by  $b_n^{\text{Sch}}$ , otherwise 0.

The ISPG, UEPA, and SPA processes can be considered as sub-models that are developed based on the SIM by a certain modification including fixing some variables and linearizing some constraints. To clarify the specific modification, the details are provided in the following sections.

The iterative process ends when the relative change in  $C^{\text{P}}$  is smaller than the pre-set gap allowance, and the schedule plan with the minimum  $C^{\text{P}}$  is selected as the outputted schedule plan. Otherwise, the next iteration begins. It is possible that a transit paradox [40] may occur. In a transit paradox, the operators implement an adjustment for service improvement that has the unintended effect of worsening the system performance in terms of  $C^{\text{P}}$ . When a transit paradox occurs, the proposed method opts for the schedule plan with the best performance rather than the schedule plan in the final iteration.

Moreover, the line plan for train services is designed based on the information of passenger flows. However, this information may be missing in some cases. For example, the passenger flow on a newly added train service is unknown. To solve this problem, the passenger flow in such a case is assumed to be zero, and the newly added train service is assumed to stop at all passing stations. This line plan is likely to provide more feasible itineraries than a line plan on which newly added train services skip some stations. Furthermore, these newly added train services will use train type  $f^*$ , which has the highest fixed cost and the highest capacity. Thus, we set that train services will not use other train types ( $\forall v \in V_f; \forall n \in N^{\text{New}}; \forall n' \in N$ ):

$$b_{vnn'}^{\text{TU}} = b_{vn'n}^{\text{TU}} = 0, \forall f \in (F - f^*). \quad (41)$$

This assumption ensures that the operating cost remains within  $\varepsilon$ . When the schedule plan is input into the UEPA process, the

TABLE II  
DA-ISPGR ALGORITHM

<b>Step 1.</b> Add all of the train services into $N^{\text{Uns}}$ , and empty $N^{\text{Sch}}$ .
<b>Step 2.</b> Set TU $v = 1$ .
<b>While</b> $C^{\text{O}} < \varepsilon$ and $N^{\text{Uns}}$ is not empty, <b>do</b>
<b>Step 3.</b> Run the train-service assignment to determine which train services are served by $v$ and move these train services from $N^{\text{Uns}}$ to $N^{\text{Sch}}$ .
<b>Step 4.</b> Consider the next TU: $v = v + 1$ .
<b>End</b>

passenger flows can be determined, and TUs can be reassigned according to the onboard passenger flow to minimize the operating cost. Some TUs may be replaced with TUs of a different type, and the fixed costs of these TUs are lower than that of  $f^*$ . Therefore, the operating cost will not exceed  $\varepsilon$  and the modified solution given by the SPA will remain feasible.

### A. ISPG

The ISPG process can be considered as a modified version of the SIM that excludes  $CS^{\text{P}}$  and  $CS^{\text{L}}$ , but adds some more constraints including  $CS^{\text{LH}}$  and Constraints (41) and (42).

$$A_i = \bar{A}_i, \forall r \in R, \forall i \in I_r. \quad (42)$$

As indicated by Constraint (42), passenger flow on itinerary  $i$ ,  $A_i$ , is set as a fixed value ( $\bar{A}_i$ ). Because all train services used in the ISPG process are newly added,  $\bar{A}_i$  is zero according to the aforementioned assumption. The decision variables include  $b_n^{\text{Sch}}$ ,  $b_{sn}^{\text{Stop}}$ ,  $b_{nn's}^{\text{Arr}}$ ,  $b_{nn's}^{\text{Dep}}$ ,  $t_{ns}^{\text{Arr}}$ ,  $t_{ns}^{\text{Dep}}$ ,  $b_{np}^{\text{Track}}$ ,  $b_{vnn'}^{\text{TU}}$ , and  $\Pi_{fss'}^{\text{TR}}$ . Because the passenger flows are unknown and the ISPG process aims to obtain a feasible solution as a starting point for the later processes, which will improve the solution, this study does not set a specific objective for the ISPG process. That is, when the ISPG process finds the first feasible solution, the process can stop. Thus, the ISPG process can be formulated as follows.

min0,  
s.t. Constraint III-A.1, (41), and (42), constraint sets  $CS^{\text{LH}}$ ,  $CS^{\text{T}}$ , and  $CS^{\text{S}}$ .

Although the ISPG process is relatively simpler than the SIM model, it remains difficult to solve the ISPG process directly. Hence, a decomposition approach, DA-ISPGR, is proposed. In the DA-ISPGR approach, the original problem is divided into several subproblems in each of which a TU is used, and the train circulation plan and timetables of its served train services are determined simultaneously.

1) *DA-ISPGR*: The DA-ISPGR algorithm is shown in Table II. If  $N^{\text{Uns}}$  is empty or the insertion of a new TU violates the operating cost constraint, the algorithm stops; otherwise, a new TU is added, and train services are assigned to it. If  $v$  serves more than one train service, these train services can be considered as predecessors ( $n^{\text{Pre}}$ ) or posteriors ( $n^{\text{Post}}$ ). For example, TU 1 in Fig. 2 serves T2 before serving T4, and we designate T2 as  $n^{\text{Pre}}$  and T4 as  $n^{\text{Post}}$ .

The train-service assignment in Table II can then follow an ascending order or a descending order. In an ascending order, a train service is first selected as  $n^{\text{Pre}}$ ; subsequently,



TABLE III  
TRAIN-SERVICE ASSIGNMENT FOR A TRAIN UNIT

<b>Step 1.</b> Add all unscheduled train services into $N^{Uns}$ and all scheduled train services into $N^{Sch}$ .
<b>Step 2.</b> Select a train service as $n^{Post}$ .
<b>Step 3.</b> Generate a timetable for $n^{Post}$ , and move it from $N^{Uns}$ to $N^{Sch}$ .
<b>If the solution is not found or <math>C^O &gt; \varepsilon</math>, then</b>
<b>Step 3a:</b> Remove $n^{Post}$ from $N^{Sch}$ , and return to Step 2.
<b>End</b>
<b>Step 4.</b> Determine predecessors of $n^{Post}$ .
<b>Step 4a:</b> Select a train service as $n^{Pre}$ . If no predecessors are found, end the algorithm; otherwise, go to Step 4b.
<b>Step 4b:</b> Move $n^{Pre}$ from $N^{Uns}$ to $N^{Sch}$ , and generate a timetable for $n^{Pre}$ .
<b>If the solution is found and <math>C^O \leq \varepsilon</math>, then</b>
<b>Step 4c:</b> $n^{Post} = n^{Pre}$ . Return to Step 4a.
<b>else</b>
<b>Step 4d:</b> Move $n^{Pre}$ from $N^{Sch}$ to $N^{Uns}$ . Return to Step 4a.
<b>End</b>

the posterior services are selected one by one. In contrast, in a descending order, a train service is first selected as  $n^{Post}$ ; subsequently, the predecessors are selected one by one. In this study, the train-service assignment follows a descending order, and a train service can be selected for the assignment if (a) it is unscheduled; (b) while looking for a predecessor of  $n^{Post}$ , its destination terminal is the original terminal of  $n^{Post}$ ; and (c) its ideal arrival time at the destination is the latest when compared with those of the other train services that satisfy the two abovementioned rules. The ideal departure and arrival times can be given by the operators based on their previous experience or calculated by a timetabling model (whose solution is denoted the M-MinPC solution) that can minimize  $C^P$  without consideration of TU scheduling [27]. The DA-ISPG algorithm attempts to use information from the M-MinPC solution when selecting the train services for  $N^{Uns}$  because the M-MinPC solution has already minimized the total generalized passenger cost. Alternatively, the train-service assignment and the selection of train services for  $N^{Uns}$  can follow a random order or other specific rules. This can be easily altered in the algorithm. The train-service assignment for a TU is summarized in Table III.

The timetable generation process for train service  $n$  used in the train-service assignment (TTG-TSA) process can be defined as follows:

$$\begin{aligned} \min & \beta^+ \times \max(t_{ns_n^{Ori}}^{Dep} - T_n^{Ideal}, 0) + \beta^- \\ & \times \max(T_n^{Ideal} - t_{ns_n^{Ori}}^{Dep}, 0) + (t_{ns_n^{Des}}^{Arr} - t_{ns_n^{Ori}}^{Dep}) \\ \text{s.t.} & CS^{LH}, CS^T, \text{ and} \end{aligned} \quad (43)$$

$t_{ns}^{Arr} + T^{LT} \leq t_{n's}^{Dep}$ ; if train services  $n$  and  $n'$  are served by the same TU, and  $n$  is the predecessor of  $n'$ ;

$$s = s_n^{Des}; \forall n' \in N^{Sch}; \quad (44)$$

$$\begin{cases} t_{n's}^{Arr} = \bar{t}_{n's}^{Arr}, \forall n' \in N^{Fix}, \forall s \in S_n \\ t_{n's}^{Dep} = \bar{t}_{n's}^{Dep}, \forall n' \in N^{Fix}, \forall s \in S_n \end{cases} \quad (45)$$

where  $T_n^{Ideal}$  is the ideal time departing from the origin terminal  $s_n^{Ori}$  of  $n$  and is set to equal that in the M-MinPC solution;

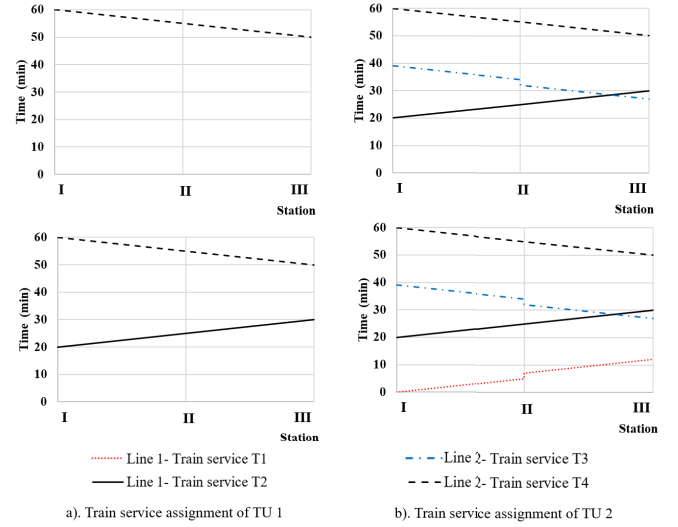


Fig. 4. DA-ISPG for the small-scale example.

$\beta^+$  and  $\beta^-$  are the penalties for the differences between  $T_n^{Ideal}$  and  $t_{ns_n^{Ori}}^{Dep}$ ; and in the TTG-TSA process,  $N^{Fix} = N^{Sch} - n$ .

The difference between this solution and the M-MinPC solution is minimized by objective (43) because the M-MinPC solution suggests that scheduling  $n$  at this spatiotemporal point may reduce the generalized passenger cost. Moreover, the running time of  $n(t_{ns_n^{Des}}^{Arr} - t_{ns_n^{Ori}}^{Dep})$  is minimized to reduce the variable cost. Constraint (44) ensures that the layover time ( $T^{LT}$ ) between two train services served by the same TU is adequate. Except for  $n$ , the train services in  $N^{Sch}$  have already been scheduled, and the corresponding arrival and departure times are known from the previous calculations and are used to set  $\bar{t}_{n's}^{Arr}$  and  $\bar{t}_{n's}^{Dep}$  for Constraint (45). Hence, the TTG-TSA process focuses on finding a feasible timetable for  $n$ .

$C^O$  in Tables II and III is calculated using the train circulation plan generation (TCPG) model as follows:

$$\begin{aligned} \min & C^O \\ \text{s.t.} & CS^S, \text{ constraints(41), (42), and(45)}. \end{aligned}$$

In the TCPG model,  $N^{Fix} = N$ , and  $\bar{t}_{n's}^{Arr}$  and  $\bar{t}_{n's}^{Dep}$  are set based on the result outputted by the TTG-TSA process.

2) *A Simple Example for the DA-ISPG:* To demonstrate the DA-ISPG algorithm, we used the simple example given in Fig.1 and assumed the schedule plan in Fig. 2 to be the M-MinPC solution. First, TU 1 was selected and the train-service assignment was run to determine which train services were served by TU 1. Four train services were incorporated in the set of  $N^{Uns}$ , and T4 had the latest arrival time in the M-MinPC. Thus, T4 was selected as  $n^{Post}$ , and its timetable was generated (as shown in the upper left corner of Fig. 4).

Second, the operating cost of the presented schedule plan was less than the budget, so we added the predecessor of T4. The destinations of T1 and T2 were the same as the original station of T4. In the solution for M-MinPC, the arrival time of T2 was later than that of T1, and T2 was

thus selected.  $T^{LT}$  was assumed to be 20 min, and then the timetable of T2 was generated (as shown in the bottom left corner of Fig. 4). Because of the limitations of  $T^{LT}$  and the operation window, no predecessor of T2 could be found, and thus we turned to TU 2 to generate its circulation plan. If adding an additional TU results in an operating cost that exceeds the budget, the DA-ISPG stops. Otherwise, train services are selected for TU 2. Accordingly, T3 was first selected and then T1 was selected. As this example shows, the operating cost is checked whenever a new TU or train service is added, such that the outputted schedule plan can be easily controlled to ensure that the operating cost constraint holds.

### B. UEPA

The UEPA process focuses on the passenger assignment, so the timetable is fixed based on the output of the ISPG or SPA processes, and used to calculate the fixed values for the waiting time and the uncongested cost. Furthermore, the respective constraints in the SIM can be deleted to form the model used in the UEPA process. That is, the UEPA process can be expressed as a problem that identifies the  $A_i$ ,  $\tilde{c}_r$ , and  $c_{ns's'}^{Sur}$  subject to  $CS^P$ .

$$\begin{aligned} & \min C^P, \\ & \text{s.t. Constraint set } CS^P. \end{aligned}$$

This problem can be formulated as an LP model after the mathematical transformations and solved by the column-generation method, as described in previous studies [27], [37], [38]. The column-generation method starts by solving the LP model with the set of the optimal itineraries only. Then iteratively, itineraries that have the potential to reduce  $C^P$  are added. If adding new itineraries would no longer reduce  $C^P$ , the procedure terminates.

### C. LPD

The LPD determines the values of  $b_n^{Sch}$  and  $b_{sn}^{Stop}$  ( $\forall s \in S_n; \forall n \in N$ ) in the SIM and thus other processes can use  $b_n^{Sch}$  and  $b_{sn}^{Stop}$  as fixed parameters to calculate values of other variables in the SIM. The LPD sets the  $N^{Sch}$ ,  $N^{Uns}$ ,  $S_n^{Stop}$ , and  $S_n^{Skip}$  based on the passenger assignment outputted by the UEPA process, constraint set  $CS^{LH}$ , and Constraints (21) and (22). The LPD may cause some itineraries become infeasible because the used train services are not scheduled or the used stations are skipped, and passengers who choose these itineraries in this UEPA process will be assigned to another feasible itinerary in the next UEPA process.

### D. SPA

Based on the SIM, the SPA process fixes  $A_i$  ( $\forall r \in R$ ,  $\forall i \in I_r$ ),  $b_n^{Sch}$ , and  $b_{sn}^{Stop}$  ( $\forall s \in S_n; \forall n \in N$ ) according to the output of the UEPA process and the design of the LPD, respectively. Furthermore, the SPA process excludes  $CS^P$  except Constraint (18) which ensures that the itineraries used

in the UEPA process are feasible in the SPA solution. Thus, the SPA process can be formulated as follows.

$$\begin{aligned} & \min C^P, \\ & \text{s.t. Constraint III-A.1, (18), and (42),} \\ & \quad \text{constraint sets } CS^{LH}, CS^T, \text{ and } CS^S. \end{aligned}$$

The SPA adjusts the timetable and the train circulation plan simultaneously, which is a NP-hard problem. To increase the computation speed, the SPA process first runs the timetable adjustment (TTA) model which adjusts the timetable with keeping the train circulation plan, and then uses the TCPG model to check whether the train circulation plan can be further improved to reduce  $C^O$  after the adjustment. The TTA model is formulated as follows.

$$\begin{aligned} & \min C^P, \\ & \text{s.t. } CS^{LH}, CS^T, \text{ constraints (18), (42), (44), and} \\ & \quad c^{Var} \times \sum_{n \in N, s=s_n^{Ori}, s'=s_n^{Des}} (t_{ns'}^{Arr} - t_{ns}^{Dep}) + C^{Fixed-L} \\ & \quad + C^{TR-L} \leq \varepsilon. \end{aligned} \quad (46)$$

The number of used TUs and the respective train circulation paths are fixed. Thus,  $C^{Fixed}$  and  $C^{TR}$  equal the values in the last iteration ( $C^{Fixed-L}$  and  $C^{TR-L}$ , respectively). However, the train running time and layover time can be changed, so Constraint (46) requires that the changes of train running time will not violate the operating cost constraint (III-A.1). The TTA model decides values of  $b_{nn's'}^{Arr}$ ,  $b_{nn's'}^{Dep}$ ,  $t_{ns}^{Arr}$ ,  $t_{ns}^{Dep}$ , and  $b_{np}^{Track}$ . It can be solved by a decomposition approach for train timetable adjustment (DA-TTA).

The DA-TTA approach decomposes the original problem into several subproblems, and a timetable is generated for each line. To avoid the spatiotemporal conflict between line  $l'$  and other lines, the subproblem for  $l'$  fixes other lines and focuses on generating a timetable for  $l'$ . That is, a subproblem for  $l'$  can be formulated as a new model (M-DA-TTA) that modifies the TTA model by adding Constraint (45) in which  $N^{Fix} = N - N_{l'}$  and the fixed values  $\tilde{t}_{n's}^{Arr}$  and  $\tilde{t}_{n's}^{Dep}$  are obtained in the previous calculation. After each subproblem, the TCPG model is used to check whether the train circulation plan could be further improved to reduce  $C^O$ . If it could, the updated train circulation plan is used for a further adjustment. In contrast, if the DA-TTA decomposed the problem in terms of TUs, the train circulation plan would be fixed and thus TUs would be selected one-by-one. That is, the train circulation plan could not be updated with a better plan during the adjustment. Thus, the DA-TTA was set to decompose the problem in terms of line. The DA-TTA algorithm is given in Table IV.

Although the DA-TTA approach cannot always arrive at the global optimum, it reduces computation time because the M-DA-TTA model has a smaller solution set and can thus be solved in a considerably shorter time. Moreover, the DA-TTA approach improves the solution iteratively until a local optimum is found.

### E. TSI

The TSI process aims to add some new train services into the output of the SPA process under the operating cost

TABLE IV  
DA-TTA ALGORITHM

```

While the gap between the latest two iterations' objectives is not
smaller than the gap allowance, do
  for each  $l'$  in  $L$  do
    Step 1. Run the M-DA-TTA model for  $l'$ .
    Step 2. Update the recorded values of the variables of  $l'$ .
  End
End

```

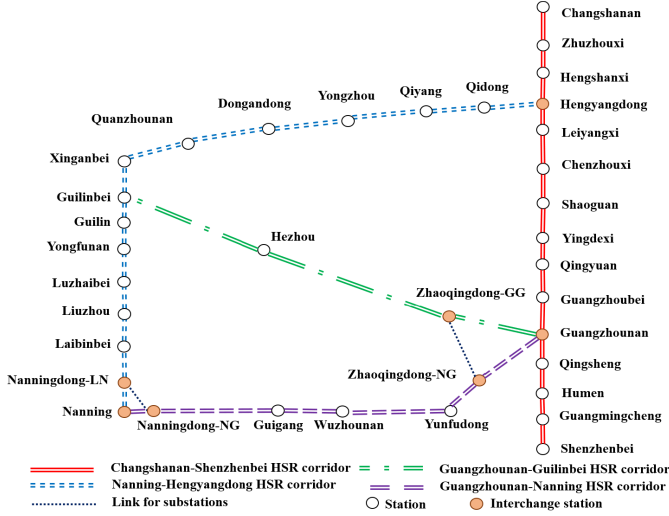


Fig. 5. Abstracted network of south china's HSR.

constraint. When fixing the timetable for scheduled train services, the TSI process adds unscheduled train services into  $N^{\text{New}}$  one by one and identifies a feasible solution by using the TTG-TSA process and the TCPG model.

## V. NUMERICAL EXAMPLES

The computation described in Section V was completed on a personal computer equipped with an Intel i7-4790U central processing unit (3.60 GHz). The settings were applied: (a)  $T^{\text{LT}} = 15$  min; (b)  $T^{\text{H}} = 4$  min; (c)  $c^{\text{Var}} = 20$  O-units/min, where O-unit denotes the unit operating cost; (d)  $\Pi^{\text{Min}} = 1$  passenger; (e) the gap allowance for stopping the CPLEX and iteration was 0.1%; and (f) the penalty for schedule differences was set to  $\beta^+ = 2\beta^- = 0.02$ . Based on daily observations and previous research [40], [41], we set the same weight coefficients for all passengers ( $\forall r \in \mathbf{R}$ ), as follows: (a)  $w_{r1} = 1.0$ ; (b)  $w_{r2} = 1.8$ ; (c)  $w_{r3} = 2.0$ ; (d)  $w_{r4} = 10.0$  min; (e)  $w_{r5} = 1.0$ ; and (f)  $w_{r6} = 2.5$  min/yuan.

Thirty-four main stations were selected from the HSR network in South China (Fig. 5). Nanningdong and Zhaoqingdong stations have two yards, one for each of the two HSR corridors they serve; accordingly, one yard was modeled as a substation as stated in Section III. In this manner, a 36-station abstract network comprising four main HSR corridors was built in which all depots could serve all train types. The network was used to test the heuristic from three aspects, the efficiency, optimality, and the application to a large-size problem, in the following sections.

TABLE V  
DETAILS OF THE DIFFERENT CASES

Fixing $\Pi^{\text{OD}}$ , $\Pi^{\text{Pas}}$ , and $\varepsilon$ as 1, 200, and 30,000 O-units, respectively						
Case	Scale factor	Maximum number of scheduled train services	Problem size		Gap in $C^{\text{P}}$ (%)	Gap in $C^{\text{Unc}}$ (%)
			Number of variables	Number of constraints		
1	1	8	12,043	58,806	0.00	0.00
2	2	16	49,212	261,579	0.00	0.00
3	3	24	111,519	609,984	0.00	0.00
4	4	32	198,964	1,105,461	0.00	0.00
5	5	40	311,547	1,749,450	0.00	0.00

Fixing the scale factor and $\varepsilon$ as 1 and 18,000 O-units, respectively						
Case	$\Pi^{\text{OD}}$	$\Pi^{\text{Pas}}$	Problem size		Gap in $C^{\text{P}}$ (%)	Gap in $C^{\text{Unc}}$ (%)
			Number of variables	Number of constraints		
6	5	200	20,239	268,608	0.00	0.00
7	10	200	34,276	664,020	0.00	0.00
8	15	200	43,308	919,355	0.00	0.00
9	20	200	53,308	1,199,875	0.00	0.00
10	25	200	65,229	1,473,548	0.00	0.00
11	25	225	65,229	1,473,548	5.69	0.67
12	25	250	65,229	1,473,548	5.69	1.19

Note: Aside from the information of running and dwell times collected from the official website of China Railway, other data were assumed. Eight lines were initially set, and each line ran one train service from a corridor's one end to its other end. The number of train services on a line can be increased by multiplying it with a scale factor. Two train types with different capacities (800 and 1000 passengers) and fixed costs (600 and 610 O-units) were assumed. We randomly generated the OD pairs. The uncongested cost of a residual itinerary was 1500 min. Moreover, gaps in this table were computed as follows:  $\text{Gap} = \frac{\text{Heuristic-BBA}}{\text{Heuristic}}$ .

## A. Optimality and Efficiency Test

### 1) Problem Setting of the Optimality and Efficiency Test:

The model with the UE condition is not a convex problem, and a TTP is already an NP-hard problem whose globally optimal solution is difficult to be found. Thus, we transformed this problem into a simplified problem that relaxes the train capacity constraint (when the train capacity is sufficiently large that the passengers always select the itinerary with the lowest uncongested cost):

$$\begin{aligned}
 & \min C^{\text{P}}, \\
 & \text{s.t. Constraint III-A.1, constraint sets } CS^{\text{P}}, CS^{\text{L}}, \\
 & \quad CS^{\text{T}}, \text{ and } CS^{\text{S}}, \text{ and} \\
 & \quad \Pi_f^{\text{Cap}} = M, f \in \mathbf{F}. \tag{47}
 \end{aligned}$$

Constraint (47) allows the non-linear constraints to be linearized even if the objective is quadratic. The simplified problem was thus solved using the branch-and-bound algorithm (BBA) via CPLEX, and  $C^{\text{P}}$  in its optimal solution was the lower bound for the studied problem. Thus, the BBA was selected as the method against which to benchmark the heuristic. Different cases were set to test the heuristic, as shown in Table V.

2) *Optimality Test:* Table V shows the results. We explored the factors that influenced the gaps in  $C^{\text{P}}$  as follows.

First, the effect of problem size was tested. To increase the problem size, the number of train services running on a line was increased in cases 1–5, and the number of OD pairs ( $\Pi^{\text{OD}}$ ) was increased in cases 6–10. The gaps in  $C^{\text{P}}$  and the



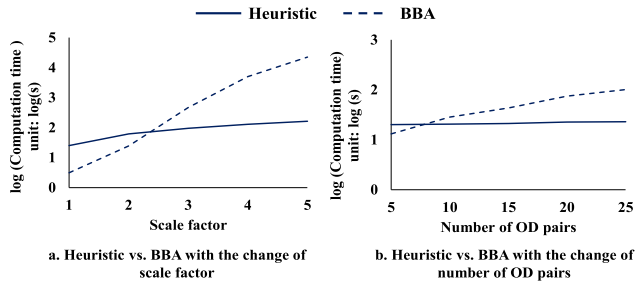


Fig. 6. Trend of computation times with increase in problem size.

total uncongested cost ( $C^{\text{Unc}}$ ) in these cases illustrated that the heuristic maintained good solution quality as the problem size increased. Second, the influence of train capacity was analyzed in cases 10–12, in which the number of passengers in an OD pair ( $I^{\text{Pas}}$ ) was increased. As indicated by the results, the gaps in  $C^{\text{P}}$  and  $C^{\text{Unc}}$  were affected by the increase in  $I^{\text{Pas}}$  because a higher value of  $I^{\text{Pas}}$  implies a higher probability that a greater number of train services do not have sufficient seats. When some passengers could not use the itinerary with the lowest uncongested cost because of insufficient train capacity, the studied problem assigned these passengers to an alternative or artificial itinerary, whereas the simplified problem continued to assign all passengers to the itinerary with the lowest uncongested cost. Both the alternative and artificial itineraries usually have higher uncongested costs, and therefore, the gap in  $C^{\text{Unc}}$  will not be zero. In cases 11 and 12, the congested cost could not have been zero, and the gaps in  $C^{\text{P}}$  would thus have been larger than the gaps in  $C^{\text{Unc}}$ . Overall, the gaps in  $C^{\text{Unc}}$  and  $C^{\text{P}}$  in Table V are below 2% and 6%, respectively. Hence, the test results demonstrate that the solutions obtained using the heuristic tend to be close to the global optima of the simplified problems, indicating that the heuristic can generate solutions similar to the global optima of the studied problems.

3) *Computation Time Test*: The computation time is generally affected by the number of train services on the network for scheduling and the number of OD pairs. Therefore, when comparing the computation times of the BBA and the heuristic, we selected cases 1–5 and cases 6–10 to test the influences of these two factors. In these cases, the BBA found the global optimum, and the heuristic obtained solutions with the same objective values. Fig. 6 shows the log<sub>10</sub> values of the computation times for clarity. The computation times of the BBA were shorter than those of the heuristic when the problem size was small. The BBA tackled only one ILP problem, whereas the heuristic tackled several ILP problems, one of which was for a possible train service assignment to a TU. However, Fig. 6 shows that the computation time of the BBA increased more quickly than that of the heuristic as the problem size increased, indicating the abilities of the heuristic to solve large problems are better than those of the BBA.

## B. Large-Scale Applicability

1) *Problem Setting of a Large-Scale Example*: Between 06:15 and 23:30, 240 high-speed train services run on the

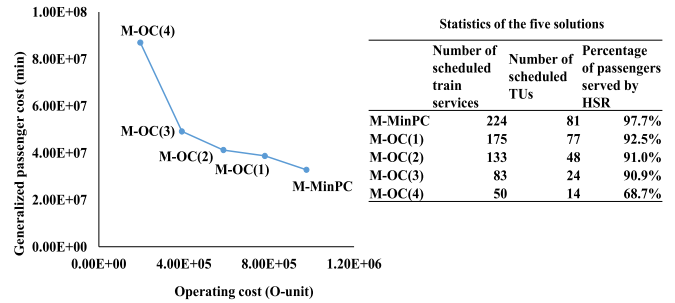


Fig. 7. Solutions for the south china HSR network.

network that was collected from the official website of China Railway (<http://www.12306.cn/mormhweb/>). Because of differences in zone attractiveness and trip time restrictions, 9076 OD pairs were generated. As passenger volume data were unavailable, passenger volumes were set based on daily observation. The uncongested cost of a residual itinerary was 3000 min. Additionally, because real TU information was unavailable, we assumed that there were three train types (Type 1, Type 2, and Type 3) with different capacities (1000, 1300, and 1600 passengers, respectively) and fixed costs (600, 800, and 1000 O-units, respectively). The operating budget can be set based on previous operational experience and the allowable resources. Alternatively, a range of budget levels can be set and analyzed to find a proper solution that is in keeping with the allowable resources. We demonstrated this method with the following example. First, we ran the timetabling model suggested by Xie *et al.* [27] to obtain a timetable minimizing the passenger cost, which we denoted the M-MinPC timetable. Next, we ran the TCPG model to compute the minimum operating cost ( $C^{\text{O-MinPC}}$ ) to support the M-MinPC timetable. We then set four operating cost boundaries—80% of  $C^{\text{O-MinPC}}$ , 60% of  $C^{\text{O-MinPC}}$ , 40% of  $C^{\text{O-MinPC}}$ , and 20% of  $C^{\text{O-MinPC}}$ —and denoted their respective models M-OC(1), M-OC(2), M-OC(3), and M-OC(4).

2) *Solutions of the Large-Scale Example*: Fig. 7 presents the best-found solutions that we determined for the M-MinPC timetable model and each of the above four models (i.e., M-MinPC, M-OC(1), M-OC(2), M-OC(3), and M-OC(4)), and related statistical details.

It can be seen that as the operating budget decreased, the generalized passenger cost initially increased slowly and then rapidly. This trend is attributable to the fact that the decrease in the operating budget led to a decrease in the numbers of scheduled train services and scheduled TUs, which resulted in fewer passengers being served. Setting a higher operating boundary generally enabled a solution to be found that had a lower generalized passenger cost than that for lower operating boundaries. However, operators should not necessarily adopt a higher operating boundary, as extra investment may occasionally fail to generate a proportionally higher return. For example, the solution for M-OC(1) required a 33.33% increase in the operating cost relative to the solution for M-OC(2), but the corresponding reduction in the generalized passenger cost (for the solution for M-OC(1) corresponding to that for M-OC(2)) was only 6.08%. The latter result accounts for the

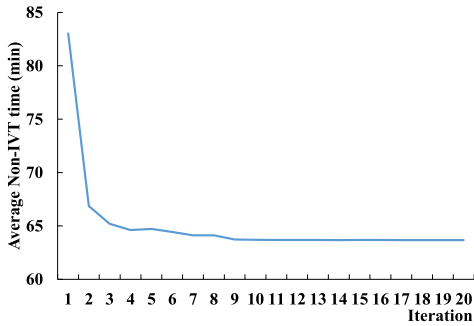


Fig. 8. Average Non-IVT time in each outer iteration of M-OC(2).

nearly horizontal line linking these two solutions in Fig. 7. Thus, the solution for M-OC(2) appeared to be superior to that for M-OC(1). Therefore, we suggest that if a practical operation does not have a hard budget, a range of budget levels should be set and their respective solutions determined and compared, as this will enable a proper solution to be found that results in a relatively low passenger cost and operating cost.

3) *Analysis of M-OC(2)*: The proposed heuristic is an iterative method. Feedback from the passenger assignment is generated in each outer iteration and used in the SPA, which has an inner iteration to improve the schedule plan to reduce the passenger costs. For example, in M-OC(2), there was a decreasing trend in both the total generalized passenger cost and the average non-IVT time (i.e., the sum of the waiting and walking time). The average non-IVT time in the first 20 iterations of the M-OC(2) without a pre-set criterion is shown in Fig. 8. The average non-IVT time decreased by 30.43%. After eight outer iterations, the gap of the total generalized passenger cost decreased to 0.1% (i.e., the pre-set criterion for stopping was met), and the heuristic stopped. However, the average non-IVT time only converged at the 10th iteration; to reach this convergence point, we could use a lower pre-set criterion or allow more outer iterations to run. On average, one outer iteration took 1.52 h, and 29.23% of the computation time was used to optimize the schedule plan. Generally, the ISPG and SPA processes were completed within 3.42 min and 26.66 min, respectively. These long computation times resulted from the complexity of the passenger assignment. In a future study, we will consider how to improve the efficiency of passenger assignment.

To demonstrate that our model and method can manage different types of trains, we changed some settings. We added Type 4 trains, which had the same capacity and fixed cost as Type 1 trains, but we assumed that Type 1 and Type 4 trains needed different equipment in depots. All depots were assumed to serve all train types, except for the depot serving Shenzhenbei Station, which was assumed to serve only Type 2, Type 3, and Type 4 trains. We reran M-OC(2) to identify a new solution, which we denoted Solution 2; thus, we denoted the original solution of M-OC(2) as Solution 1. Both solutions had a similar number of used TUs (gap  $\approx 2.08\%$ ), average running times per TU (gap  $\approx 2.20\%$ ), and average non-IVT

time (gap  $\approx 1.10\%$ ). Solution 1 used 39 Type 1 TUs and Solution 2 used 10 Type 1 TUs and 27 Type 4 TUs. The depot serving Shenzhenbei Station had a unique structure: Type 4 vehicles were used in Solution 2 to take over the work of the Type 1 vehicles that were related to this depot and used in Solution 1.

4) *Comparison of M-OC(1), M-OC(2), and M-MinPC*: The M-MinPC solution is obtained via the traditional hierarchical approach which first designs a timetable and then find a train circulation plan to support the timetable. In this approach, a timetable which can fully use the network capacity to provide a good service to passengers, but the respective operating cost may be too high. For example, if the operators want to control the operating cost within 80% of  $C^{O-\text{MinPC}}$ , the operating cost of the M-MinPC solution exceeds the limit by 25%. Thus, the traditional hierarchical approach usually includes a further step to delete train services from the M-MinPC solution to meet the requirement of the operating cost.

Our approach integrates the consideration of operating cost into planning, so that an outputted solution meets operating cost requirements and attempts to maintain service quality. For example, when the operating budget was 80% of  $C^{O-\text{MinPC}}$ , our approach generated a solution that met the requirement (i.e., the solution for M-OC(1)). Compared with the solution for M-MinPC, in the solution for M-OC(1), the number of served passengers was decreased by 6.5%, and the average waiting time was increased by 8.9 min. Similarly, when the operating budget was 60% of  $C^{O-\text{MinPC}}$ , our approach generated the solution for M-OC(2). Compared with the solution for M-MinPC, in the solution for M-OC(2), the number of served passengers was decreased by 8.1%, and the average waiting time was increased by 19.9 min. This decrease in service quality was due to the reduction in the operating cost. However, given that the latter was 40% ( $1 - 60\%$ ), this decrease in service quality may be acceptable in real-life operations.

This comparison indicates the advantages of our approach to generate a schedule plan to balance the needs of operators and passengers. In summary, our large-scale example demonstrates that our approach can be applied to a practical-sized train problem.

## VI. CONCLUSION

A schedule-based model is proposed in this study to solve a new integrated problem that comprises the LPP, TTP, VSP, and the UE passenger assignment problem. To fulfill the requirements of operators and passengers, the model minimizes the generalized passenger costs under the operating cost constraint. To obtain high-quality solutions within a reasonable time, a new heuristic was developed. The model and algorithm were applied to numerical examples. The results show that the heuristic can obtain the similar results as the BBA at considerably faster computation speeds. Furthermore, the algorithm was able to handle an integrated problem that involved 9076 OD pairs, 240 high-speed train services, and 36 stations forming a four-corridor HSR network. Therefore,

the model and the algorithm can be used for a practical-sized train planning problem.

However, in this study, we made some assumptions that may not hold in real applications. For example, we assumed that double tracks were used between stations, TUs and crews were sufficient to support schedule plans, and passengers can always find a feasible access/egress mode in time. In future studies, we will relax these assumptions to construct a planning tool that can be applied in various practical scenarios.

## REFERENCES

- [1] L. Zhang, S. Wang, and X. Qu, "Optimal electric bus fleet scheduling considering battery degradation and non-linear charging profile," *Transp. Res. E, Logistics Transp. Rev.*, vol. 154, Oct. 2021, Art. no. 102445.
- [2] Y. Wang, S. Zhu, A. D'Ariano, J. Yin, J. Miao, and L. Meng, "Energy-efficient timetabling and rolling stock circulation planning based on automatic train operation levels for metro lines," *Transp. Res. C, Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103209.
- [3] Y. I. Silva-Soto and O. J. Ibarra-Rojas, "Timetabling with flexible frequencies to synchronise groups of bus lines at common stops," *Transportmetrica A, Transp. Sci.*, vol. 17, no. 4, pp. 978–1001, Dec. 2021.
- [4] L. K. Nielsen, L. Kroon, and G. Maróti, "A rolling horizon approach for disruption management of railway rolling stock," *Eur. J. Oper. Res.*, vol. 220, no. 2, pp. 496–509, Jul. 2012.
- [5] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints," *Transp. Res. B, Methodol.*, vol. 76, pp. 117–135, Jun. 2015.
- [6] V. Guihaire and J.-K. Hao, "Transit network design and scheduling: A global review," *Transp. Res. A, Policy Pract.*, vol. 42, no. 10, pp. 1251–1273, Dec. 2008.
- [7] J. Parbo, O. Nielsen, and C. Prato, "Passenger perspectives in railway timetabling: A literature review," *Transp. Rev.*, vol. 36, no. 4, pp. 500–526, 2016.
- [8] S. Bunte and N. Kliewer, "An overview on vehicle scheduling models," *Public Transp.*, vol. 1, no. 4, pp. 299–317, 2009.
- [9] S. Zhao, H. Yang, and Y. Wu, "An integrated approach of train scheduling and rolling stock circulation with skip-stopping pattern for urban rail transit lines," *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103170.
- [10] M. Michaelis and A. Schöbel, "Integrating line planning, timetabling, and vehicle scheduling: A customer-oriented heuristic," *Public Transp.*, vol. 1, pp. 211–232, Aug. 2009.
- [11] G. Laporte, F. A. Ortega, M. A. Pozo, and J. Puerto, "Multi-objective integration of timetables, vehicle schedules and user routings in a transit network," *Transp. Res. B, Methodol.*, vol. 98, pp. 94–112, Apr. 2017.
- [12] X. Dou, Q. Meng, and K. Liu, "Customized bus service design for uncertain commuting travel demand," *Transportmetrica A, Transp. Sci.*, vol. 17, no. 4, pp. 1405–1430, Dec. 2021.
- [13] P. Mo, L. Yang, Y. Wang, and J. Qi, "A flexible metro train scheduling approach to minimize energy cost and passenger waiting time," *Comput. Ind. Eng.*, vol. 132, pp. 412–432, Jun. 2019.
- [14] L. Kroon, G. Maróti, and L. Nielsen, "Rescheduling of railway rolling stock with dynamic passenger flows," *Transp. Sci.*, vol. 49, no. 2, pp. 165–184, May 2015.
- [15] Ş. Yıldırım and B. Yıldız, "Electric bus fleet composition and scheduling," *Transp. Res. Part C: Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103197.
- [16] V. Cacchiani, J. Qi, and L. Yang, "Robust optimization models for integrated train stop planning and timetabling with passenger demand uncertainty," *Transp. Res. B, Methodol.*, vol. 136, pp. 1–29, Jun. 2020.
- [17] J. Xie, S. C. Wong, S. Zhan, S. M. Lo, and A. Chen, "Train schedule optimization based on schedule-based stochastic passenger assignment," *Transp. Res. E, Logistics Transp. Rev.*, vol. 136, Apr. 2020, Art. no. 101882.
- [18] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds," *Transp. Res. B, Methodol.*, vol. 41, no. 3, pp. 320–341, Mar. 2007.
- [19] A. Schöbel, "An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation," *Transp. Res. C, Emerg. Technol.*, vol. 74, pp. 348–365, Jan. 2017.
- [20] A. Nuzzolo and U. Crisalli, "The schedule-based modeling of transportation systems: Recent developments," in *Schedule-based Modeling of Transportation Networks*. Boston, MA, USA: Springer, 2009, pp. 1–26.
- [21] J. Xie, S. C. Wong, and S. M. Lo, "Three extensions of Tong and Richardson's algorithm for finding the optimal path in schedule-based railway networks," *J. Adv. Transp.*, vol. 2017, Jan. 2017, Art. no. 9216864.
- [22] X. Wu, H. Dong, and C. K. Tse, "Multi-objective timetabling optimization for a two-way metro line under dynamic passenger demand," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4853–4863, Aug. 2021.
- [23] Y.-H. Chang, C.-H. Yeh, and C.-C. Shen, "A multiobjective model for passenger train services planning: Application to Taiwan's high-speed rail line," *Transp. Res. B, Methodol.*, vol. 34, no. 2, pp. 91–106, 2000.
- [24] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach," *Transp. Res. B, Methodol.*, vol. 67, pp. 235–263, Sep. 2014.
- [25] Y. Yue, S. Wang, L. Zhou, L. Tong, and M. R. Saat, "Optimizing train stopping patterns and schedules for high-speed passenger rail corridors," *Transp. Res. C, Emerg. Technol.*, vol. 63, pp. 126–146, Feb. 2016.
- [26] L. P. Veelenturf, L. G. Kroon, and G. Maróti, "Passenger oriented railway disruption management by adapting timetables and rolling stock schedules," *Transp. Res. C, Emerg. Technol.*, vol. 80, pp. 133–147, Jul. 2017.
- [27] J. Xie, S. Zhan, S. C. Wong, and S. M. Lo, "A schedule-based timetable model for congested transit networks," *Transp. Res. C, Emerg. Technol.*, vol. 124, Mar. 2021, Art. no. 102925.
- [28] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transp. Res. C, Emerg. Technol.*, vol. 36, no. 11, pp. 212–230, 2013.
- [29] X. Yang, B. Ning, X. Li, and T. Tang, "A two-objective timetable optimization model in subway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1913–1921, Oct. 2014.
- [30] H. Sun, J. Wu, H. Ma, X. Yang, and Z. Gao, "A bi-objective timetable optimization model for urban rail transit based on the time-dependent passenger volume," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 604–615, Feb. 2019.
- [31] Y. Bai, Q. Hu, T. K. Ho, H. Guo, and B. Mao, "Timetable optimization for metro lines connecting to intercity railway stations to minimize passenger waiting time," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 79–90, Jan. 2021.
- [32] X. Sun, S. Zhang, H. Dong, Y. Chen, and H. Zhu, "Optimization of metro train schedules with a dwell time model using the Lagrangian duality theory," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1285–1293, Jun. 2015.
- [33] S. Zhan, S. C. Wong, P. Shang, and S. M. Lo, "Train rescheduling in a major disruption on a high-speed railway network with seat reservation," *Transportmetrica A, Transp. Sci.*, pp. 1–36, Feb. 2021, doi: 10.1080/23249935.2021.1877369.
- [34] X. Guo, J. Wu, H. Sun, R. Liu, and Z. Gao, "Timetable coordination of first trains in urban railway network: A case study of Beijing," *Appl. Math. Model.*, vol. 40, nos. 17–18, pp. 8048–8066, Sep. 2016.
- [35] M. B. Khan and X. Zhou, "Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 81–89, Mar. 2010.
- [36] S. Zhan, S. C. Wong, P. Shang, Q. Peng, J. Xie, and S. M. Lo, "Integrated railway timetable rescheduling and dynamic passenger routing during a complete blockage," *Transp. Res. B, Methodol.*, vol. 143, pp. 86–123, Jan. 2021.
- [37] W. Y. Szeto, Y. Jiang, K. I. Wong, and M. Solayappan, "Reliability-based stochastic transit assignment with capacity constraints: Formulation and solution method," *Transp. Res. C, Emerg. Technol.*, vol. 35, pp. 286–304, Oct. 2013.
- [38] G. Xu, H. Yang, W. Liu, and F. Shi, "Itinerary choice and advance ticket booking for high-speed-railway network services," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 82–104, Oct. 2018.
- [39] Z. Hou, H. Dong, S. Gao, G. Nicholson, L. Chen, and C. Roberts, "Energy-saving metro train timetable rescheduling model considering ATO profiles and dynamic passenger flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2774–2785, Jul. 2019.



- [40] W. Y. Szeto and Y. Jiang, "Transit assignment: Approach-based formulation, extragradient method, and paradox," *Transp. Res. B, Methodol.*, vol. 62, pp. 51–76, Apr. 2014.
- [41] C. O. Tong and A. J. Richardson, "A computer model for finding the time-dependent minimum path in a transit system with fixed schedules," *J. Adv. Transp.*, vol. 18, no. 2, pp. 145–161, Mar. 1984.



**Jiemin Xie** received the Ph.D. degree in transportation engineering from The University of Hong Kong, Hong Kong.

She is currently an Assistant Professor with the School of Intelligent Systems Engineering, Sun Yat-sen University. Her current research interests include analysis of passenger behaviors and transit planning.



**Shuguang Zhan** received the Ph.D. degree in transportation planning and management from Southwest Jiaotong University, Chengdu, China.

He is currently a Professor with the School of Automotive and Transportation Engineering, Hefei University of Technology. His current research interests include analysis of passenger behaviors and railway planning.



**Sze Chun Wong** received the Ph.D. degree in transport studies from University College London, London, U.K.

He is currently a Chair Professor with the Department of Civil Engineering, The University of Hong Kong. His current research interests include dynamic transit assignment problems, traffic equilibrium problems, optimization of traffic signal settings, road safety, and urban taxi services.



**Siu Ming Lo** received the Ph.D. degree in architecture from The University of Hong Kong, Hong Kong.

He is currently a Professor with the Department of Architecture and Civil Engineering, City University of Hong Kong. His current research interests include urban and land use planning, spatial planning for pedestrian movement and evacuation modeling, and building development.