RESEARCH ARTICLE

A practical model for the train-set utilization: The case of Beijing-Tianjin passenger dedicated line in China

Yu Zhou^{1☯}, Leishan Zhou^{1☯}, Yun Wang^{1☯*}, Xiaomeng Li^{1‡}, Zhuo Yang^{2‡}

1 MOE Key Laboratory for Urban Transportation Complex System Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, **2** Department of Civil, Environmental, and Infrastructure Engineering, Volgenau School of Engineering, George Mason University, Fairfax, Virginia, United States of America

☯ These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

* 13114248@bjtu.edu.cn

OPEN ACCESS

Citation: Zhou Y, Zhou L, Wang Y, Li X, Yang Z (2017) A practical model for the train-set utilization: The case of Beijing-Tianjin passenger dedicated line in China. PLoS ONE 12(5): e0175698. <https://doi.org/10.1371/journal.pone.0175698>

Editor: Tieqiao Tang, Beihang University, CHINA

Received: October 21, 2016

Accepted: March 28, 2017

Published: May 10, 2017

Copyright: © 2017 Zhou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work is financially supported by “China Railways Corporation” (No. 2016X005-D), and “the Fundamental Research Funds for the Central Universities” (2015YJS094). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: This work received financial support from the “China Railways Corporation”. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Abstract

As a sustainable transportation mode, high-speed railway (HSR) has become an efficient way to meet the huge travel demand. However, due to the high acquisition and maintenance cost, it is impossible to build enough infrastructure and purchase enough train-sets. Great efforts are required to improve the transport capability of HSR. The utilization efficiency of train-sets (carrying tools of HSR) is one of the most important factors of the transport capacity of HSR. In order to enhance the utilization efficiency of the train-sets, this paper proposed a train-set circulation optimization model to minimize the total connection time. An innovative two-stage approach which contains segments generation and segments combination was designed to solve this model. In order to verify the feasibility of the proposed approach, an experiment was carried out in the Beijing-Tianjin passenger dedicated line, to fulfill a 174 trips train diagram. The model results showed that compared with the traditional Ant Colony Algorithm (ACA), the utilization efficiency of train-sets can be increased from 43.4% (ACA) to 46.9% (Two-Stage), and 1 train-set can be saved up to fulfill the same transportation tasks. The approach proposed in the study is faster and more stable than the traditional ones, by using which, the HSR staff can draw up the train-sets circulation plan more quickly and the utilization efficiency of the HSR system is also improved.

Introduction

During the past few years, China has been in a period when high-speed railway (HSR) developed rapidly, with the operation mileage increasing from 404 kilometers in 2008 to over 19 thousand kilometers at the end of 2015. With the gradually formed HSR network, the mobility of society, including people and cargo, is increasingly dependent on the HSR. Even though the imbalance between transport supply and demand has been slightly mitigated by introducing HSR, great efforts are still needed to improve the transport capability of HSR, so as to satisfy

the explosively increasing mobility demand. Generally, the transport capability of HSR can be improved from two aspects. Firstly, speeding up the infrastructure construction can raise the upper bound of capacity. Secondly, improving the operation and management of HSR is an effective way to make the most of the existing infrastructure. Considering the long construction period, it is impractical to increase the HSR operation mileage within a short time. Therefore, enhancing the utilization efficiency of HSR is an achievable and key solution to improve the transport capacity and mitigate the social contradiction of imbalance between transport supply and demand.

As the carrying tools of HSR, train-sets play a vital role in the process of completing society mobility task, the utilization efficiency of which is one of the most important factors influencing the transport capacity of HSR. Therefore, how to improve the train-set utilization efficiency has been a research hotspot in railway transport for several decades. The utilization efficiency of train-sets is directly dependent on the train-set utilization plan, which is made in a given train graph by a number of train-sets to complete the trip tasks. When a train-set utilization plan is drawn up, the amount of train-sets in demand as well as the working time and idle time of each train-set can be obtained. Obviously, by following a good train-set utilization plan, the proportion of train-sets' working time in each day can be dramatically increased. Thus, a number of train-sets can be saved to serve for other train schedules, which means more trips can be added to the train graph without considering the limitation of available train-sets.

During the past few decades, extensive research has been carried out worldwide to make an attempt to draw up a high-quality train-set utilization plan, leading to the development of various models and techniques. In Europe, the train-set utilization optimization problem mainly focused on the coupling and uncoupling arrangements of train-sets to carry out the trip tasks in the train graph. Schrijver [1] firstly paid attention to the train-set scheduling problem. He proposed a model based on the minimum cost flow theory and solved it by the software CPLEX. Abbink et al. [2] explored the train marshalling problem during the peak period, and the optimization goal was set to minimize the shortage number of train seats. Alfieri et al. [3] addressed the train-set utilization problem on a single train line in a single day. An integer programming model was described to obtain the circulation of rolling stock considering the order of the train units in the compositions. Fioole et al. [4] put forward a mixed integer programming model on basis of previous research achievements, and the improved branch and bound algorithm was applied to obtain the optimization solution. Barnhart et al. [5], Peeters and Kroon [6], and Holmberg & Yuan [7] used the Dantzig-Wolfe decomposition approach [8] to solve the train-set circulation problem, instead of the classical flow decomposition method used in other network flow problems. Maróti and Kroon [9] proposed an integer programming model with the consideration of maintenance constraint. Moreover, he designed a heuristic algorithm to solve the model and obtained a feasible solution. However, the operational mode of train-sets in China is quite different from that in European countries. In China, the train-sets are used as a whole, so there is no need to consider the problem of coupling and uncoupling. Zhao et al. [10] discussed the different utilization modes of train-sets comprehensively and they found that the mode of unrestrained section can improve the utilization efficiency. Nie et al. [11] proposed an integer programming model considering the deadhead trains condition, and applied the Hungary algorithm to solve the problem. Wang et al. [12] set up a train-set utilization model to complete the trip tasks in a cyclic train timetable, and a column generation algorithm was used to obtain the optimal solution. Then Wang et al. [13] improved this model, and designed a method of branch and price, and the findings have inspired latter research. Zhao [14] believed that the train-set utilization problem can be regarded as a typical NP hard problem, and it cannot be solved directly by ready-made

software. Characterized by easy-to-understand and fast, a large number of heuristic algorithms have been applied to solve the train-set utilization problem. Yang et al. [15] built a train-set connection network model aiming to maximize the train-set utilization efficiency, and designed a genetic algorithm to solve the model. Chen et al. [16] proposed a multi-objective optimization model to maximize the equilibrium degree of train-set utilization. Considering the number of train-sets in use, a simulated annealing algorithm was designed to solve the model. Hong et al. [17] put forward an integer programming model considering the maintenance and train graph constraints, and a heuristic approach was applied to solve this problem.

Recent research on the train-set utilization problem mainly focuses on developing new methodologies and introducing rescheduling procedures. The train-set utilization problem has been proved to be a TSP (Traveling Salesman Problem) with some special constraints. Based on TSP, Zhang [18] built a train-set utilization model in a railway network by taking both the running time constraint and running distances constraint into consideration, and an ant colony algorithm was used to solve the model; Zhou [19] proposed an integer programming model to obtain the train-set utilization plan, considering the constraint of Level 1 maintenance standard, and an improved ant colony algorithm was designed to solve the model. Thorlacius [20] proposed an integrated rolling stock planning model that took into account all practical requirements simultaneously for rolling stock planning at DSB S-tog. By using a hill climbing heuristic algorithms to solve this model, the efficiency of new rolling stock plan was increased by 2%. Jørgen [21] designed two approaches, one of which is Mixed Integer Linear Program (MILP) solved by CPLEX and the other is column and row generation approach, and both approaches are sufficiently fast in real-time setting. Lai [22] put forward a hybrid heuristic process to solve this problem for Taiwan Railway Administration (TRA). The efficiency of rolling stock use was increased by about 5% and the solution time was reduced significantly from 3h to 11.2s. Tatsushi [23] proposed a column generation and Lagrangian relaxation heuristics, and good lower and upper bounds for 300 trips were found within reasonable computing time. Due to disturbances or disruptions from the real world, the train-set circulation plan need to be rescheduled in order to maintain a feasible rolling stock circulation, and generally, this procedure should be carried out in a short time. Some researchers studied the train-set utilization problem in the rescheduling procedures. Cadorso [24] proposed a new approach to obtain better and more robust circulations of the rolling stock train units, solving the rolling stock assignment while accounting for the train routing problem. By designing a heuristic based on the benders decomposition, more robust and efficient solutions were obtained. Wagenaar [25] proposed three different models for rescheduling, namely the Extra Unit Type model, the Shadow-Account model and the Job-Composition model. All models were tested on instances of Netherlands Railways. Results showed that the models were able to take maintenance appointments into account efficiently. Cacchiani [26] presented an overview of recovery models and algorithms for real-time railway disturbance and disruption management, and the application of these models and algorithms in real-life railway systems would be instrumental in improving the quality of provided railway services, leading to an increased utilization of the involved railway systems. Table 1 provides a systematic comparison of key model components and solution methods in the existing research.

Although there is a comprehensive body of literature on the train-set utilization problem, no results of previous research can be directly used into the practical production. There are two major reasons. First, limited studies take the maintenance constraints into consideration, which is actually a key factor influencing the compilation of the train-set utilization plan. Second, even if the maintenance is considered, the solutions of models are obtained by applying heuristic algorithms, which is unstable and inaccurate. Therefore, to realize the objective that the train-set utilization plan can be generated by computers automatically and used in practice,

Table 1. Comparison of key model components and solution methods in the existing train-set utilization research.

Modeling scenarios	Modeling features	Objective Function	Solution Method	Publication
MT, RD, NM, SB	IP	Minimize train-set costs	Solve by CPLEX	Scrijver [1]
ST, UUR, NM, SB	IP, AP	Minimize connection time	Hungarian Algorithm	Zhao et al. [14]
MT, NM	IP	Minimize absence of passenger	Solve by CPLEX	Abbink et al. [2]
ST, M, UFR	IP, MFP, TM	Minimize transition costs	Solve by CPLEX	Maróti and Kroon [27]
RD, NM, UFR	IP, MFP	Minimize train-sets and running distance	Preprocess and solve by CPLEX	Alfieri et al. [3]
MT, RD, NM	MIP	Minimize absence of passenger and shunting	Improved Branch and Bound, CPLEX	Fioole et al. [4]
RD, NM, UFR	IP	Minimize absence of passenger and shunting	D-W Decomposition Algorithm	Peeters and Kroon. [6]
RD, NM	IP	Minimize absence of passenger and shunting	Heuristic Algorithm and CPLEX	Maróti [28]
MT, M, UUR	NM, TSP	Minimize connection time	Computer Simulation	Geng et al. [29]
ST, M	IP	Minimize train-sets	Crossover heuristic Algorithm	Hong et al. [17]
M, UUR, MB	MOIP, TSP	Minimize train-sets and maintenance	Hierarchical Heuristic Algorithm	Miao et al [30]
ST, M, UUR	IP, TSP	Minimize connecting time	Improved Ant Colony Algorithm	Zhou et al. [19]
ST, M, UUR, MB	IP, TSP,	Minimize connecting time	Model Convert, Two Stage Method	Present Study

Modeling scenarios: ST-Single Train-Set Type; MT-Multiple Train-set Type; SB-Sing Base; MB-Multiple Base; RD-Reconnection and Decomposition; NM-No Maintenance; M-Maintenance; UUR-Using in Uncertain Region; UFR-Using in Fixed Region

Modeling characteristics: IP-Integer Programming; AP-Assignment Problem; SM-Simulation Model; MFP- Multi-Commodity Flow Problem; MIP- Mixed Integer Programming; TM-Transition Model; NM-Network Mode; TSP-Traveling Salesman Problem; MOIP-Multi-Objective Integer Programming

<https://doi.org/10.1371/journal.pone.0175698.t001>

an integer programming model considering the maintenance constraints is built to maximize the train-set utilization efficiency. Specially, a two-stage method is proposed in the model solving process in this study to obtain a unique solution. With this method, the train-set utilization plan can be obtained by a computer-generated system.

The remainder of this paper is organized as follows. Section 2 describes the basic situation of the train-set utilization problem. In Section 3, the optimization model of the train-set utilization plan is developed. Section 4 introduces the solution method of the model. In Section 5, a case study based on the Beijing-Tianjin HSR line is performed to illustrate the model application. Finally, Section 6 provides the conclusions marks.

Problem description

The train-set utilization plan is drawn up to identify the work arrangements of train-sets according to the given train graph as well as the rules and regulations of maintenance. For train-sets, except for some necessary operations in the railway stations, the main work includes two parts: one is to complete all the trip tasks in the given train graph; and the other is to carry out maintenances. Therefore, the train-set utilization plan should figure out the following aspects: the number of needed train-sets; the specific train-set undertaking its corresponding trip; the sequence of the trips undertaken by the same train-set; and the time and location each train-set is maintained. Different train-set utilization plans can be made to complete the trip tasks in a given train graph. Therefore, in this study, the optimization problem of the train-set utilization plan is to complete all the trip tasks in a given train graph by utilizing the least train-sets on the premise of satisfying the requirements of the maintenance.

Fig 1 illustrates the framework of the train-set utilization optimization problem. Firstly, some basic information should be prepared as the input parameters, including: the train graph, the rules and regulations of maintenance and station working. Secondly, in order to model the train-set utilization plan, a train-set utilization network is represented by a directed graph. Thirdly, an integer programming model is established to minimize the total connecting

Fig 1. The framework of the train-set utilization optimization problem.

<https://doi.org/10.1371/journal.pone.0175698.g001>

Fig 2. The relationship between the train graph and the train-set utilization plan.

<https://doi.org/10.1371/journal.pone.0175698.g002>

time, considering the spatial constraints, the time constraints, the maintenance constraints, and the unicity constraints. Finally, a two-stage approach is designed to solve the problem, and an optimal train-set utilization plan is obtained as the output.

Fig 2 illustrates the relationship between the train graph and the train-set utilization plan. Fig 2(a) is a simple sample of a given train graph, including three stations and eight trips. After adding the train-sets' connection relationships and maintenance information, Fig 2(b) represents the train-set utilization plan. In this sample, to complete the eight transportation tasks, it needs at least three train-sets. Particularly, Trip G82 and G89 are undertaken by one train-set; Trip G84, G87, and G88 are undertaken by the second train-set; and Trip G83, G86, and G91 are undertaken by the last train-set. The train-set undertaking Trip G83 should be maintained at the inspection and repair depot near Station B after completing the task of Trip G86.

Train-set utilization plan modeling

Train-set utilization network representation

In order to model the train-set utilization plan, a train-set utilization network is represented by a directed graph $G(V, E)$. The node set V is a union of subsets V_T and V_M , which represent trips and maintenances, respectively. Each node $i \in V_T$ is defined as a tuple $(n_i, s_i^d, s_i^a, t_i^d, t_i^a, d_i, t_i)$, where n_i indicates the train number, s_i^d indicates the departure station, s_i^a indicates the arrival station, t_i^d indicates the departure time, t_i^a indicates the arrival time, d_i indicates the running distance, and t_i indicates the running time. Each node $k \in V_M$ is defined as a tuple (φ_k, τ_k^a) , where φ_k indicates the inspection and repair depot, τ_k^a indicates the time train-set departing from the inspection and repair depot to arrival station after maintenance. As shown in Fig 3, there are three kinds of arcs in the train-set utilization network, including connection arcs (i.e., trip $\rightarrow$ trip), maintenance arcs (i.e., trip $\rightarrow$ maintenance) and arcs after maintenance (i.e., maintenance $\rightarrow$ trip). Therefore, the arc set E consists of three disjoint sets E_C , E_Q and E_H , where E_C represents directional arcs between two trip nodes, i.e., connection arcs, E_Q represents directional arcs from trip node to maintenance node, i.e., maintenance arcs and E_H represents directional arcs from maintenance node to trip node, i.e., arcs after maintenance.

Fig 3. Three kinds of arcs in the train-set utilization network.

<https://doi.org/10.1371/journal.pone.0175698.g003>

The weight of connection arcs ω_{ij}^C means the connection time between Trip i and Trip j . Generally, when the arrival station of Trip i is different from the departure station of Trip j , the connection time should be defined as infinite. It means that the two trips cannot be connected due to the huge waste of train-set empty running. As shown in Fig 4, when two trips can be connected, the values of ω_{ij}^C is determined by the sequential order of the two trips' departure time and arrival time. If the departure time of Trip j is after the arrival time of Trip i , the ω_{ij}^C is equal to $t_j^d - t_i^a$, otherwise, the ω_{ij}^C is equal to $1440 + t_j^d - t_i^a$, i.e., the train-set will undertake Trip j the next day. Based on above analyses, the values of ω_{ij}^C can be calculated as Eq (1).

$$\omega_{ij}^C = \begin{cases} t_j^d - t_i^a & t_j^d - t_i^a \geq 0, s_j^d = s_i^a \\ 1440 + t_j^d - t_i^a & t_j^d - t_i^a \leq 0, s_j^d = s_i^a \\ +\infty & s_j^d \neq s_i^a \end{cases} \quad (1)$$

The weight of maintenance arcs ω_{ik}^Q means the duration from the moment train-set arriving at its arrival station, t_i^a , to the moment train-set returning back to its arrival station after maintenance, τ_k^a . As shown in Fig 5, when $\tau_k^a - t_i^a \geq 0$, the value of ω_{ik}^Q is equal to $\tau_k^a - t_i^a$. When $\tau_k^a - t_i^a \leq 0$, it means that the train-set can only undertake Trip j the next day, and the value of ω_{ik}^Q is equal to $1440 + \tau_k^a - t_i^a$. Therefore, the values of ω_{ik}^Q can be calculated as Eq (2).

$$\omega_{ik}^Q = \begin{cases} \tau_k^a - t_i^a & \tau_k^a - t_i^a \geq 0 \\ 1440 + \tau_k^a - t_i^a & \tau_k^a - t_i^a \leq 0 \end{cases} \quad (2)$$

The weight of arcs after maintenance ω_{ki}^H means the duration from the moment train-set returning back to its arrival station after maintenance, τ_k^a , to the moment train-set departing from the station, t_i^d . As shown in Fig 6, when $t_i^d - \tau_k^a \geq 0$, the value of ω_{ki}^H is equal to $t_i^d - \tau_k^a$.

Fig 4. The weight of connection arcs.

<https://doi.org/10.1371/journal.pone.0175698.g004>

When $t_i^d - \tau_k^a \leq 0$, it means that the train-set can only undertake Trip i the next day, and the value of ω_{ki}^H is equal to $1440 + t_i^d - \tau_k^a$. Therefore, the values of ω_{ki}^H can be calculated as Eq (3).

$$\omega_{ki}^H = \begin{cases} t_i^d - \tau_k^a & t_i^d - \tau_k^a \geq 0 \\ 1440 + t_i^d - \tau_k^a & t_i^d - \tau_k^a \leq 0 \end{cases} \quad (3)$$

Analyses of constraints

To undertake a trip task, the train-set should satisfy four specific requirements, including the spatial constraint, the time constraint, the maintenance constraint and the unicity constraint. The detail analyses are as follows:

1. Spatial constraints. As mentioned above, an empty running train-set needs to be dispatched to connect two trips. When the departure station of the latter trip is different from the arrival station of its former trip, it will cause huge waste. Therefore, as shown in Eq (4), when a train-set undertake two adjacent trips, the departure station of the latter trip, s_j^d , is restricted to

Fig 5. The weight of maintenance arcs.

<https://doi.org/10.1371/journal.pone.0175698.g005>

be the same with the arrival station of the former trip, s_i^a , where $i \leq j$.

$$s_i^a = s_j^a \quad i \leq j \quad (4)$$

2. The time constraints. The time constraints means that the duration of connection arcs, maintenance arcs and arcs after maintenance should respectively satisfy the requirement of the minimal necessary working procedure duration between two tasks, as shown in Fig 7.

When a train-set arrives at the station after completing the trip task, some preparation work need to be carried out before departing for the next trip, such as cleaning, pollution discharge, etc. The duration needed for such preparation work is defined as operation time, T_o . Therefore, to ensure that two trips can be undertaken by one train-set, the connection time ω_{ij}^C of the two trips should be longer than the operation time T_o , as shown in Eq (5).

$$\omega_{ij}^C \geq T_o \quad (5)$$

When a train-set completes a trip task and reaches to the maintenance standard, it should

Fig 6. The weight of arcs after maintenance.

<https://doi.org/10.1371/journal.pone.0175698.g006>

be maintained in the inspection and repair depot. As shown in Fig 8, four essential procedures will be followed by the train-set to complete the maintenance task. Table 2 lists the detailed information about the four procedures along with their duration definitions.

To ensure a trip node can be connected to a maintenance node, there should be enough time for the train-set to go through these procedures. Therefore, Eq (6) restrains that the weight of maintenance arc ω_{ik}^Q should be larger than the duration of the process T_z .

$$\omega_{ik}^Q \geq T_z \quad (6)$$

Once completing the maintenance task, train-sets can be utilized to undertake the trip tasks. However, as mentioned above, a series of preparation work should be carried out before departing to ensure the train-set working normally, as shown in Fig 9. Therefore, in order to connect a maintenance node with a trip node, Eq (7) restrains that the weight value of arc after maintenance ω_{ki}^H should be larger than the duration of preparation work T_c .

$$\omega_{ki}^H \geq T_c \quad (7)$$

Fig 7. The relationship between the operation time and the weight of connection arcs.

<https://doi.org/10.1371/journal.pone.0175698.g007>

Fig 8. The relationship between the process duration and the weight of maintenance arcs.

<https://doi.org/10.1371/journal.pone.0175698.g008>

Table 2. Four essential procedures.

	Procedures	Duration	Definition
1	Some basic inspection will be carried out after the train-set arriving at the station.	Inspection time	T_a
2	The train-set will departure from the arriving station to the inspection and repair depot.	Running time	T_b
3	The maintenance will be carried out in the inspection and repair depot	Maintenance time	T_m
4	The train-set will return back from the inspection and repair depot to the station.	Running time	T_b
	Total	Duration of the process	$T_z = T_a + 2T_b + T_m$

<https://doi.org/10.1371/journal.pone.0175698.t002>

3. The maintenance constraints. For safe operation concern, the train-sets must be strictly complied with the maintenance rules and standards. In China, the train-sets' maintenance standards are divided into five levels according to the accumulated operation time and mileages. In this study, it is assumed that the train-set utilization plan is drawn up by treating one day as a cycle. Generally, the operation time standard of Standard Level 2 to Standard Level 5 is more than a month. Therefore, only the Standard Level 1 needs to be considered in the study. The Standard Level 1 requires that when the accumulated operation time reaches up to 48 hours or the accumulated operation mileages reaches up to 4,000 km, the train-set must be maintained. Additionally, in practical applications, a 10% fluctuation compared with standard is allowable (e.g., Train-sets should be maintained when the accumulated operation mileages are within a range from 3600 km to 4400 km).

When undertaking several trips, the accumulated operation time and mileages of the train-set are $\sum_{i=1}^n T_i$ and $\sum_{i=1}^n d_i$ respectively, where n is the number of undertaking trips. If ΔS_m and

Fig 9. The relationship between the duration of preparation work and the weight of arcs maintenance.

<https://doi.org/10.1371/journal.pone.0175698.g009>

ΔS_t represent the acceptable fluctuation ranges of operation time and mileages respectively, the accumulated operation time and mileages of each train-set should be within the range of $[S_t - \Delta S_t, S_t + \Delta S_t]$ and $[S_m - \Delta S_m, S_m + \Delta S_m]$, respectively.

$$S_t - \Delta S_t \leq \sum_{i=1}^n T_i \leq S_t + \Delta S_t \quad (8)$$

$$S_m - \Delta S_m \leq \sum_{i=1}^n d_i \leq S_m + \Delta S_m \quad (9)$$

4. The uniqueness constraints. The uniqueness involves two aspects. One is that after completing a trip task, each train-set should either undertake another trip or go for maintenance. The other one is that for each trip in a given train graph, only one train-set can be allocated to complete the task.

The binary variable x_{ij}^C is defined to describe whether the trip node i and trip node j are connected by the connection arc. The binary variable x_{ik}^Q is defined to describe whether the trip node i and maintenance node k is connected by the maintenance arc. The binary variable x_{ki}^H is defined to describe whether the maintenance node k and trip node i are connected by the arc after maintenance. If the two nodes are connected by arcs, the value of the binary variable is 1, otherwise 0, as shown in Eqs (10)–(12).

$$x_{ij}^C = \begin{cases} 1 & \text{trip node } i \text{ and trip node } j \text{ is connected} \\ & \text{by a connection arc} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$x_{ik}^Q = \begin{cases} 1 & \text{trip node } i \text{ and maintenance node } k \text{ is connected} \\ & \text{by a maintenance arc} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$x_{ki}^H = \begin{cases} 1 & \text{maintenance node } k \text{ and trip node } i \text{ is connected} \\ & \text{by a arc after maintenance} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Eq (13) restrains when a train-set completing Trip i , it should either undertake Trip j or go for Maintenance k . Eq (14) restrains that Trip j should be undertaken by a train-set either just completing Trip i or Maintenance k .

$$\sum_{j=1}^a x_{ij}^C + \sum_{k=1}^b x_{ik}^Q = 1 \quad i \neq j \text{ \& } k \neq i \quad (13)$$

$$\sum_{i=0}^a x_{ij}^C + \sum_{k=0}^b x_{kj}^H = 1 \quad i \neq j \text{ \& } k \neq j \quad (14)$$

Optimization goal

Railway passenger transport corporations aim to realize profit maximization on the premise of ensuring passengers' life and property security. Therefore, the cost is a key factor influencing the operation managers to make decisions. Due to the high acquisition cost and maintenance cost of train-sets, it is an effective measure to improve the utilization efficiency of train-sets. Thus, to complete the same number of trip tasks in a given train graph, less train-sets are needed. In a train-set utilization plan, the required amount of train-sets is equal to the value that 1440 minutes, i.e., a cycle, divides the sum of the total connecting time and trips' running time. The running time of all trips is a fixed value given in the train graph, so the required amount of train-sets is only determined by the total connecting time. Therefore, the optimization goal of train-set utilization plan is to minimize the total connecting time, which includes the following three parts: the connecting times of all connection arcs ($\sum_{i,j \in [1, a], i \neq j} \omega_{ij}^C x_{ij}^C$), all maintenance arcs ($\sum_{i \in [1, a], k \in [1, b], i \neq k} \omega_{ik}^Q x_{ik}^Q$), and all arcs after maintenance ($\sum_{i \in [1, a], k \in [1, b], i \neq k} \omega_{ki}^H x_{ki}^H$), as shown in Eq (15).

$$\min \left(\sum_{i,j \in [1, a], i \neq j} \omega_{ij}^C x_{ij}^C + \sum_{i \in [1, a], k \in [1, b], i \neq k} \omega_{ik}^Q x_{ik}^Q + \sum_{i \in [1, a], k \in [1, b], i \neq k} \omega_{ki}^H x_{ki}^H \right) \quad (15)$$

Model solution: A two-stage approach

The train-set utilization planning model aims to identify whether arcs should be connected between trips and maintenances. Among the three kinds of arcs, the connection arcs between two trips is relatively easy to obtain due to the static characteristics that all trips are given in the train graph. Inversely, when and where the maintenances should be carried out are indeterminate according to the accumulated operation time and mileages. Therefore, the difficulty in solving the train-set utilization planning problem lies in the solving process of maintenance arcs and arcs after maintenance. To solve the above difficulty, a two stage approach is proposed. As shown in Fig 10, in the first stage, all the trips in the given train graph are allocated into a set of segments and no arcs are connected between trip nodes and maintenance nodes. In the second stage, the maintenance arcs and arcs behind maintenance will be connected between different segments. By designing a solving algorithm in each stage, the optimal and unique solution of the train-set utilization planning model can be obtained, and so the compilation of train-set utilization plan can be realized in a computer-generated system.

The first stage: Segment generation

With the introduction of the segment concept, a directed graph $\hat{G}(\hat{V}, \hat{E})$ should be defined to represent the train-set utilization network. The node set $\hat{V}$ is a union of ordered subsets $\hat{V}_{SE}$ and $\hat{V}_M$, which represent segments and maintenances, respectively. Each node $i \in \hat{V}_{SE}$ is defined as a tuple $(p_i, f_i^s, f_i^t, l_i^s, l_i^t, time_i, distance_i)$, where p_i indicates the number of trips in the Segment i , f_i^s and f_i^t indicate the departure station and departure time of the first trip respectively, l_i^s and l_i^t indicate the arrival station and arrival time of the last trip respectively, $time_i$ and $distance_i$ indicate the running time and distance of all the trips respectively.

Fig 10. Basic process of the two-stage approach.

<https://doi.org/10.1371/journal.pone.0175698.g010>

The first stage aims to obtain the connection relationship between trip nodes and prepare for the next stage simultaneously. Therefore, when allocating trips into segments, the following requirements should be met.

1. For each segment, to ensure no maintenance arcs and arcs after maintenance are included, the running time and distance of all trips should be within the maintenance standards.

2. For each segment, to ensure two adjacent trips can be connected arbitrarily, the departure station of the latter trip should be the same with the arrival station of the former trip. Moreover, the dwell time between the two trips should satisfy the requirement of operation time.

Let $ST = \{1, 2, \dots, m\}$ represent the set of stations, where m is the total number of stations in the train graph. To obtain the segment sets more efficiently, all trips are classified into a group of new trip sets based on the trips' departure station. Let $DE = \{DE_1, DE_2, \dots, DE_i, \dots, DE_m\}$ represent the above new trip sets. Particularly, all the trips in set $DE_i \in DE$ departure from station i , where $1 \leq i \leq m$.

Based on the above analyses, the segment generation algorithm is designed as follows.

Algorithm 1: Segment Generation

```

while  $DE \neq \emptyset$ 
  do generate segment  $\hat{v}$ , initialize  $time = 0$ ,  $distance = 0$ ;
  for  $i = 1$  to  $0$ 
    if  $DE_i \neq \emptyset$ 
      add the first trip  $t_0$  in the set  $DE_i$  into the segment  $\hat{v}$ ;
       $distance = distance + d_0$ ,  $time = time + (t_0^a - t_0^d)$ ;
      remove  $t_0$  from  $DE_i$ ; break;
    end
  end
  while  $distance < S_m + \Delta S_m$  and  $time < S_t + \Delta S_t$ 
    do find  $DE_j$  according to the arrival station  $s_0^a$  of trip  $t_0$  in set  $DE_i$ 
    if  $DE_j \neq \emptyset$ 
      sort the set  $DE_j$  in ascending order based on the time interval between the
      arrival
      time of  $t_0$  in set  $DE_i$  and the departure time of trips in the set  $DE_j$ 
      for  $k = 0$  to  $q_j$  ( $q_j$  denotes the number of trips in the set  $ST_j$ )
        if  $t_k^d - t_0^a > T_o$ ,
           $distance = distance + d_k$ ,  $time = time + (t_k^a - t_k^d)$ 
          if  $S_m - \Delta S_m \leq distance \leq S_m + \Delta S_m$  &  $S_t - \Delta S_t \leq time \leq S_t + \Delta S_t$ 
            add  $t_k$  into  $\hat{v}$ ,  $t_0 = t_k$ , remove  $t_k$  from  $DE_j$ , break;
          else:  $distance = distance - d_k$ ,  $time = time - (t_k^a - t_k^d)$ 
        end
      end
    else
      break;
    end
  end

```

The second stage: Segment connection

There are three decision variables in the train-set circulation planning model, including x_{ij}^C , x_{ik}^Q and x_{ki}^H . In the first stage, the decision variable x_{ij}^C has been solved since the connection relationships between trip nodes are obtained. Therefore, the objective of the second stage is to calculate the decision variables x_{ik}^Q and x_{ki}^H , i.e., the connection relationship between trips and

Fig 11. Flow chart of the segment connection.

<https://doi.org/10.1371/journal.pone.0175698.g011>

maintenance. With the concept of segment, the arcs between segments is equivalent to the combination of the maintenance arc and arc after maintenance. Therefore, the second stage aims to identify the segment connection relationship.

When node $i \in \hat{V}_{SE}$ is connected to node $j \in \hat{V}_{SE}$, it means that the train-set should be maintained after completing the last trip in Segment i , and then undertake the first trip in Segment j . Therefore, to ensure that node $i \in \hat{V}_{SE}$ can be connected to node $j \in \hat{V}_{SE}$, the following requirements should be satisfied:

1. The arrival station of the last trip, l_i^s in Segment i should be the same with the departure station of the first trip, f_j^s in Segment j .
2. The time interval, ω_{ij}^s , between the last trip in Segment i and the first trip in Segment j should be longer than the duration of a series of working procedures, which is equal to the sum of T_Z and T_C . Particularly, the value of ω_{ij}^s can be obtained by Eq (16).

$$\omega_{ij}^s = \begin{cases} f_j^t - l_i^t & f_j^t - l_i^t \geq T_Z + T_C \\ 1440 + f_j^t - l_i^t & f_j^t - l_i^t < T_Z + T_C \end{cases} \quad (16)$$

Algorithm 2: Segment Connection

- Step 1. Initial segment node set $\hat{V}_{SE} = \{\hat{v}_1, \hat{v}_2 \dots, \hat{v}_n\}$, arc set $E = \emptyset$ and station set $ST = \{1, 2 \dots, m\}$, let $st = 1$.
- Step 2. Classify the segment set $\hat{V}_{SE}$ according to the last trip arrival station, $l_i^s = st$, and the first trip departure station, $f_j^s = st$; get arrival segment set $\hat{V}_{st}^a = \{\hat{v}_i | \hat{v}_i = (p_i, f_i^s, f_i^t, l_i^s = st, l_i^t, time_i, distance_i)\}$ and departure segment set $\hat{V}_{st}^d = \{\hat{v}_j | \hat{v}_j = (p_j, f_j^s = st, f_j^t, l_j^s, l_j^t, time_j, distance_j)\}$, sort the two sets by time ascending.
- Step 3. Calculate ω_{ij}^S according to the above rule, get the matrix $\{\omega_{ij}^S\}$.
- Step 4. Gain a set of optimal solutions by the Hungarian Algorithm, get the matrix $\{\omega'_{ij}^S\}$, add the zero corresponding arc into set E .
- Step 5. If $st = m$, end the algorithm; else let $st = st + 1$ and go to step 1.

Case studies

In this section, the case of the train graph of Beijing-Tianjin passenger dedicated line was adopted to test the performance of the proposed train-set utilization model and solution approach. Several evaluation indexes were applied to evaluate the quality of the computation results. Furthermore, the comparison between the ant colony algorithm and the two-stage approach was conducted to evaluate the advancements. The proposed approach is coded in C# and solved by Microsoft Visual Studio, and the computational experiments were performed on a 3.4 GHz Core i7 PC with 16 GB of RAM. The proposed model can be solved within a few seconds.

Main features of the case study

Fig 12 illustrates the topological structure of the Beijing-Tianjin passenger dedicated line, the length of which is nearly 120 kilometers. There are five stations along the line, including two endpoint stations (Beijing South Station (BJSS) and Tianjin Station (TJS)), and three middle stations (Yizhuang Station (YZS), Yongle Station (YLS), and Wuqing Station (WQS)). Moreover, only the two endpoint stations BSS and TJS are located near the repair and inspection depot. Therefore, the maintenance of train-sets can be carried out only after arriving at BJSS and TJS.

Actually, the train graph used in this study was carried out in December, 2014. As shown in Fig 13, there are 174 trips in total in the train graph, among which 86 trips are from BJSS to TJS and 88 trips are from TJS to BJSS. The total running distances of all the trips, D_{alb} is equal to 20,880 kilometers, and the total running time, T_{alb} is equal to 6,090 minutes.

The parameters related to the model and approach are listed as follows:

- Basic operation time T_o : $T_o^{BJSS} = 14$ minutes; $T_o^{TJS} = 14$ minutes.
- Minimal duration of the four essential procedures if train-sets needs to be maintained T_Z : $T_Z^{BJSS} = 261$ minutes; $T_Z^{TJS} = 257$ minutes.

Fig 12. The topological structure of Beijing-Tianjin passenger dedicated line.

<https://doi.org/10.1371/journal.pone.0175698.g012>

- Duration of preparation work before train-set departing in station T_C : $T_C^{BJSS} = T_C^{TJS} = 28$ minutes.
- Distance range of maintenance standard S_{one} : $3,600 \text{ kilometers} \leq S_{one} \leq 4,400 \text{ kilometers}$.
- Time range of maintenance standard T_{one} : $1.8 \text{ day} \leq T_{one} \leq 2.2 \text{ day}$.

Fig 13. Train graph of Beijing-Tianjin passenger dedicated line.

<https://doi.org/10.1371/journal.pone.0175698.g013>

Computational results of the train-set utilization plan

Fig 14 illustrates the computational results of the model, i.e., the train-set utilization plan of the train graph of Beijing-Tianjin passenger dedicated line carried out in December, 2014.

Table 3 lists the results of some important evaluation indexes.

The results show that to complete the above 174 trip tasks, the minimal amount of needed train-set N is 12. The total maintenance times M of all the 12 train-sets is 6 each day. The

Fig 14. Train-set utilization plan in Beijing-Tianjin passenger dedicated line.

<https://doi.org/10.1371/journal.pone.0175698.g014>

Table 3. Computational results.

Train-sets Amount	Maintenance Times	Utilization Rate	Connecting Time
12	6	46.9%	31,055 minutes

<https://doi.org/10.1371/journal.pone.0175698.t003>

average running distances of each train-set per day $D_{average} = D_{all} / N$ is 1740 kilometers. Each day, the rail line overhaul is carried out during the period 0:00–6:00. Therefore, the train-sets cannot work for six hours per day, and the maximal running time of each train-set can be calculated as follows: $1440 - 6 \times 60 = 1080$ minutes. The average utilization efficiency of each train-set $U = T_{all} / (1080 \times N)$ reaches 46.9%.

Theoretically, the minimum connection time of this case is obtained by applying Hungarian algorithm, which is equal to 25,542 minutes when maintenance constraints are not considered. Moreover, the total maintenance times of all train-sets is 4.7 each day, which can be calculated as $D_{all} / (1 + 10\%)S_{one}$. However, the theoretical results cannot be reached in practical production, because train-sets must be maintained according to the rules and regulations. Also, not all train-sets need to be maintained until their running distances reaching the upper limit of the standard. Therefore, the connection time 31,055 minutes and maintenance times 6 obtained by the proposed model and approach are very close to the theoretical values. These results prove that the proposed model and approach are feasible and the train-sets are utilized with high efficiency.

Comparison with the ant colony algorithm

Many researchers have applied the ant colony algorithm (ACA) to solve train-set utilization problem. In order to highlight the performance of this study in improving the train-set utilization efficiency, the ant colony algorithm is applied to carry out the same experiments with the same computer, and the detailed comparison results are as listed in Table 4. The results show that obviously the two-stage approach has the advantages over the traditional ACA in improving the solution quality and shortening the computation time.

On one hand, the utilization efficiency of train-sets is much higher by applying the two-stage approach. First, the amount of train-sets in use is reduced, which will save the acquisition cost significantly. Second, the total maintenance times is decreased from eight to six, saving much invalid running time from the railway stations to the inspection and repair depot. Third, the average running distances of each train-set per day is 134 kilometers longer than that of the traditional ACA.

On the other hand, the computation time of the two-stage approach is 2302 milliseconds, which is almost one-third of the traditional ACA. What should be noted is that, as the convergence speed of the traditional ACA is indeterminate, an average of 20 computation times is used. Furthermore, the computation of the two-stage approach is much more stable. Only one

Table 4. Results of the comparison between the ant colony algorithm and the two-stage approach.

	Two-stage approach	Traditional ACA	Difference
N	12	13	1
M	6	8	2
$D_{average}$ (kilometers/day)	1740	1606	134
U	46.9%	43.4%	3.5%
Computational time (millisecond)	2302	8032	5730

<https://doi.org/10.1371/journal.pone.0175698.t004>

optimal train-set utilization plan can be obtained with two-stage approach while different results are achieved by using the traditional ACA.

There are certain scope and limitations in the study. Firstly, the proposed approach is applicable for a short-term train-set circulation plan formulation. In China, there are five levels of maintenances in total, and the maintenance of level one is considered only in the study. As the maintenance of Level 2 to Level 5 take more than one month usually, it is beyond the scope of consideration in the study. Secondly, only one type of train-sets is studied by using the proposed approach, so the approach is appropriate for the new-built HSR with single type train-set, instead of the multi-type train-sets situation. As China has developed standard train-sets and tried to use the standard train-set as the main carrier in the future, the present approach will show important referential meaning in the utilization of standard train-set in the future.

Conclusion

The train-set utilization plan is drawn up to identify the work arrangements of train-sets, which is profoundly affected by the given train graph as well as the rules and regulations of maintenance. To optimize the train-sets' utilization efficiency, in this study, an integer programming model is proposed considering the spatial constraints, the time constraints, the maintenance constraints, and the unicity constraints. In the process of model solution, a two-stage approach is designed and an optimal train-set utilization plan is obtained as the output. Using the model, the case of the real train graph of Beijing-Tianjin passenger dedicated line was carried out. The results show that the model and approach proposed in this study is practical and reasonable. By comparing the results obtained by the two-stage approach and the ant colony algorithm, it could be found that to complete the same trip tasks, the former one needs less train-sets, less maintenance times, and longer average running distances for each train-set per day. In short, the utilization efficiency of train-sets is much higher by applying the two-stage approach. Moreover, the fast computation speed and good stability can contribute to the realization of a train-set utilization plan computer-generated system.

Acknowledgments

This work is financially supported by “the Fundamental Research Funds for the Central Universities” (2015YJS094).

Author Contributions

Conceptualization: YZ LSZ.

Data curation: YZ LSZ.

Formal analysis: YZ YW.

Funding acquisition: YZ LSZ.

Investigation: XML ZY.

Methodology: YZ LSZ YW.

Project administration: YZ LSZ.

Resources: YZ YW ZY.

Software: YZ LSZ.

Supervision: LSZ YW.

Validation: YZ LSZ.

Visualization: YW.

Writing – original draft: YZ LSZ.

Writing – review & editing: YZ YW ZY XML.

References

- Schrijver A. Minimum circulation of railway stock. *CWI Quarterly* 1993; 205–221.
- Abbink E, Berg BVD, Kroon L, Salomon M. Allocation of railway rolling stock for passenger trains. *Transportation Science* 2004; 38(1): 33–41.
- Alfieri A, Schrijver A. Efficient circulation of railway rolling stock. *Transportation Science* 2006; 40(3): 378–391.
- Fioole P, Kroon L, Schrijver A. A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research* 2006; 172(2): 1281–1297.
- Barnhart C, Hane CA, Vance PH. Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems. *Operations Research* 2000; 48(2): 318–326.
- Peeters M, Kroon L. Circulation of railway rolling stock: a branch-and-price approach. *Computers and Operational Research* 2006; 35(2): 538–556.
- Holmberg K, Yuan D. A multi-commodity network flow problem with side constraints on paths solved by column generation. *Inform Journal on Computing* 2003; 15(1): 42–57.
- Dantzig GB, Wolfe P. Decomposition principle for linear programs. *Operations Research* 1960; 8: 101–111.
- Maróti G, Kroon L. Maintenance routing for train units: the interchange model. *Computers and Operations Research* 2007; 34(4): 1121–1140.
- Zhao P, Yang H, Hu AZ. Research on usage of High-Speed passenger trains on uncertain railroad region. *Journal of the China Railway Society* 1997; 19(2): 15–19.
- Nie L, Zhao P, Yang H, An-Zhou HU. Study on motor trainset operation in high speed railway. *Journal Of The China Railway Society* 2001; 3(24): 1–7.
- Wang Y, Liu J, Miao J. Column generation algorithm based optimization method for maintenance scheduling of multiple units. *China Railway Science* 2009; 31(2): 115–120.
- Wang Y, Liu J, Miao J. Optimization of the circulation plan for multiple units based on adjustable train path. *China Railway Science* 2013; 33(4): 112–119.
- Zhao P, Tomii N. Train-set scheduling and an algorithm. *Journal of the China Railway Society* 2003; 1–7.
- Yang J, Yang H, Lu HB. Application of genetic algorithm in optimizing DMU circulating model. *Railway Transport and Economy* 2004; 26(7): 65–68.
- Chen H. Research on the relative problems of motor train-set scheduling system. Ph.D Thesis 2007; Beijing Jiao Tong University, Beijing.
- Hong SP, Kim KM, Lee K, Park BH. A pragmatic algorithm for the train-set routing: the case of korea high-speed railway. *Omega* 2009; 37(3): 637–645.
- Zhang C. Study on the train units operation problems of passenger dedicated rail network. Ph.D Thesis 2010; JiaoTong University, Beijing.
- Zhou Y, Zhou LS, Wang Y. Using improved ant colony algorithm to investigate EMU circulation scheduling problem. *Discrete Dynamics in Nature and Society* 2014; 1–13.
- Thorlacius P., Larsen J., Laumanns M., An integrated rolling stock planning model for the Copenhagen suburban passenger railway. *Journal of Rail Transport Planning and Management*, 2015, 5:240–262
- Jørgen T., Joris C., Lucas P., and Leo G., A comparison of two exact methods for passenger railway rolling stock (re)scheduling. *Transportation Research Part E*, 2016, 91:15–32.
- Lai Y., Fan D., and Huang K., Optimizing rolling stock assignment and maintenance plan for passenger railway operations. *Computer & Industrial Engineering*, 2015, 85:284–295.
- Tatsushi N., Akiyoshi O., Masahiro I, Satoru T., and Kenji U., A Combined column generation and heuristics for railway short-term rolling stock planning with regular inspection constraints. *Computers and Operations Research*, 2017, 81: 14–25.

24. Cadarso L., Marín A., Improving robustness of rolling stock circulations in rapid transit networks, *Computers and Operations Research*, 2014, 51:146–159.
25. Wagenaar J., and Kroon L., Maintenance in railway rolling stock rescheduling for passenger railways. *Erim Report*, 2015, Rotterdam.
26. Cacchiani V., Huisman D., Kidd M., Kroon G., Toth P., Veelenturf L., and Wagenaar J., An overview of recovery models and algorithms for real-time railway rescheduling. *Transportation Research Part B*, 2014, 63: 15–37.
27. Maróti G., Kroon L. Maintenance routing for train units: the transition model. *Transportation Science* 2005; 39(4): 518–525.
28. Maróti G. Research models for railway rolling stock planning, Ph.D Thesis 2006; Technische Universiteit Eindhoven.
29. Geng JC, Xiao RG, Quan NS, Niu HX. Research on periodicity of motor train-set scheduling for special lines for passenger traffic. *Journal of the China Railway Society* 2006; 28(4): 17–21.
30. Miao J, Wang Y, Yang Z. Research on the optimization of EMU circulation based on optimized connecting network 2010; 32(2): 1–7.