Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 7637298, 15 pages
https://doi.org/10.1155/2022/7637298

Research Article

Path-Based Approach for Expanding Rail Transit Network in a Metropolitan Area

Anjun Li[1,2] **Dian Wang**[1,2] **Qiyuan Peng**[1,2] **and Lisha Wang**[3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China*
[3]*School of Architecture and Urban Planning, Chongqing University, Chongqing 400030, China*

Correspondence should be addressed to Lisha Wang; wang_lisha@outlook.com

Received 9 May 2022; Accepted 25 July 2022; Published 5 September 2022

Academic Editor: Yuan Gao

Rail transit network design is an important strategic problem in determining the layout of infrastructure and improving operating performance. A core transit network with multiclass rail transit systems has been constructed in many metropolitan areas worldwide. In this study, we aimed to expand an existing network to shorten travel time and improve service quality under the restriction of limited transport supply. We formulate the studied problem as a mixed-integer linear model to obtain optimal construction links, the number of trains required on each link, and the path selected by each traveler such that the weighted sum of total costs from the perspective of travelers, operators, and investors is minimized. The formulated model is path-based, where feasible paths for each traveler are generated to describe the full door-to-door journey, including the first/last mile, transfers, and multiclass transit modes. Owing to the complexity of the network design problem and because it is impractical to enumerate all feasible paths for each traveler in real-size problems, we propose a column generation-based algorithm to find both tight lower bounds and good-quality solutions efficiently by considering only a subset of feasible paths. We prove that the pricing subproblem in column generation can be decomposed into multiple shortest path problems, which can be solved efficiently and separately, based on *O*/*D* pairs instead of individual travelers. A rail transit network along a metropolitan corridor was studied as an example. Multiple computational experiments were conducted, and the results illustrate the validity and practicality of the proposed methodology for solving the problem.

1. Introduction

To alleviate travel issues in urban areas and mitigate climate change, many metropolitan areas worldwide have constructed or upgraded rail transit systems, such as underground metros, overground commuter railways, and elevated monorails [1]. Increasing mobility and longer trips achieved by rail transit have also promoted the growth of metropolitan areas.

Because it is a huge investment to construct a rail line and incurs a large cost to operate rail transit systems, network design is an important process in the strategic stage. The problem consists of selecting nodes and links from a potential or underlying network to construct stations and the connections between them [2]. Network design issues also pervade the full hierarchy of strategic, tactical, and operational decision-making processes [3]. According to Farahani et al. [4], strategic decisions are related to the infrastructure of the network; tactical decisions are concerned with the effective utilization of infrastructure and resources in the network; and operational decisions are related to traffic flow and demand management. Citizens are encouraged to use mass rapid transit systems to relieve traffic congestion and reduce emissions in metropolitan areas. However, the convenience and level of service of rail transit affect people's preference to use it. Therefore, network design planners need to consider not only the location of facilities (represented by nodes and links) to control the investment but also the transport supply and service on the physical network to satisfy travel demand as much as possible.

The urban transportation network design problem is complicated and practical enough to warrant ongoing research [4, 5]. Generally, network design problems can be classified into two groups: networks planned from scratch and additions or extensions of lines in an already functioning network [2]. We focus on the latter in this study because an existing rail transit network usually needs to be expanded with time-varying travel demand, particularly along overloaded corridors. In this study, we propose a rail transit network design model, especially for making decisions regarding additions or extensions of links in an existing network with sufficient travel demand. Inspired by Li et al. [6], this study has two motivations. First, from the perspective of modeling methods, we would like to fully describe and efficiently find door-to-door travel paths for travelers in a multiclass rail transit network or even a multimodal transportation network. Both the links to be constructed and the paths to be selected are the decision variables. Second, in practice, we would like to improve the level of service of the current rail transit network by adding new infrastructure smartly with the proposed method. Shortening the travel time is an important objective.

The remainder of this paper is organized as follows. First, we review previous studies on rapid transit network design models in Section 2. The problem description is presented in Section 3. Section 4 presents a mixed-integer linear programming model for the studied rail transit network design problem. Section 5 describes the development of the column generation-based algorithm. Section 6 presents a case study and discusses the results. Finally, conclusions and future research directions are presented in Section 7.

2. Literature Review

Bussieck et al. [7] stated that railroad network planning decisions are mainly based on political reasoning and reviewed several mathematical programming methods for designing a railroad network. Laporte et al. [8] proposed a basic rapid transit network design (RTND) problem that focuses on designing a core network to maximize trip coverage. As a strategic stage of long-term planning, several players, such as politicians, urban planners, engineers, management consultants, and citizen groups, should be involved in the decision process [1]. Marín and García-Ródenas [9] included budget constraints in the network design model. Bagloee and Ceder [10] considered more details regarding transit networks, such as multiclass transit vehicles and system capacity. García et al. [11] and Cadarso and Marín [12] considered transfers at stations. From the perspective of system performance, Laporte et al. [13, 14] focused on the effectiveness and robustness of alternative routes at the planning stage. To describe the unserved demand by the transit network, An and Lo [15] considered flexible services, such as dial-a-ride or taxi. Cadarso and Marín [12] emphasized the importance of future impacts on mobility and congestion. Gutiérrez-Jarpa et al. [16] further proposed a multiobjective model that is conducive to a postoptimization analysis for effectiveness, efficiency, and equity concerns.

When planning a rapid transit network, the concept of "mobility as a service" should be considered from the viewpoint of mobility and accessibility [17]. Bagloee and Ceder [10] integrated the network design problem and frequency-setting analysis. Canca et al. [2, 18] included the network design, line planning, and a number of carriages required for each line in the optimization model. Peng et al. [19] proposed a non-linear and non-convex model to determine rail transit projects over multiple time periods and jointly optimize the headways. It was desirable to analyze the transit network development in a multimodal transportation system. Huang et al. [20] investigated a multimodal transit network design method and solved the optimization model with a heuristic algorithm, where passengers' route-choice behavior is described in the lower-level problem. Therefore, travel paths have received increasing attention in recent years. A door-to-door trip should include the first and last mile travels to rapid transit journeys for users' full travel paths. Laporte and Pascoal [21] and Gutiérrez-Jarpa et al. [16, 22] determined the shortest paths from the origin node to the destination node for each O/D pair. Gutiérrez-Jarpa et al. [16] considered the detailed passenger access or egress time between nodes and stations in the model. Chai et al. [23] studied the urban rail transit network design problem using a neighborhood search algorithm and applied a utility function for passenger path selection. Zhou et al. [24] solved the line-planning problem with passenger path assignment by using commercial linear programming solvers. However, all passengers in the same O/D pair choose the same path. More efficient algorithms need to be developed to deal with a large number of nodes or links in a multimodal transit network.

Table 1 summarizes and compares the key modeling components in existing research on the RTND problem. As shown in Table 1, most existing network design studies are modeled by links and address travel demand as an aggregated demand. However, it is difficult to distinguish the path selection among travelers with the same O/D pair. Moreover, the rapid transit network design problem has proven to be NP-hard [13, 14]. Heuristic algorithms are mostly used to solve non-linear models, which are always without bounding capacity [26]. Although heuristic methodologies can solve real-size problems, the quality of the resulting solutions cannot always be guaranteed without any lower-bounding technique.

In this study, we conclude that the contributions are threefold. First, we develop a path-based model that can describe the entire journey of each traveler, including many detailed factors, such as transfers and multiclass transit modes. This helps to assign travel paths for individual travelers. Even travelers with the same O/D pair can select different paths. Second, network design and capacity are interrelated [16]. In our model, vehicle supply is considered and matched with travel demand. Both the number of trains on each link during a certain period and the number of passengers on each train are constrained. Third, to handle the large network design problem, we designed a column generation-based algorithm to efficiently compute both tight lower bounds and high-quality solutions.

Table 1: Summary table of RTND models.

Publication	Formulation	Route/mode choice	Flow type	Capacity	Solution method
Laporte et al. [8, 14]	Linear, link-based	Aggregated	Demand	—	—
Marín and García-Ródenas [9]	Non-linear, link-based	Aggregated	Demand	—	Approximate of non-linear function
Bagloee and Ceder [10]	Route-based	Aggregated	Demand, supply	Constrained	Heuristic methodology
Laporte and Pascoal [21]	Linear, path-based	—	—	—	Modular heuristic
Cadarso and Marín [12]	Non-linear, link-based	Aggregated	Demand	—	Linear relaxation
Gutiérrez-Jarpa et al. [22]	Linear, corridor-based	Aggregated	Demand	—	An exact methodology within a heuristic
Gutiérrez-Jarpa et al. [16]	Linear, path-based	Aggregated	Demand	—	ϵ-constraint method
Canca et al. [18]	Non-linear, link-based	Aggregated	Demand, supply	Constrained	Branch-and-bound
Canca et al. [2, 25]	Non-linear, link-based	Aggregated	Demand, supply	Constrained	Neighborhood search algorithm
Huang et al. [20]	Non-linear, link-based	Aggregated	Demand, supply	Constrained	Artificial bee colony algorithm
Chai et al. [23]	Non-linear, link-based	Aggregated	Demand	Constrained	Neighborhood search algorithm
Peng et al. [19]	Non-linear, link-based	Aggregated	Demand, supply	Constrained	Genetic algorithm

3. Problem Description

3.1. Macroscopic Perspective. The core networks of rail transit have been constructed in many metropolitan areas around the world. The existing functioning network consists of multiclass rail transit systems. The planned links can form a fully connected network between the important nodes. However, in a complex built environment, feasible candidate links are limited. Therefore, several candidate links can be provided externally as inputs. In Figure 1, the solid lines represent rail transit links already under operation. The purple dashed lines represent the candidate links to be built.

With externally given candidate links to be designed, the rail transit network design problem in our study determines the optimal new links to be constructed in an existing network, the number of trains scheduled on each link, and the path of each traveler so that the given travel demand can be fulfilled. The main inputs include: (1) the network of available and proposed infrastructures represented by nodes i, j and link (i, j); (2) the estimated or observed travel demand represented by the number of travelers for each O/D pair r; (3) construction cost $f(i, j)$ of candidate link (i, j); (4) travel time on each link (i, j), which will be converted to travel cost $c_p(k)$ for traveler p and operating cost $c_v(i, j)$ in the model; and (5) capacity of passengers on each train $\mathrm{Cap}_{\mathrm{pax}}(i, j)$ and capacity of trains on each link $\mathrm{Cap}_{\mathrm{veh}}(i, j)$. Moreover, we only consider links (which can be regarded as segments of lines) connecting important stations. In other words, there may be some non-significant intermediate stations on the links. To consider the possible transport supply or demand request between those intermediate stations, a parameter of baseline resource $br_{i,j}$ is introduced for each link (i, j).

The objective is to implement one or more rail transit links from the candidate links to minimize the weighted sum of the total costs with respect to travelers, operators, and investors. Each traveler is assigned a path to achieve the system's optimum condition. Because the feasible seats and space of each train are limited, the number of onboard passengers on each link $\mathrm{Cap}_{\mathrm{pax}}(i, j)$ is constrained. Meanwhile, owing to the technical conditions of railways, the number of trains on each link during a certain time period (e.g., hour or day) $\mathrm{Cap}_{\mathrm{veh}}(i, j)$ is also strictly constrained.

3.2. Network Formation and Path Generation. To describe a complete journey for travelers, we extend the physical network with different types of nodes and links, where a station is no longer represented by a single node. In other words, each traveler starts and ends the journey at the demand nodes. Each travel path (from origin to destination) must pass through the demand nodes, station entrances, and platforms. Figure 2 illustrates a graph including one transfer station and two demand nodes. Travelers from node 1 to node 10 need to access station A and board a train; then transfer to station B and board another train; and finally, deboard the train at station C and finish the last mile to the destination. Specifically, the node sequence of the path from node 1 to node 10 is 1-2-3-4-5-6-7-8-9-10. If a new link is proposed to connect nodes 3 and 8 directly, the node sequence of another feasible path can be 1-2-3-8-9-10. For each O/D pair, a virtual link is created to describe travel for unserved demand or an alternative travel mode such as driving. An initial feasible path pool can be generated as a k-shortest path problem with flow-balance constraints.

4. Methodology

In this section, we propose a mixed-integer linear programming model for expanding rail transit networks to improve transport capacity and service quality. The goal is to

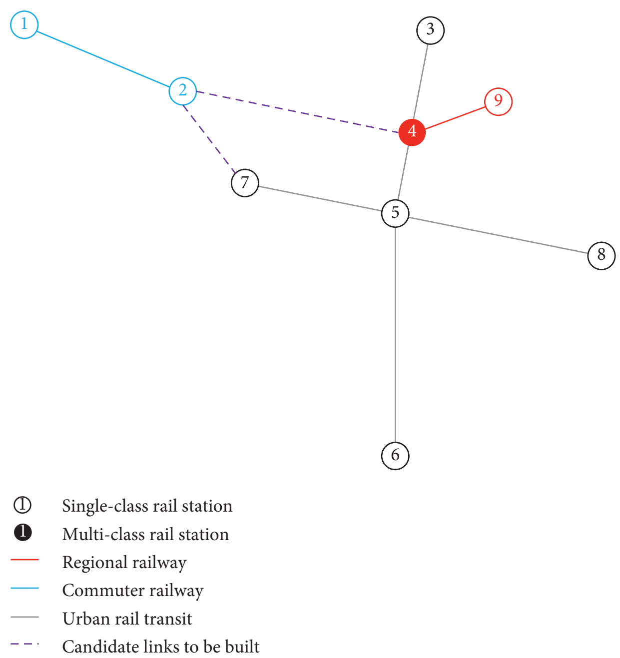

FIGURE 1: Illustration graph of a rail transit network in a metropolitan area.

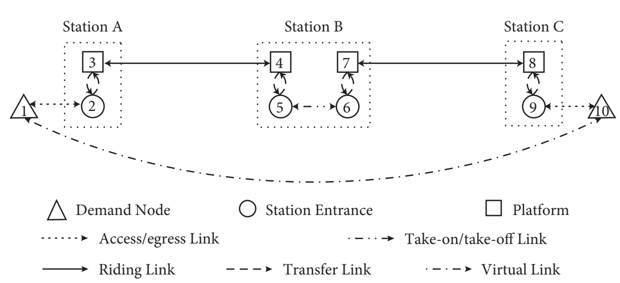

FIGURE 2: Illustration graph of a sample physical network.

enhance the existing rail transit network by adding new links that minimize the total generalized costs. The optimal links to be designed are determined, and each traveler is assigned a feasible travel path. To facilitate the model formulation, we assume the following simplifications: (1) the OD demand is given and fixed. Travelers choose the path with minimum generalized cost. This is a common assumption in previous studies [2, 12, 24, 27, 28]. (2) The set of candidate links for expanding the rail transit network is prespecified [19, 27]. The construction costs of candidate links are given and fixed. We need to determine the link(s) that should be constructed. (3) To ensure transport safety and a high level of service simultaneously, operators or authorities always impose the minimum and maximum headways of rail transit lines in a metropolitan area [29, 30]. The average headway is widely used to compute the waiting time in a high-frequency transit network [31–36]. Therefore, we set the waiting time at a platform node as constant and added it to the travel time of the outgoing link(s).

4.1. Notations. To facilitate model formulation, the main sets, parameters, and decision variables used in the model are formally defined in Table 2.

4.2. Objective Function. The rail transit network design model formulation is defined with respect to travelers, operators, and investors. Hence, we deal with a multi-objective optimization model. As usual, we minimized a positive linear combination of the different costs.

$$\min z = \alpha \cdot z_p + \beta \cdot z_v + \gamma \cdot z_c, \tag{1}$$

$$z_p = \sum_{p \in P} \sum_{k \in K} \left[\mu_p(k) \cdot c_p(k) \cdot x(k) \right], \tag{2}$$

$$z_v = \sum_{(i,j) \in L_R} \left[c_v(i,j) \cdot y(i,j) \right], \tag{3}$$

$$z_c = \sum_{(i,j) \in L_R} \left[f(i,j) \cdot z(i,j) \right]. \tag{4}$$

The minimization of the total travel costs for all travelers in the network, expressed by (2), is one of the main components of the objective function. The operating cost in equation (3) was also minimized from the perspective of the agencies. Finally, the total construction cost was minimized using equation (4). Because the units of the three types of costs are different, the values of coefficients α, β, and γ are relative. In other words, for different coefficients, a larger one may not indicate that the cost is more important than the other.

4.3. Constraints. As previously mentioned, there are five types of constraints: path selection, maximum number of onboard passengers on each link, minimum and maximum number of trains on each link, number of new links (optional), and variable domains.

$$\sum_{k \in K} \left[\mu_p(k) \cdot x(k) \right] = 1, \forall p \in P, \tag{5}$$

$$\sum_{p \in P} \sum_{k \in K} \left[\mu_p(k) \cdot \delta_{i,j}(k) \cdot x(k) \right] \leq \mathrm{Cap}_{\mathrm{pax}}(i,j) \cdot y(i,j) + br_{i,j}, \forall (i,j) \in L_R, \tag{6}$$

$$y(i,j) - \mathrm{Cap}_{\mathrm{veh}}(i,j) \cdot z(i,j) \leq 0, \forall (i,j) \in L_R, \tag{7}$$

$$y(i,j) \geq m \cdot \mathrm{Cap}_{\mathrm{veh}}(i,j), \forall (i,j) \in L_R, \tag{8}$$

$$n_{\min} + n_E \leq \sum_{(i,j) \in L_R} z(i,j) \leq n_{\max} + n_E, \tag{9}$$

$$x(k) \in \{0,1\}, \tag{10}$$

$$y(i,j) \in \mathbb{Z}, \tag{11}$$

Table 2: Definition of key sets, parameters, and decision variables.

Notation	Description
Sets	
N	Set of physical nodes in the network, indexed by i, j, and n.
L	Set of physical links in the network, indexed by (i, j).
L_R	Subset of riding links for trains (including candidate links for construction).
K	Set of feasible paths for all O/D pairs, indexed by k.
P	Set of travelers, indexed by p.
R	Set of O/D pairs, indexed by r.
Parameters	
$\mu_p(k)$	Binary constant, which is equal to 1 if path k is a feasible path of traveler p, and 0 otherwise. We note that a path is dedicated to a unique traveler.
$c_p(k)$	Generalized travel cost for traveler p through path k.
$c_v(i, j)$	Train operating cost on link (i, j).
$f(i, j)$	Construction cost of link (i, j).
$\delta_{i,j}(k)$	Utilization coefficient of link (i, j) for path k (=1, if the link is a part of path k; = 0, otherwise).
$\text{Cap}_{\text{pax}}(i, j)$	Passenger carrying capacity on each train.
$\text{Cap}_{\text{veh}}(i, j)$	Train passing capacity on each link during a certain time period.
$br_{i,j}$	Baseline resource on link (i, j) (>0, if additional supply is provided on the link; <0, if demand request is on the link).
α	Coefficient in the objective function for total travelers' cost.
β	Coefficient in the objective function for operating cost.
γ	Coefficient in the objective function for construction cost.
Variables	
$x(k)$	1, if path k is selected by a traveler; 0, otherwise
$y(i, j) \in \mathbb{Z}$	Number of trains needed on link (i, j) to satisfy travel demand
$z(i, j)$	1, if link (i, j) is constructed; 0, otherwise

$$z(i, j) \in \{0, 1\}. \tag{12}$$

Constraint (5) guarantees that each traveler has a path and chooses only one path to perform the journey. It can be a realistic path in the physical network. It can also be a virtual path to achieve travel for unserved demand or to describe an alternative travel mode such as driving. Because resources (available space and/or seats) are limited, the passenger carrying capacity of each train is strictly constrained in equation (6). The capacity parameter $\text{Cap}_{\text{pax}}(i, j)$ can be given according to the number of seats and load factor (which can be variable according to authority regulations or government policies such as restrictions during COVID-19). Note that we only consider the travel demand or transport supply between the terminals of the links. A line can be formed by more than one link. The parameter $br_{i,j}$ is introduced to describe the intermediate demand or supply of links (i, j). Meanwhile, owing to the technical conditions of railways (such as communication signals and train control equipment), the number of trains on each link of rail transit during a certain period is also constrained by equation (7). To guarantee the investment benefit, if link (i, j) is constructed, the number of trains needed on the link should be no less than a specific proportion m of the link's train passing capacity $\text{Cap}_{\text{veh}}(i, j)$, as expressed in equation (8). Constraint (9) limits the number of new links if external effects exist. Finally, constraints (10)–(12) ensure variable domains for $x(k)$, $y(i, j)$, and $z(i, j)$, respectively.

In some realistic situations, the construction cost may be strictly constrained owing to the limit of investment B. In this case, the third term $\gamma \cdot z_c$ in objective (1) is substituted with a constraint (13) and removed from the objective function.

$$\sum_{(i,j)\in L_R} [f(i, j) \cdot z(i, j)] \le B. \tag{13}$$

5. Column Generation-Based Algorithm

The network design problem is NP-hard [13, 14], and it is nearly impractical to enumerate all feasible paths for all O/D pairs in a real-size network owing to a large number of variables. Fortunately, we find that in any feasible solution of our path-based formulation, only a small part of the variables (equal to the number of travelers) are not equal to 0. This motivated us to design a column generation-based algorithm to search for tight lower bounds and high-quality solutions. Column generation is an efficient algorithm for solving linear programming with numerous variables by considering only a subset of variables and identifying new hopeful variables by solving a particular pricing subproblem. Column generation has been successfully applied to solve some transportation problems, for example, Borndörfer et al. [37], Park et al. [38], and Capelle et al. [39] for line planning and location-routing problems.

Following the column generation method proposed by Wang et al. [26], we first relaxed the binary and integer variables in the proposed network design model to be continuous and regarded it as the master model in column generation. Second, we constructed a restricted master model by considering only a subset of paths K_c for all travelers. After solving the restricted master model to optimality (e.g., state-of-the-art commercial solvers), the resulting optimal dual values were used to construct a pricing subproblem model. The pricing subproblem model aimed to identify a new path by minimizing the reduced cost of this new path. If a new path with a corresponding negative reduced cost was identified, we added a new path to K_c, updated the restricted master model based on K_c, and proceeded to the next column generation iteration. Otherwise, the optimal solution from the restricted master model is also optimal for the master model according to the duality theory of linear programming. The associated optimal objective function value of the restricted master model (denoted as RMM_OFV) provided a lower bound for the network design model. At this point, we attempted to compute an integer solution for the network design model using the paths identified by column generation and terminated the algorithm 1. A flow chart of the column generation-based algorithm is shown in Figure 3.

5.1. Restricted Master Model and the Dual Problem. We constructed the restricted master model by relaxing constraints (10)–(12) and replacing K in equations (2), (5), and (6) with K_c ($K_c \subset K$). Without considering the strict constraint (13) of the construction budget, the restricted master model can be formulated as follows. Constraints (7)-(9) keep the same as described in subsection 4.3. So we do not repeat them here.

$$\begin{aligned} \min z' &= \alpha \cdot z_p{'} + \beta \cdot z_v + \gamma \cdot z_c, \\ z_p{'} &= \sum_{p\in P} \sum_{k\in K_c} \left[\mu_p(k) \cdot c_p(k) \cdot x(k)\right], \\ z_v &= \sum_{(i,j)\in L_R} \left[c_v(i,j) \cdot y(i,j)\right], \\ z_c &= \sum_{(i,j)\in L_R} \left[f(i,j) \cdot z(i,j)\right]. \end{aligned} \tag{14}$$

Subject to

$$\sum_{k\in K_c} \left[\mu_p(k) \cdot x(k)\right] = 1, \quad \forall p \in P, \tag{15}$$

$$\sum_{p\in P} \sum_{k\in K_c} \left[\mu_p(k) \cdot \delta_{i,j}(k) \cdot x(k)\right] \leq \mathrm{Cap}_{\mathrm{pax}}(i,j) \cdot y(i,j) + br_{i,j}, \quad \forall (i,j) \in L_R, \tag{16}$$

Let π_p ($\forall p \in P$) and $\varphi_{i,j}$ ($\forall (i,j) \in L_R$) be the dual variables corresponding to equations (15) and (16), respectively; let $\omega_{i,j}$, $\varepsilon_{i,j}$ ($\forall (i,j) \in L_R$) be the dual variables corresponding to constraints (7) and (8), respectively; constraint (9) is divided into two parts to describe the limits of minimum and maximum number of new links; and let $\lambda_{\min}$, $\lambda_{\max}$ be the corresponding dual variables.

The dual problem of the restricted master model can be expressed as a linear model represented by equations (17)–(25). The optimal dual values of the restricted master model (denoted as $X = \{\pi_p^*, \varphi_{i,j}^*, \omega_{i,j}^*, \varepsilon_{i,j}^*, \lambda_{min}^*, \lambda_{\max}^* | \forall p \in P, \forall (i,j) \in L_R\}$) were obtained by solving the dual problem to optimality.

$$\max \sum_{p\in P} \pi_p + \sum_{(i,j)\in L_R} \left[br_{i,j} \cdot \varphi_{i,j}\right] + \sum_{(i,j)\in L_R} \left[m \cdot Cap_{veh}(i,j) \cdot \varepsilon_{i,j}\right] + \left(n_{min} + n_E\right) \cdot \lambda_{min} + \left(n_{max} + n_E\right) \cdot \lambda_{max}. \tag{17}$$

Subject to

$$\sum_{p\in P} \left[\mu_p(k) \cdot \pi_p\right] + \sum_{p\in P} \sum_{(i,j)\in L_R} \left[\mu_p(k) \cdot \delta_{i,j}(k) \cdot \varphi_{i,j}\right] \leq \alpha \cdot \sum_{p\in P} \left[\mu_p(k) \cdot c_p(k)\right], \forall k \in K_c, \tag{18}$$

$$-\mathrm{Cap}_{\mathrm{pax}}(i,j) \cdot \varphi_{i,j} + \omega_{i,j} + \varepsilon_{i,j} \leq \beta \cdot c_v(i,j), \forall (i,j) \in L_R, \tag{19}$$

$$-\mathrm{Cap}_{\mathrm{veh}}(i,j) \cdot \omega_{i,j} + \lambda_{\min} + \lambda_{\max} \leq \gamma \cdot f(i,j), \forall (i,j) \in L_R, \tag{20}$$

$$\varphi_{i,j} \leq 0, \forall (i,j) \in L_R, \tag{21}$$

$$\omega_{i,j} \leq 0, \forall (i,j) \in L_R, \tag{22}$$

$$\varepsilon_{i,j} \geq 0, \forall (i,j) \in L_R, \tag{23}$$

$$\lambda_{\min} \geq 0, \tag{24}$$

$$\lambda_{\max} \leq 0. \tag{25}$$

5.2. Pricing Subproblem and Reduced Cost. The reduced cost was obtained by solving the pricing subproblem model to optimality. Constraint (18) was constructed using paths, and each path was only feasible for a unique traveler. We introduced decision variables μ_p and $\delta_{i,j}$ for the new path to be identified for traveler p, which was similar to $\mu_p(k)$ and $\delta_{i,j}(k)$ for path k (as described in Table 2). Specifically, $\mu_p = 1$ if the new path can be used by traveler p, and 0 otherwise; $\delta_{i,j} = 1$ if the new path passes through link (i,j), and 0 otherwise. Hence, the pricing subproblem model for each traveler p can be expressed as the following mixed-integer linear model, based on the optimal dual values provided by solving (the dual of) the restricted master model to optimality.

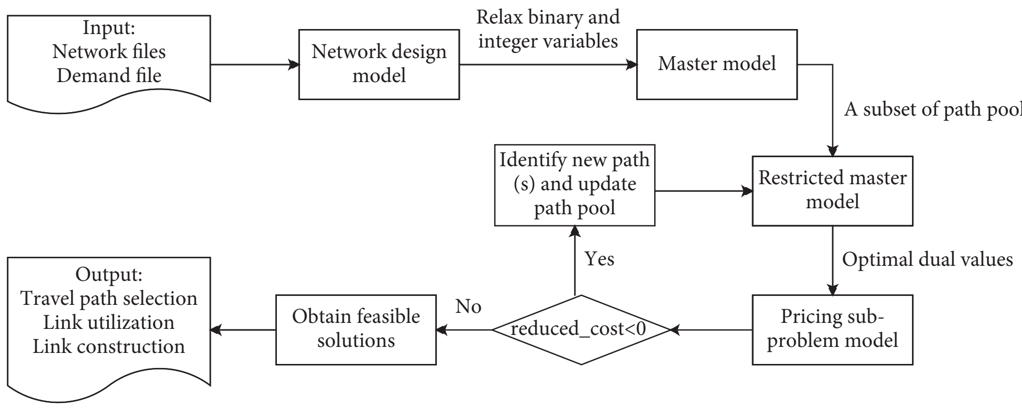

Figure 3: Flow chart of the column generation-based algorithm.

$$\min \alpha \cdot c_p - \pi_p^* \cdot \mu_p - \sum_{(i,j)\in L_R} \left(\varphi_{i,j}^* \cdot \delta_{i,j} \cdot \mu_p\right),$$
$$c_p = \sum_{(i,j)\in L} \left[tc_{i,j}(p) \cdot \delta_{i,j}\right]. \quad (26)$$

Subject to

$$\sum_{(i,j)\in L_{out}(n)} \delta_{i,j} - \sum_{(i,j)\in L_{in}(n)} \delta_{i,j} = \begin{cases} 1, & n = N_O(p) \\ -1, & n = N_D(p) \\ 0, & \forall n \cup \{N_O(p) \cdot N_D(p)\} \end{cases} \quad (27)$$

where $L_{\text{out}}(n)$ is the subset of outbound links for node n and $L_{\text{in}}(n)$ is the subset of inbound links for node n.

A new path for traveler p can be identified by minimizing the objective function (26). In equation (26), $tc_{i,j}$ denotes the travel cost through link (i, j). Constraint (27) is the flow-balance constraint to ensure a complete and continuous path for traveler p from origin node $N_O(p)$ to destination node $N_D(p)$.

Note that the number of travelers is always very large in real-size scenarios. If the pricing subproblem model is constructed traveler-by-traveler, it will increase the computational burden. Even the memory of a computer is insufficient for extremely large computations. We note that for travelers with the same O/D pair, the only difference in reduced cost, as expressed in (26), is the second term $\pi_p^* \cdot \mu_p$, which is not related to the identified new path. Meanwhile, feasible paths for each O/D pair are available for all travelers with the same O/D pair. Therefore, we construct the pricing subproblem model for each O/D pair r as the shortest path problem, whose origin node is $N_O(r)$ and destination node is $N_D(r)$. After identifying feasible paths, it is easier and more practical to calculate the reduced cost for each traveler. Then, we can perform a disaggregated assignment in the network design model for each traveler (even though travelers with the same O/D pair may choose different paths to achieve optimality).

$$\min \alpha \cdot \sum_{(i,j)\in L} \left[tc_{i,j}(p) \cdot \delta_{i,j}\right] - \sum_{(i,j)\in L_R} \left(\varphi_{i,j}^* \cdot \delta_{i,j} \cdot \sum_{p\in P} \mu_p\right). \quad (28)$$

Subject to

$$\sum_{(i,j)\in L_{out}(n)} \delta_{i,j} - \sum_{(i,j)\in L_{in}(n)} \delta_{i,j} = \begin{cases} 1, & n = N_O(r), \\ -1, & n = N_D(r), \\ 0, & \forall n \cup \{N_O(r) \cdot N_D(r)\}. \end{cases} \quad (29)$$

The number of O/D pairs was controllable, and they were independent. Thus, it is easier to handle, even with parallel computing. By solving the subproblem for each O/D pair to optimality, a new path can be identified. All travelers with the same origin and destination nodes can use the new path(s) of the O/D pair. We record the objective function value of the resulting optimal/feasible solution for the O/D pair r as PSM_OFV(r).

To check whether to terminate the algorithm, we need to calculate the reduced cost for each traveler by adding a constant term (i.e., $\pi_p^* \cdot \mu_p$) to PSM_OFV(r). The reduced cost S_p for traveler p with the new path is expressed in equation (30), where P_r is the traveler set with the same origin and destination nodes as the O/D pair r.

$$S_p = PSM\_OFV(r) - \pi_p^* \cdot \mu_p, \forall p \in P_r. \quad (30)$$

If $S_p < 0$ for any traveler p, the new path will benefit the master problem. Therefore, one or more paths for the corresponding O/D pairs were added to the current path pool, K_c. Each new path is feasible for all travelers with the same O/D pair. Subsequently, we updated the restricted master model and resolved the updated restricted master model to obtain a better solution. Otherwise, if the reduced costs for all travelers are non-negative, the current optimal solution of the restricted master model is already optimal for the master model. Thus, the termination condition was reached.

5.3. Algorithm Procedure. When the algorithm terminates, the optimal solutions of the restricted master model may be infeasible for the network design model owing to the relaxation of constraints (10)–(12). In the case of infeasibility, we solve the restricted master model (updated in the last column generation iteration) using constraints (10)–(12) to obtain a feasible solution. The complete algorithm procedure and pseudocode of the column generation-based algorithm are provided in Algorithm 1.

In the column generation-based algorithm, we first need to initialize a path pool as a feasible solution. Thus, we constructed virtual paths for each *O*/*D* pair by connecting the origin and destination nodes through a virtual node. The capacity of the virtual paths was set to be very large to satisfy all unserved demands. It can also be regarded as an alternative transportation mode to achieve a journey with an appropriate path travel cost. To encourage travelers to use the designed rail transit system as much as possible, the travel cost for the virtual paths can be set to be very large. In other words, if a traveler selects a virtual path, the large travel cost is regarded as a penalty. To further improve the quality of the solutions, a few feasible paths in the physical network can be obtained by the shortest path algorithm or heuristic algorithm and added to the initial path pool before generating new paths.

6. Case Study

6.1. Network Formation. We consider a travel corridor in a metropolitan area as an example, as shown in Figure 4, to conduct computational experiments. Multiclass rail transit lines with high travel demand exist along the corridor. The transportation mode of access to or egress from rail transit stations is assumed to be single. Several feasible plans have been proposed as candidate links for construction to shorten the total travel time and provide additional transport supply. Specifically, link (3, 4) is a suburban/metropolitan railway; links (7, 8), (8, 10), and (8, 12) are urban subway lines; to avoid the disturbance of long-distance travel, link (15, 20) is a virtual high-speed railway link representing travelers for long-distance intercity journeys with high-speed railways; links (4, 7), (4, 10), (10, 12), and (10, 15) are new candidate links proposed to be selected, which is an external input. In detail, link (4, 7) can be regarded to invest in some infrastructure renovation for run-through service between the metropolitan line (3, 4) and urban subway line (7, 8). Links (4, 10) and (10, 12) are suitable for construction as new candidate links to connect the metropolitan railway link (3, 4) and the high-speed railway terminal node 15. If link (10, 15) is constructed, trains can run directly through the metropolitan lines and the virtual high-speed line. Travelers can board trains for long-distance travel without transfers. It is also proven that network design is a foundation for tactical and operational planning, which is beneficial to train operation and management.

To make the illustration figure brief, we regard all access/egress, take-on/take-off, and transfer links as walking links in the example. All walking and riding links in the network are bidirectional, with the same travel time for both directions. The taking-on and taking-off links between the station entrance and platform have different travel times owing to the different processes. For example, railway and subway stations in China require security checks before boarding.

6.2. Demand and Parameters. It is assumed that there are a total of 1,000 travelers with 30 *O*/*D* pairs and the same value of time (equal to 60 CNY/h). The travel demand between demand nodes and the construction costs of the candidate links are listed in Table 3, respectively. Table 4 shows the baseline resource on each existing riding link is assumed to be 60, which indicates that there are 60 demand requests for short-haul travel on each link. We assume that the operating cost (which may include the maintenance expense) is related only to the operating time. Thus, we define the operating cost $c_v(i, j)$ per train on link (i, j) as equal to the travel time on the link, which is multiplied by β for the objective function of the generalized costs. In the initialization process, we constructed a virtual path and provided at most three feasible paths for each *O*/*D* pair as the initial path pool. The virtual path directly connects the origin and destination nodes. It is generated to guarantee that each traveler has at least one feasible path to finish the trip. To encourage travelers to use the designed rail transit network as much as possible, the travel cost for the virtual paths was set to 9999. The feasible paths given in the initialization process are used to quickly obtain a good solution.

Let us consider the time period of the case study to be per hour. Four scenarios were designed: scenarios 1 and 2, where the number of trains on each riding link $\mathrm{Cap}_{\mathrm{veh}}(i, j)$ is sufficiently large and the coefficient of construction cost γ is small; scenario 3—based on acenario 1—we increased the coefficient of construction cost γ; and scenario 4—based on scenario 3—we decreased the number of trains on each riding link $\mathrm{Cap}_{\mathrm{veh}}(i, j)$ per hour. We assumed that the capacity of passengers for each train on each link, $\mathrm{Cap}_{\mathrm{pax}}(i, j)$ is identical. Additionally, the capacity of trains on each link $\mathrm{Cap}_{\mathrm{veh}}(i, j)$ is also identical. In the case study, the units of traveler, operator, and construction costs are minute, 100 CNY, and 100 million CNY, respectively. Therefore, the coefficients, α, β, and γ, not only reflect the weights of the three types of costs but also have a function of normalization.

Table 5 compares the parameters of the four scenarios. The coefficients of traveler cost per CNY, operator cost per 100 CNY, and construction cost per 100 million CNY are given in rows 1–3, respectively. Here, we focus only on the sensitivity of construction cost. A larger coefficient γ means that decision-makers are more sensitive to construction costs. Row 4 sets the proportion m to guarantee that the investment benefit is 0.2 for all four scenarios. The capacity of passengers for each train on each link and the capacity of trains on each link per hour are provided in rows 5–6. Row 7 presents the solution. Scenario 1 was directly solved without the proposed algorithm, and scenarios 2–4 were solved using a column generation-based algorithm. We did not limit the number of new links in this case study.

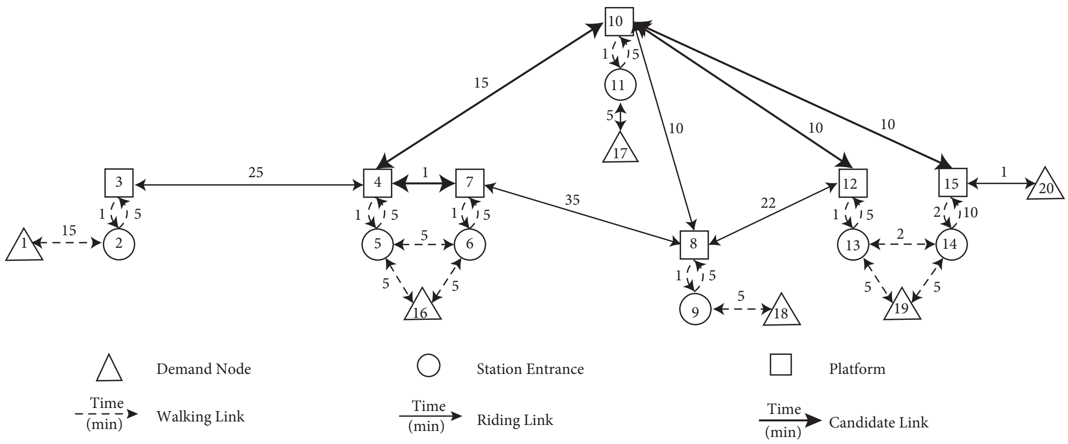

Figure 4: A network along a metropolitan corridor for traveling with multiclass rail transit.

Input: Nodes, links, and travel demand in the network.
Output: Path selected by each traveler $x(k)$, number of trains needed on each riding link $y(i,7)$, and links to be constructed $z(i,7)$.

//Initialization
Create all nodes and all links.
Set PathPool = 0.
for $r \in R$ **do**
    Construct a virtual path from origin node to destination node for OD pair r to finish the trip through a virtual node.
    PathPool = PathPool U Virtual Path.
    Perform a shortest path algorithm or heuristic algorithm to obtain one or a few feasible paths.
    PathPool = PathPool U Feasible Path.
    Create path list Ke with traveler's attribute.
    Create path-link incidence matrix and traveler-path selection matrix.
//Loop
while True **do**
    //Restricted master model for network design with relecation
    Construct restricted master model based on Ke and solve it to optimality.
    Record the optimal objective function value RM M_OF'V, the optimal solutions $2r(k)$ $(\forall k \in K_C)$, $y(i,j)$ $(\forall (i,j) \in L_R)$ $z(i,j)$ $(\forall (i,j) \in L_R)$, and the optimal dual values $X = \{\pi_p^*, \phi_{i,j}^*, \omega_{i,j}^*, \varepsilon_{i,j}^*, \lambda_{\min}^*, \lambda_{\max}^* | \forall p \in P, \forall (i,j) \in L_R\}$
    //Pricing subproblem model for finding one or more new paths
    Set flag = 0.
    for $r \in R$ **do**
        Get the new path's utilization incidence p, for traveler p.
        Construct pricing subproblem model for O/D pair r based on optimal dual values X and solve it to optimality.
        Record the optimal objective function value PSM_OFV with the new path and the optimal solutions of path-link incidence $\delta_{i,j}$ $(\forall (i,j) \in L_R)$.
        //Reduced cost for each traveler
        for $p \in P$ **do**
            Calculate the reduced cost S_p, with the new path.
            if $S_p < 0$ and NewPath not in PathPool **then**
            PathPool = PathPool ∪ New Path.
            flag = flag+1.
    if flag = 0 **then**
    break
    else
        Update path list K_C.
        Update path-link incidence matrix and traveler-path selection matrix.
LB = RMM_OFV.
//Feasible solutions for the master problem
Add $x(k) \in \{0,1\}$ $(\forall k \in K)$, $y(k) \in \mathbb{Z}$, $z(i,j) \in \{0,1\}$ $(\forall (i,j) \in L_R)$, to restricted master model, and solve the restricted master model to optimality.
Record the optimal objective function value Best_OFV, and the optimal solutions Bestx.
return LB, Best_OFV, BestX.

Algorithm 1: Column generation-based algorithm.

Table 3: Travel demand between demand nodes.

From node	To node					
	1	16	17	18	19	20
1	—	15	15	40	15	20
16	15	—	45	65	15	40
17	20	40	—	50	35	40
18	45	55	50	—	45	50
19	10	20	25	30	—	20
20	30	40	35	60	15	—

Table 4: Construction costs of candidate links.

Link	Construction cost/100 million CNY
(4, 7)/(7, 4)	2.5
(4, 10)/(10, 4)	12.5
(10, 12)/(12, 10)	15
(10, 15)/(15, 10)	20

Table 5: Comparison of the four scenarios.

Scenario	1	2	3	4
(1) α	1	1	1	1
(2) β	10	10	10	10
(3) γ	10	10	200	200
(4) m	0.2	0.2	0.2	0.2
(5) $\text{Cap}_{\text{pax}}(i, j)$	100	100	100	100
(6) $\text{Cap}_{\text{veh}}(i, j)$	5	5	5	2
(7) Solution	—	Column generation-based algorithm		

6.3. Results' Analysis. The mixed-integer linear programming model for solving the rail transit network design problem was implemented using Python 3.6, and the Gurobi Optimizer (version 9.1.1) was used to obtain the optimal solutions. The experiment was carried out on a personal computer with an Intel Core i7-9750H CPU at 2.59 GHz and 16 GB of RAM, running Windows 10 Pro 64-Bit. The results for the four scenarios are summarized in Table 6.

Row 1 shows the total number of variables used to solve the network-design model. Rows 2–4 report the optimal objective function value, lower bound, and computation time, respectively. Row 5 shows the relative gap between the objective function value and the lower bound. Rows 6–9 list the optimal results for the total travel time for all travelers, number of *O*/*D* pairs whose travel time is reduced, optimal construction links, and total construction cost, respectively. Compared with scenario 1, the number of variables in scenarios 2–4 was significantly reduced with the proposed column generation-based algorithm. In a larger network with more travel demand, the reduction in variables and computation time will be more significant.

Figure 5 illustrates the total travel time for all *O*/*D* pairs. This shows an overall improvement in the rail transit network and link performance. Under the same conditions of *O*/*D* demand, link capacity, and train capacity as those in scenarios 1 and 2, the total travel time for 1,000 travelers is 60,075 min by solving the model without any new links added to the existing network. In total, 24 out of the 30 *O*/*D* pairs benefited from the new links. To benefit travelers the most, we set the value of γ to be relatively small in scenarios 1 and 2. With the new links (4, 10)/(10, 4), (12, 10), and (10, 15)/(15, 10) constructed in scenario 1, the total travel time is reduced by 34.5%. Comparing scenarios 1 and 2, even though the objective value of scenario 1 is slightly better because of an unconstructed link (10, 12) in scenario 2, the total travel time is less in scenario 2 from the perspective of travelers.

With regard to investment, we increased the value of γ to 200 in scenario 3. The optimal construction links are (4, 10)/(10, 4) and (10, 15)/(15, 10). The total construction cost decreased, but the total travel time increased by approximately 9%, compared to scenarios 1 and 2. This is because some travelers from node 18 (or 19) to node 19 (or 18) cannot use link (10, 12)/(12, 10) to shorten the travel time. However, all travelers can use the shortest paths with four new links. In fact, link (15, 10) can have the same function as link (12, 10) if the enter procedures on link (14, 15) can be simplified.

When considering the link capacity constraints in scenario 4, only two trains can operate on each link per hour. The optimal construction links are the same as those in scenario 3. However, some travelers through links (4, 10)/(10, 4) and (8, 10)/(10, 8) have to be reassigned to other paths. Therefore, not all travelers can use the shortest paths for some *O*/*D* pairs. Specifically, only 20 out of 50 travelers from 18 to 20 can use the shortest path: 18-9-8-10-15-20. The remaining 30 travelers had to travel through path 18-9-8-12-13-14-15-20, including a transfer process. For travelers from 20 to 18, only 15 out of 60 can use the shortest path through links (15, 10) and (10, 8). Others have to use link (12, 8) with a transfer. All travelers from 16 to 18 and from 18 to 16 have to select another path(s) through link (7, 8)/(8, 7) in scenario 4, instead of using the shortest path(s) through link (4, 10)/(10, 4) in scenario 3. In all four scenarios, all travelers can be served by the rail transit network. Virtual paths are not used.

Figure 6 shows the volume of travelers on each riding link in the already functioning network. Compared with the existing network, the volumes on the transit links (7, 8)/(8, 7) and (8, 12)/(12, 8) are reduced by approximately 80% in scenario 1. The relief of these links could increase local transport capacity and improve the level of service for intermediate demand. Undoubtedly, the robustness of the entire rail transit network will be enhanced simultaneously. When focusing on the four links under scenarios 3 and 4, the volumes increased slightly in scenario 4 owing to capacity constraints. This also shows that some travelers cannot use the shortest path(s), which proves the significance of the disaggregated assignments.

Next, we keep the input physical network the same and set the values of α, β, and γ to be 1, 10, and 200, respectively, which are the same as in scenarios 3 and 4. Based on Table 3, we increased the travel demand for each *O*/*D* pair by multiplying the same adjustment factor by greater than one. Specifically, the total travel demand increased from 1,000 to 10,000 travelers. For every 1,000 additional travelers, calculate the adjustment factor of travel demand and perform

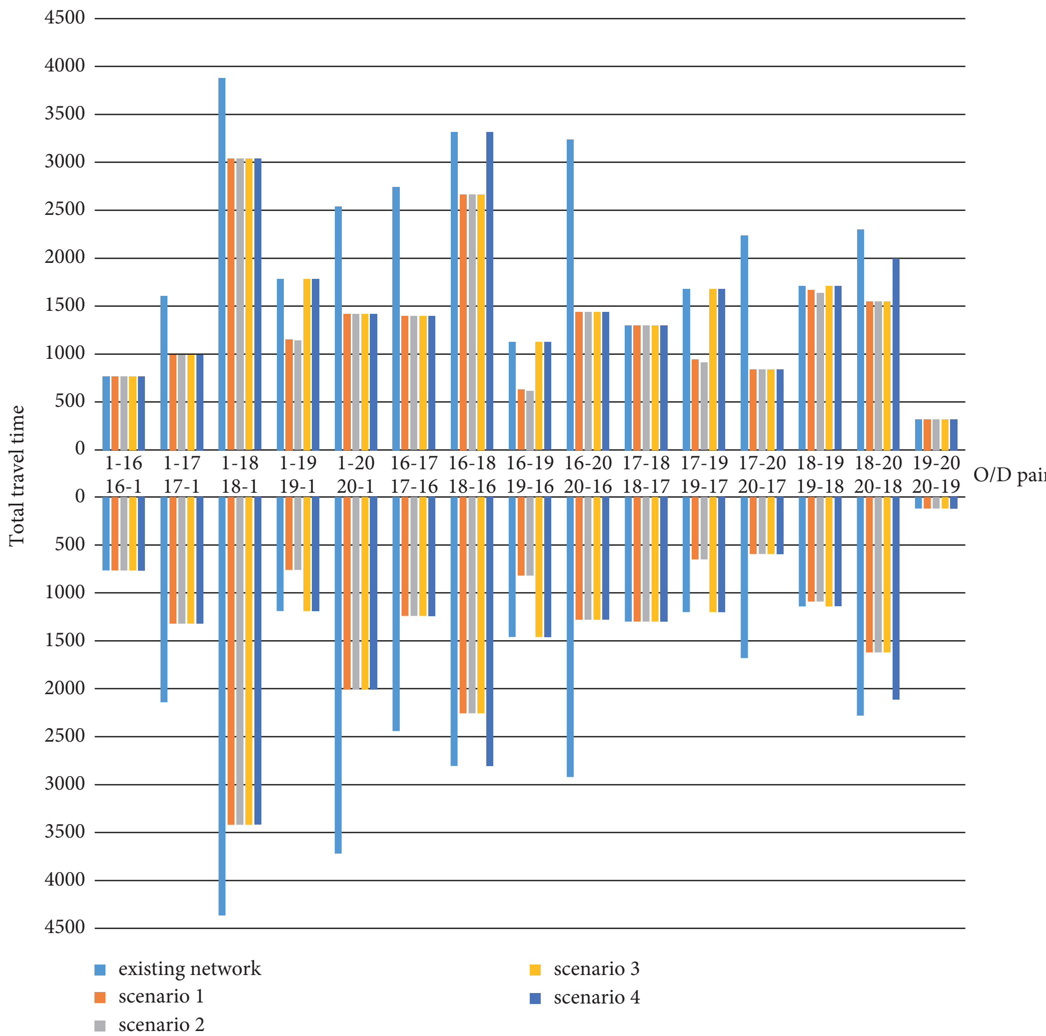

Figure 5: Total travel time for all O/D pairs.

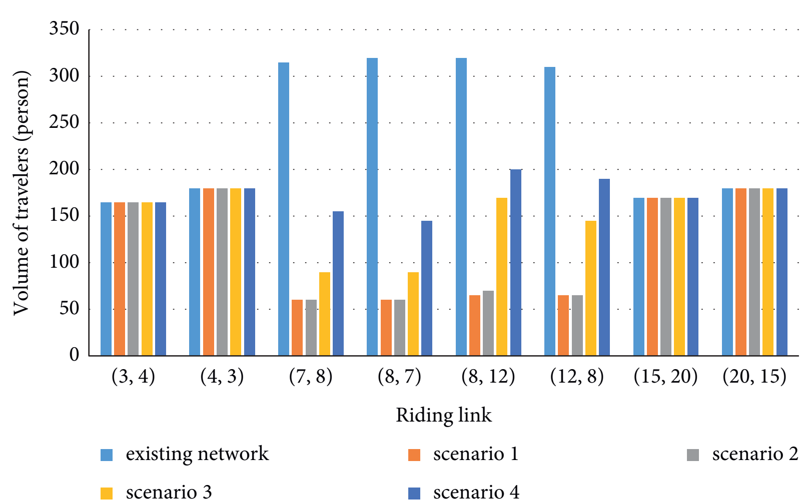

Figure 6: Volume of travelers on riding links in the already functioning network.

Table 6: Comparison of the results under four scenarios.

Scenario	1	2	3	4
(1) Number of variables	568,100	115,100	115,100	117,100
(2) Objective value	44,450	44,505	60,560	62,805
(3) Lower bound	44,448	43,032.2	48,699	56,484
(4) Computation time (seconds)	48.4	27.1	26.8	30.3
(5) Gap (%)	0.0000	3.3092	19.5855	10.0645
(6) Total travel time (minute)	39,370	39,275	42,940	45,085
(7) Number of *O*/*D* pairs with reduced travel time	24	24	16	14
(8) Optimal construction link	(4, 10) (10, 4) (12, 10) (10, 15) (15, 10)	(4, 10) (10, 4) (10, 12) (12, 10) (10, 15) (15, 10)	(4, 10) (10, 4) (10, 15) (15, 10)	(4, 10) (10, 4) (10, 15) (15, 10)
(9) Total construction cost/100 million CNY	80	95	65	65

Table 7: Comparison of the results with two solution approaches.

Instance name	Directly solve the model		Solve the model with column generation-based algorithm				Usage of virtual path	Improvement (%)
	Number of variables (10,000)	Objective function value	Number of variables (10,000)	Lower bound	Objective function value	Gap (%)		
p1000v10	56.81	56,975	11.51	45,791.5	60,560	24.39	No	6.29
p1000v20	56.81	56,975	11.51	44,285.2	60,560	26.87	No	6.29
p2000v10	113.61	98,990	23.01	90,479	104,250	13.21	No	5.31
p2000v20	113.61	98,990	23.01	87,466.5	104,250	16.10	No	5.31
p3000v10	170.41	141,250	34.51	135,166.5	145,905	7.36	No	3.30
p3000v20	170.41	141,250	34.51	130,648	145,905	10.46	No	3.30
p4000v10	227.21	183,460	43.61	180,364	187,780	3.95	No	2.35
p4000v20	227.21	183,210	43.61	173,829	187,480	7.28	No	2.33
p5000v10	284.01	234,340	54.51	229,704	235,670	2.53	No	0.57
p5000v20	284.01	225,170	54.51	217,010.2	229,045	5.25	No	1.72
p6000v10	340.81	1,275,390	66.61	1,270,040	1,275,460	0.42	Yes	0.01
p6000v20	340.81	267,320	65.41	260,191.5	270,600	3.85	No	1.23
p7000v10	397.61	4,797,370	77.71	4,791,410	4,797,370	0.12	Yes	0.00
p7000v20	397.61	309,280	76.31	303,373	312,205	2.83	No	0.95
p8000v10	454.41	8,327,460	88.81	8,321,202	8,327,350	0.07	Yes	0.00
p8000v20	454.41	352,700	87.21	347,724	355,120	2.08	No	0.69
p9000v10	511.21	11,858,300	99.91	11,852,222	11,858,300	0.05	Yes	0.00
p9000v20	511.21	400,765	98.11	395,926.5	402,525	1.64	No	0.44
p10000v10	568.01	15,488,000	111.01	15,482,478	15,488,000	0.04	Yes	0.00
p10000v20	568.01	450,570	109.01	444,679	450,560	1.31	No	0.00

two instances under two different train passing capacities $Cap_{veh}(i, j)$ for each link (i, j). Table 7 shows the key information of each instance solved using two different approaches, which can reflect the effect of the column generation-based algorithm. The first column, "Instance name," in the table indicates the total travel demand and train passing capacity for each link. For example, "p1000v10" means that 1,000 travelers need to travel on the network and at most 10 trains running on each rail transit link during a time period. The fifth column, "Lower bound," is the objective function value of the restricted master model in the column generation process. The improvement between the objective function values obtained with and without the column generation-based algorithm was calculated in the last column. When directly solving the model using the Gurobi solver, the default convergence condition can be satisfied for all instances.

As shown in Table 7, when $Cap_{veh}(i, j)$ is 10, some travelers must travel along the virtual path. In other words, some of the travel demands cannot be satisfied by the rail transit network owing to capacity constraints. The gap and improvement shown in Table 7 further prove that the column generation-based algorithm can obtain high-quality solutions. Thus, the proposed algorithm is feasible for solving large-scale problems. Figure 7 illustrates the number of variables with and without the designed algorithm when the link capacity $Cap_{veh}(i, j)$ was 20. Obviously, as the travel demand increases, the column generation-based algorithm performs better in controlling the number of variables more significantly. According to the test, when the number of travelers continuously increases to 28,000,

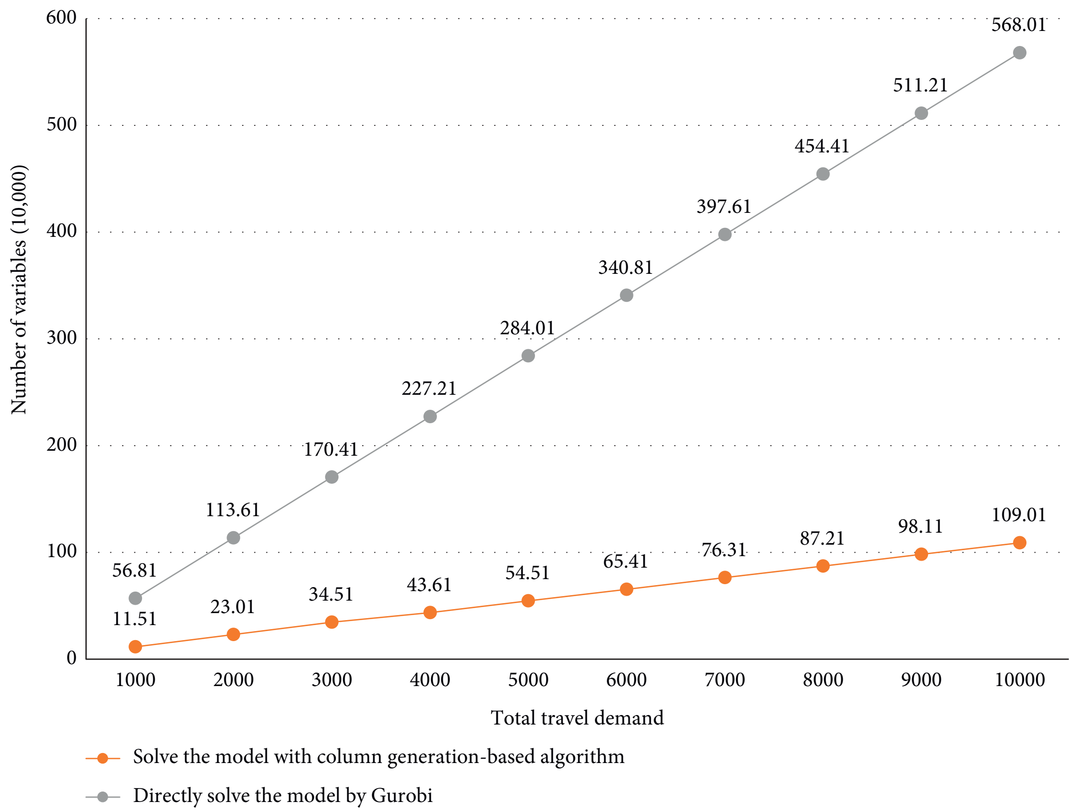

Figure 7: Comparison of a number of variables with two solution approaches.

the personal computer will be out of memory if it directly solves the model by Gurobi and cannot obtain the optimal solution. However, for a metropolis, it is normal to have a daily ridership larger than 100,000.

Therefore, the proposed model and designed algorithm have wide applicability and are feasible for solving large-scale problems, which could significantly control the number of variables and guarantee a high-quality solution with tight lower bounds. Note that the algorithm is required to build and solve the restricted master model and pricing subproblem model multiple times. In the case study, mixed-integer linear programming models were solved with "gurobipy" in Python. The calculation may be accelerated with other programming languages.

7. Conclusions

In this study, we investigated a rail transit network design problem for expanding the rail transit network in a metropolitan area. To reduce travel time for the full journey and improve the level of service, new facilities are proposed. Meanwhile, we construct paths from origin to destination for each traveler, including the first/last mile, transfers, and multiclass transit modes. A mixed-integer linear programming model was developed to minimize the weighted sum of the total costs for travelers, operators, and investors. The optimization model simultaneously determines the travel path selected by each traveler, the links to be constructed, and the number of trains required for each link. Therefore, the network design from the supply side can precisely match the travel demand, which helps effectively utilize the infrastructure.

In terms of the solution method, the network design problem is NP-hard, and it is impractical to enumerate all feasible paths for each traveler in real-size scenarios. Taking advantage of the path-based model, we propose a column generation-based algorithm to identify new paths iteratively in pricing subproblems using O/D pairs. It can control the number of variables significantly and find tight lower bounds and high-quality (near-optimal) solutions efficiently. A series of experiments were conducted along a metropolitan corridor. The comparative results revealed the validity and practicality of the proposed algorithm for solving the path-based rail transit network design problem.

The model can be further extended to a time-space network with time-dependent demand and train schedules. Future work can also consider a queue-theoretic volume-delay function [40] for path assignment, instead of the strict capacity constraints of passengers onboard. To speed up the computation for larger complicated networks, an efficient heuristic algorithm and shortest path algorithm may be required for initialization and finding new paths, respectively.

Data Availability

The data used to support the findings of this study are described in the main body of the paper. Detailed codes are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant no. 2017YFB1200701).

References

[1] G. Laporte, J. Mesa, F. Ortega, and F. Perea, "Planning rapid transit networks," *Socio-Economic Planning Sciences*, vol. 45, pp. 95–104, 2011.

[2] D. Canca, A. De-Los-Santos, G. Laporte, and J. A. Mesa, "Integrated railway rapid transit network design and line planning problem with maximum profit," *Transportation Research Part E: Logistics and Transportation Review*, vol. 127, pp. 1–30, 2019.

[3] T. L. Magnanti and R. T. Wong, "Network design and transportation planning: models and algorithms," *Transportation Science*, vol. 18, pp. 1–55, 1984.

[4] R. Z. Farahani, E. Miandoabchi, W. Y. Szeto, and H. Rashidi, "A review of urban transportation network design problems," *European Journal of Operational Research*, vol. 229, no. 2, pp. 281–302, 2013.

[5] G. L. Jia, R. G. Ma, and Z. H. Hu, "Review of urban transportation network design problems based on CiteSpace," *Mathematical Problems in Engineering*, pp. 1–22, 2019.

[6] A. Li, D. Wang, and Q. Peng, "A path-based methodology for rail transit network design in a metropolitan area," in *Proceedings Of the 9th International Conference on Railway Operations Modelling and Analysis*, pp. 2532–2549, Beijing, China, June2021.

[7] M. R. Bussieck, T. Winter, U. T. Zimmermann, Discrete optimization in public rail transport," *Mathematical Programming*, vol. 79, no. 1-3, pp. 415–444, 1997.

[8] G. Laporte, Á. Marín, J. A. Mesa, and F. A. Ortega, "An Integrated Methodology for the Rapid Transit Network Design Problem," in *Proceedings of the Algorithmic Methods for Railway Optimization*, pp. 187–199, Bergen, Norway, September2004.

[9] Á. Marín and R. García-Ródenas, "Location of infrastructure in urban railway networks," *Computers & Operations Research*, vol. 36, no. 5, pp. 1461–1477, 2009.

[10] S. A. Bagloee and A. A. Ceder, "Transit-network design methodology for actual-size road networks," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1787–1804, 2011.

[11] R. García, G. A. Astolfi, A. Marín, J. A. Mesa, F. A. Ortega, Analysis of the Parameters of Transfers in Rapid Transit Network Design," in *Proceedings of the 5th Workshop On Algorithmic Methods And Models For Optimization Of Railways. Schloss Dagstuhl-Leibniz-Zentrum für Informatik*, 2006.

[12] L. Cadarso and Á. Marín, "Improved rapid transit network design model: considering transfer effects," *Annals of Operations Research*, vol. 258, no. 2, pp. 547–567, 2015.

[13] G. Laporte, J. A. Mesa, and F. Perea, "A game theoretic framework for the robust railway transit network design problem," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 447–459, 2010.

[14] G. Laporte, A. Marin, J. A. Mesa, and F. Perea, "Designing robust rapid transit networks with alternative routes," *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 54–65, 2011.

[15] K. An and H. K. Lo, "Service reliability–based transit network design with stochastic demand," *Transportation Research Record*, vol. 2467, no. 1, pp. 101–109, 2014.

[16] G. Gutiérrez-Jarpa, G. Laporte, V. Marianov, and L. Moccia, "Multi-objective rapid transit network design with modal competition: the case of Concepción, Chile," *Computers & Operations Research*, vol. 78, pp. 27–43, 2017.

[17] D. A. Hensher and C. Mulley, "Special issue on developments in Mobility as a Service (MaaS) and intelligent mobility," *Transportation Research Part A: Policy and Practice*, vol. 131, pp. 1–4, 2020.

[18] D. Canca, A. De-Los-Santos, G. Laporte, and J. A. Mesa, "A general rapid network design, line planning and fleet investment integrated model," *Annals of Operations Research*, vol. 246, no. 1-2, pp. 127–144, 2014.

[19] Y. T. Peng, Z. C. Li, and P. Schonfeld, "Development of rail transit network over multiple time periods," *Transportation Research Part A: Policy and Practice*, vol. 121, pp. 235–250, 2019.

[20] D. Huang, Z. Liu, X. Fu, and P. T. Blythe, "Multimodal transit network design in a hub-and-spoke network framework," *Transportmetrica: Transportation Science*, vol. 14, no. 8, pp. 706–735, 2018.

[21] G. Laporte and M. M. Pascoal, "Path based algorithms for metro network design," *Computers & Operations Research*, vol. 62, pp. 78–94, 2015.

[22] G. Gutiérrez-Jarpa, G. Laporte, and V. Marianov, "Corridor-based metro network design with travel flow capture," *Computers & Operations Research*, vol. 89, pp. 58–67, 2018.

[23] S. Chai, Q. Liang, and S. Zhong, "Design of urban rail transit network constrained by urban road network, trips and land-use characteristics," *Sustainability*, vol. 11, no. 21, p. 6128, 2019.

[24] Y. Zhou, H. Yang, Y. Wang, and X. Yan, "Integrated line configuration and frequency determination with passenger path assignment in urban rail transit networks," *Transportation Research Part B: Methodological*, vol. 145, pp. 134–151, 2021.

[25] D. Canca, A. De-Los-Santos, G. Laporte, and J. A. Mesa, "An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem," *Computers & Operations Research*, vol. 78, pp. 1–14, 2017.

[26] D. Wang, J. Zhao, A. D'Ariano, and Q. Peng, "Simultaneous node and link districting in transportation networks: model, algorithms and railway application," *European Journal of Operational Research*, vol. 292, no. 1, pp. 73–94, 2021.

[27] L. Tong, X. Zhou, and H. J. Miller, "Transportation network design for maximizing space–time accessibility," *Transportation Research Part B: Methodological*, vol. 81, pp. 555–576, 2015.

[28] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, and J. C. Muñoz, "Planning, operation, and control of bus transport systems: a literature review," *Transportation Research Part B: Methodological*, vol. 77, pp. 38–75, 2015.

[29] F. Shi, S. Zhao, Z. Zhou, P. Wang, and M. G. Bell, "Optimizing train operational plan in an urban rail corridor based on the maximum headway function," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 51–80, 2017.

[30] V. Guihaire and J.-K. Hao, "Transit network design and scheduling: a global review," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1251–1273, 2008.
[31] N. Van Oort and R. van Nes, "Regularity analysis for optimizing urban transit network design," *Public transport*, vol. 1, no. 2, pp. 155–168, 2009.
[32] H. Sun, J. Wu, L. Wu, X. Yan, and Z. Gao, "Estimating the influence of common disruptions on urban rail transit networks," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 62–75, 2016.
[33] S. Saidi, S. C. Wirasinghe, L. Kattan, and S. Esmaeilnejad, "A generalized framework for complex urban rail transit network analysis," *Transportmetrica: Transportation Science*, vol. 13, no. 10, pp. 874–892, 2017.
[34] Q. Fu, R. Liu, and S. Hess, "A review on transit assignment modelling approaches to congested networks: a new perspective," *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 1145–1155, 2012.
[35] X. Jiang, X. Guo, and B. Ran, "Optimization model for headway of a suburban bus route," *Mathematical Problems in Engineering*, pp. 1–6, 2014.
[36] D. Hörcher and A. Tirachini, "A review of public transport economics," *Economics of Transportation*, vol. 25, Article ID 100196, 2021.
[37] R. Borndörfer, M. Grötschel, and M. E. Pfetsch, "A column-generation approach to line planning in public transport," *Transportation Science*, vol. 41, no. 1, pp. 123–132, 2007.
[38] B. H. Park, Y. I. Seo, S.-P. Hong, and H.-L. Rho, "Column generation approach to line planning with various halting patterns—application to the Korean high-speed railway," *Asia Pacific Journal of Operational Research*, vol. 30, no. 04, Article ID 1350006, 2013.
[39] T. Capelle, C. E. Cortés, M. Gendreau, P. A. Rey, and L.-M. Rousseau, "A column generation approach for location-routing problems with pickup and delivery," *European Journal of Operational Research*, vol. 272, no. 1, pp. 121–131, 2019.
[40] X. Zhou Simon, Q. Cheng, X. Wu et al., "A meso-to-macro cross-resolution performance approach for connecting polynomial arrival queue model to volume-delay function with inflow demand-to-capacity ratio," *Multimodal Transportation*, vol. 1, no. 2, Article ID 100017, 2022.