IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 7, 2021, accepted February 17, 2021, date of publication February 22, 2021, date of current version March 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3060816

Agent-Based Optimizing Match Between Passenger Demand and Service Supply for Urban Rail Transit Network With NetLogo

JIAMIN ZHANG
College of Transportation, Shandong University of Science and Technology, Qingdao 266590, China
e-mail: zjmsunrise@126.com

ABSTRACT Both passenger demand and service supply are among the most important factors that determine the performance of urban rail transit system. It is not easy to find out optimal solution for the match between the passenger demand and service supply with traditional methods, due to the complexity of the combinatorial intelligent supply — demand matching problem. In order to get the comprehensively optimal matching degree, this paper transforms the multi-criteria problem into the distributed artificial intelligence optimization by using multi-agent dynamic interaction technique. On the demand side, the dynamic passenger traffic demand with agents is modelled from perspective of boundedly rational travel decision. On the supply side, the dynamic service supply of train traffic with agent is modelled. The headway time is designated as the main decision variable, for the key link between the passenger demand and service supply is the headway time in different time-of-day intervals. To make the passenger demand more closely matched with service supply in urban rail transit network system at the reasonable travel cost and operational cost, the calculation formula for matching degree is proposed, along with the distributed system architecture for agent-based matching mechanism, and the negotiation-based iterative mechanisms for balancing. The proposed methods are validated on the simulation platform NetLogo. The simulation results emphasize the importance of representing the supply side and the demand side jointly/interactively. These findings are meaningful for policies on both development of efficient capacity usage strategies of urban rail transit network and provision of high level of service for passengers.

INDEX TERMS Agent, dynamic passenger traffic demand, dynamic service supply of train traffic, intelligent supply — demand matching, NetLogo, urban rail transit.

I. INTRODUCTION

Urban Rail Transit (abbreviated as URT) system is a particular kind of homogeneous railway system. Punctuality, robustness, regularity (i.e. gaps between two successive departures) and comfort are critical performance measures for URT systems in terms of deviations, randomness, uncertainties and overcrowdings. Nowadays, four of the top ten busiest metro systems globally are in China [1], resulting in that many travelers have to experience the significant recurrent congestion during peak hours on a daily basis. Thus, the encouragement of "reasonable supply, controlled demand and better utilization of resources" has been attracting more and more attentions among URT researchers and practitioners [2]. The global problem faced by the public transport (e.g. urban rail transit) agencies consists of determining how to offer a good-quality service to the passengers while maintaining reasonable asset and operating costs. URT network performance measurement generally depend on the matching degree between the passenger demand and service supply. Both supply variations and demand variations are two sources of uncertainty for all modes of transport [3], which make URT operators difficult even impossible to match them totally.

The match or balance between the passenger demand and service supply is a key issue that the rail infrastructure managers and railway undertakings have to face. In the

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyue Xu.

conditions of networked operations, the entire URT system shares both the service and the ridership. Before the transport service resources are configured for the trip makers, how they utilize a given transit system should be clarified, e.g., the route choice behaviour. A transit assignment model is useful in estimating or predicting such a travel behaviour [4]. As well known, usually transport supply should and could be moreover optimized based on the corresponding demand and capacity [5]. The interrelationship between passenger demand and service supply in the process of transit line planning is a chicken-egg problem [6], i.e., they interact mutually.

The traffic assignment to route-choice modelling can be regarded as the prediction towards the group behavior of individuals. On the demand side, the transit assignment problem is defined as the mapping of passenger demand on a given transit network. The transit system performance depends on the interaction between travel demand and transit network supply. The network flow algorithm cannot be applied to a system optimal dynamic traffic assignment problem with multiple OD pairs [7]. Due to the operational costs and capacity constraints, the reasonable rail managers generally incline to offer a limited number of daily vehicle services. While statistics show that about a quarter of stations in Beijing URT network adopt the passenger flow control measures during peak hours. Based on the analysis of 72 mass rapid transit (MRT) stations in Wuhan, China, Guo and Huang [8] obtained four principal components to explain the potential linkage to MRT ridership, one of which includes the station attributes in the network. Xu *et al.* [9] proposed a method to evaluate and improve the service quality of crowded metros from the point of service components, combining Bayesian network, structural equation modeling, and importance-performance analyses. From a market point of view, service supply or capacity should be oriented to satisfy peak demands, while from an operation plan or timetable standpoint, its considerations are necessary to define the train frequency and train paths trying to fulfil travel demand on a given infrastructure, by maintaining the desired level of service to passengers.

The order parameter is the decisive factor of the system's structure and function. For the matching problem between passenger demand and service supply, the headway time is the order parameter of the URT system. The problem of headway optimization is also referred to as transit network frequencies setting problem. Designating the cost reduction and service level improvement as the objectives, Schmaranzer *et al.* [10] presented a simulation-based headway optimization for urban mass rapid transit networks, for which the population-based evolutionary algorithms and different solution encoding variants are applied. From the transit planning branches of the transportation science, the items of route design and frequency setting compose the transit network design problem [11], which both determines to a large extent the service for the passengers and the operational costs for the operator of the system, and is regarded as one of the most intractable problems in the field of transportation due to its high degree of complexity. As the allocation of resources in a multi-route URT system, the optimal headways are affected by the passenger demands at stations [12], [13]. The main tasks on the supply side include the computation of the transport offer, e.g., the determination of the optimal train frequency/headways per line and the capacity of the trains. The settings of service frequencies determine supply capacity and have significant consequences for level of service and operational costs. One reason of success for the URT system is that the headways between the trains are chosen in such a way that the trains are equally high occupied [14].

The methods to determine dispatching headways for setting the frequencies of the transit services can be classified into two categories [15], i.e., (i) passenger load profile rule-based techniques, (ii) minimizing passenger and operator costs. The planning process for transit network spans every decision that should be taken before the operation of the URT system, which involves the subproblems as transit network design, frequency setting, transit network timetabling. Roughly speaking, its essence lies in the match between the passenger demand and the service supply.

As discussed above, a limited number of studies have been devoted to study the match between passenger demand and supply in the URT network systems. In the current operational practice, the experiences on the matching between passenger demand and service supply mainly regarding the loading factor as the single indicator, while focusing only on the single URT track line. The load factor in URT system can be quantified as the passenger load-to-train capacity ratio in a certain period (e.g. the peak period), the value of which usually varies within a threshold. When the load factor reaches or gets greater than the threshold, an overload phenomenon can be declared, and the congestion inclines to occur. The optimum state of a good match between service supply and passenger demand occurs in URT system when the observed passenger demand is accommodated meanwhile the number of vehicles in use is minimized [16].

Both resilience (i.e., the ability to return to a previous state after a disruption) and robustness (i.e., the amount of stress that can be absorbed before failure) have gathered much interest in the scientific community in recent years. For URT system, resilience and robustness associate greatly with travel time reliability and variability. A single delay could have network-wide effects if the mismatch occurs between the passenger demand and service supply, for these delays will propagate from one train to some of its connectors. The service supply determines whether the infrastructure network has sufficient capacity to meet the passenger demand. Therefore, it is important to simultaneously optimize the transit route structure and the frequency setting. Few works focus on passenger demand together with train service supply by also considering passenger control strategy in the URT network. By using the heuristic procedure, Ceder *et al.* [17] created the even-load and even-headway transit timetables through matching demand and supply. Huang *et al.* [1] focused on load balancing solutions with effective alternative routes in

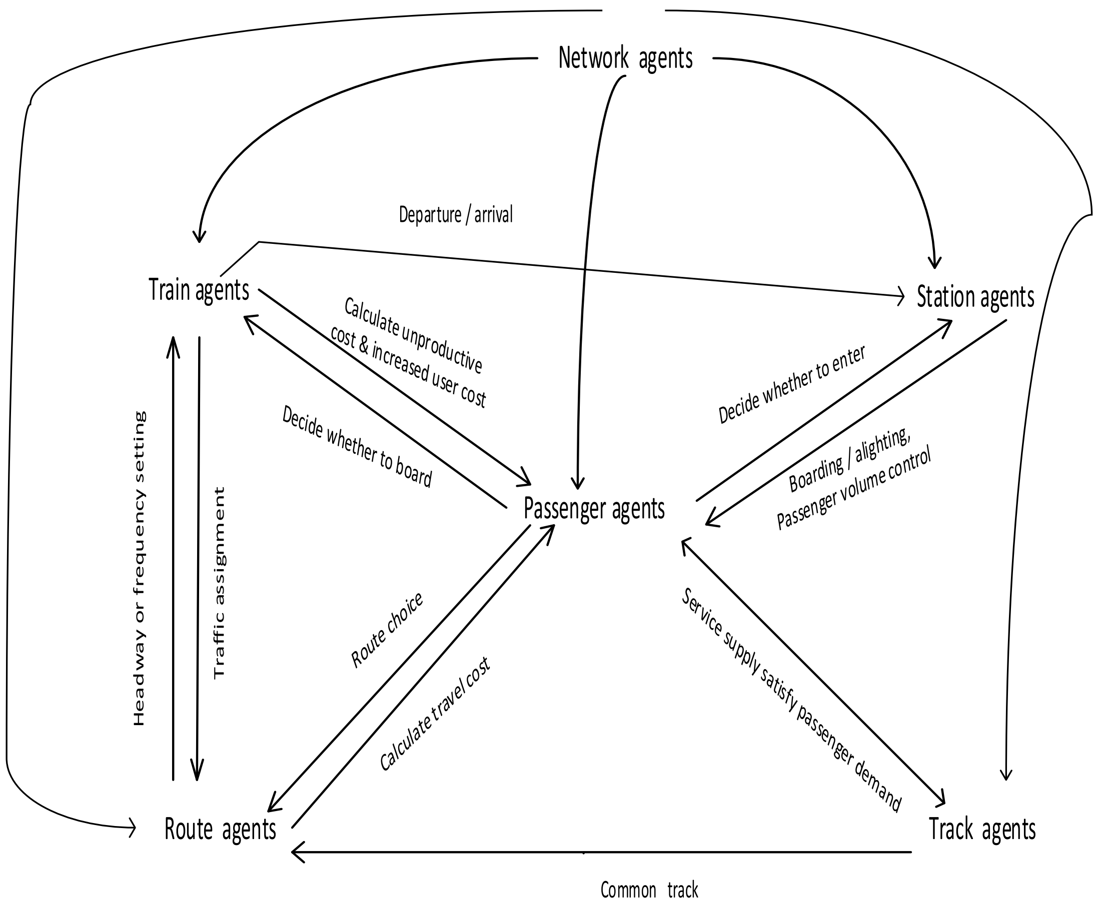

FIGURE 1. System architecture for agent-based matching mechanism.

the URT network for alleviating the congestion on the overloaded segments from the supply side. However, the method proposed cannot be applied to the case that there are no effective alternative routes. Goerigk and Schmidt [18] proposed a bilevel optimization problem with a line planning problem on the upper level and a passenger's route choice problem on the lower level.

The objective of our study is to provide a passenger demand—service supply matching method with the consideration of both the essential interactions between the train flow and passenger flow, and the infrastructure dimension of URT lines or stations. The contributions of this paper are as follows.

1. On the demand side, the dynamic passenger traffic demand with agents is modelled from perspective of passenger boundedly rational travel decision. For agent-based modelling passenger behavior, by considering both the train/vehicle congestion and infrastructure/track network capacity in URT system, the improved flow-dependent calculation method for general travel time is put forward i.e., formula (1) — formula (7), which introduced the additional waiting time (i.e. network congestion degree) on route due to URT network capacity problem (i.e. the trains traffic of URT passenger demand surpasses the infrastructure capacity supply) analogous to Bureau of Public Roads (BPR) function.

2. On the supply side, the dynamic service supply of train traffic with agent is modelled. The headway time is designated as the main decision variable, for the key link between the passenger demand and service supply is the headway time in different time-of-day intervals. Adopting the supply chain principles, the calculation method for operation cost of train agents are put forward, i.e. formula (10) — formula (13).

3. The methods for calculating the intelligent passenger demand — service supply matching degree are put forward, i.e. formula (14) — formula (17). By employing the agent technique, user utility theory (i.e., boundedly rational user equilibria), supply chain principle, and NetLogo simulation comprehensively, a system architecture for agent-based matching mechanism (Fig.1) and negotiations-based iterative/adjusting mechanisms for intelligent passenger demand—service supply matching (Fig. 2) are put forward to generate optimum matching degree, through the multi-agent dynamic interactions and equal connections between the passenger flow and the train flow, and coping with the dynamic law of match between passenger demand and service supply. The proposed methods are validated on the simulation platform NetLogo. The simulation results

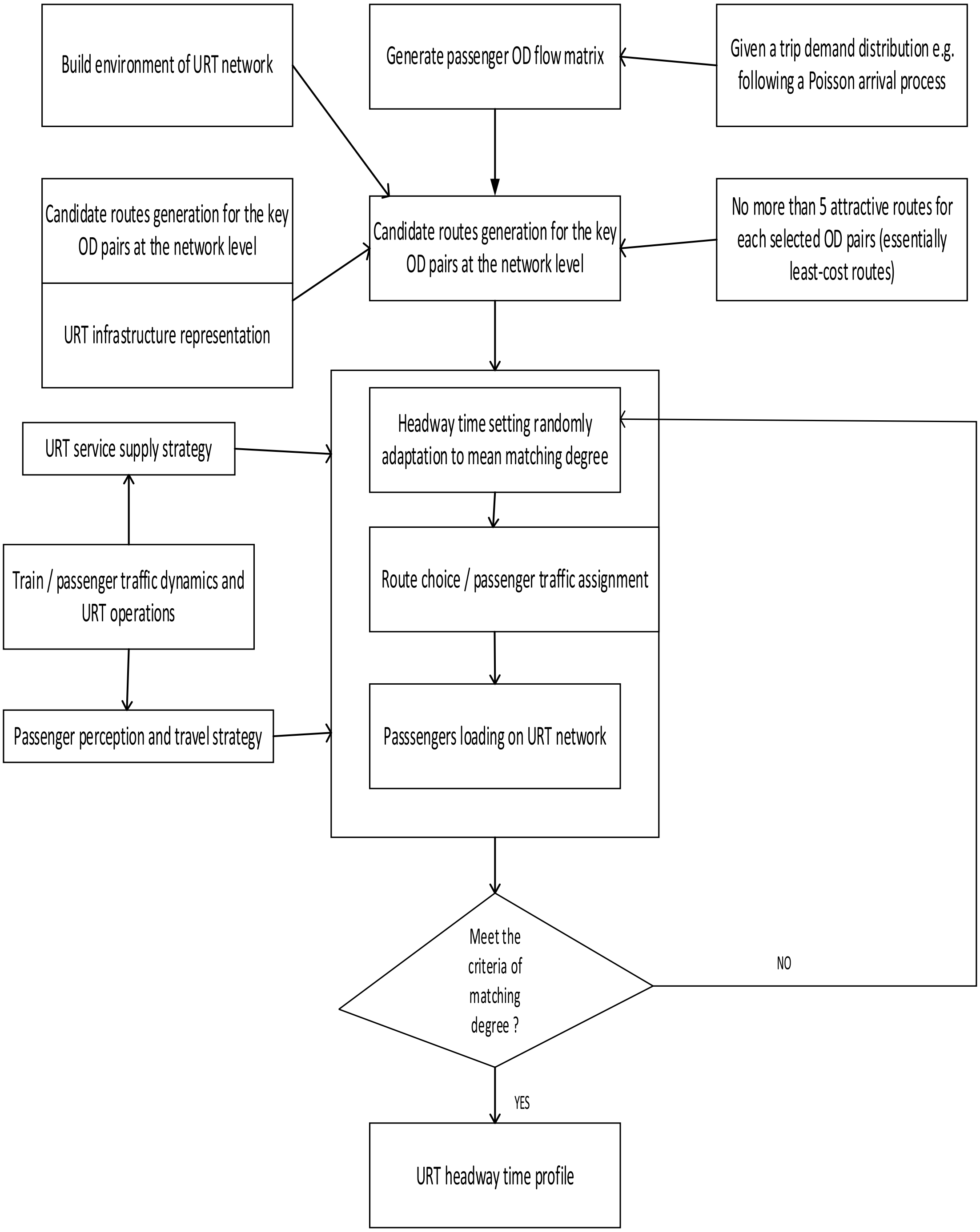

FIGURE 2. Negotiations-based iterative/adjusting mechanisms for matching passenger demand and service supply.

emphasize the importance of representing the supply side and the demand side jointly/interactively. These findings are meaningful for policies on both development of efficient capacity usage strategies of urban rail transit network and provision of high level of service for passengers.

4.The passenger assignment and train operation plan are optimized simultaneously, by modelling the dynamic passenger traffic demand with agent and the dynamic service supply of train traffic with agent, rather than cutting them separately. This study focuses on joint decisions of passenger traffic assignment and train service plan. This work is not only significant but also challenging, for it involves lots of complexities from both URT system attributes and human decision-making processes.

The remainder of this paper is organized as follows. Section II reviews and synthesizes the literature on agent technique, traffic assignment, and URT network. The problem is described in Section III. Both the dynamic passenger traffic demand and the dynamic service supply of train traffic are modelled with agent in Section IV and Section V respectively. Section VI proposes the intelligent matching method between passenger demand and service supply, including calculation of the matching degree, distributed system architecture for agent-based passenger demand and service supply matching mechanism, negotiation-based mechanisms for balancing match between passenger demand and service supply. Section VII demonstrates the numerical example through the NetLogo simulation platform. The last section draws conclusions and discusses some further research topics.

II. LITERATURE REVIEW

A. LITERATURE REVIEW ON AGENT TECHNIQUE

Agent-based method, in particular, allows to model complex systems that involve numerous autonomous and responsive elements, e.g., URT system. Multi agent systems (MAS) are a subfield of distributed artificial intelligence (DAI) technology. MAS aggregates a variety of agents, which are intelligent autonomous entities capable of observing the system environment, communicating with each other and making decisions. By capturing supply uncertainties and adaptive user decisions, Cats [19] presented a simulation framework for a multi-agent transit operations and assignment model to incorporate the interactive process between transit supply and demand, involving the integration of several components, e.g., transit operations, traffic dynamics processes, population generator, traffic and transit assignment models, real-time information processor and adaptive operations planning. BusMezzo is used as the platform for implementation, which is a stochastic event-based simulation programmed in C++ using an object-oriented programming approach.

The methods for supply and demand evolution require the further integration of operations research techniques and behavioral science models. Using an agent-based simulation method, Zhang *et al.* [20] built the artificial urban transport system to describe the passengers decision-making process of traveling route and departure time. To minimize the travel impedance, Narayan *et al.* [21] developed for the first time an integrated multimodal route choice and assignment model that allows users to combine fixed (line and schedule based) and flexible (on-demand service) public transport in a single trip or use them as exclusive modes. Considering the dynamic demand–supply interaction using an iterative learning framework, they implemented the model in an agent-based simulation framework for Amsterdam.

Due to the main sources of complexity that generally preclude finding a unique optimal solution for the transit network planning problem, e.g., non-linearity, the multi-objectives nature (e.g. minimize passenger travel times and company operating costs), and the combinatorial explosion arising from the discrete nature of the problem, generally the past approaches inclined to three categories: optimization formulations of idealized situations, OR heuristic algorithms, and practical guidelines and ad hoc procedures. An AI-based solution approach can incorporate the knowledge and expertise of transit network planners and implement efficient search techniques using AI tools. Because URT systems are generally geographically distributed in dynamically changing environments, it is well suitable for the application of an agent-based modelling (ABM) approach in the domain of traffic and transportation system [22], inspired by a learning-based approach. One key value of ABM lies in its ability to represent human behavior more realistically by incorporating the bounded rationality, agent-agent and agent-environment interactions, heterogeneity, evolutionary learning and adaptation.

B. LITERATURE REVIEW ON TRANSIT ASSIGNMENT AND URT NETWORK

Transit assignment is a process of interactions between individual passengers and transit services, which is commonly applied to estimate passenger ridership and travel times for different line and frequency plans; therefore, it plays a key role in public transport planning [23]. Traffic assignment models can forecast passengers' behavior in response to a potential supply setting, e.g., practically estimate and predict how passengers utilize transit system and choose paths, as well as explicitly model passenger flow distribution. At its core, any transit assignment models include a route choice model that describes the behavior of transit riders regarding their choices of routes to travel between trip origins and destinations. To relax the traditional 'perfect rationality' (PR) route choice paradigm assumption in the static traffic assignment problem, Di *et al.* [24] used boundedly rational user equilibria (BRUE) representing traffic flow distribution patterns, where travelers can take any route whose travel cost is within an 'indifference band' of the shortest path cost, and all the BRUE flow patterns can help predict the variation of the link flow pattern in a traffic network under the boundedly rational behavior assumption. Approaches to transit assignment [25] are broadly divided into schedule and frequency-based assignment. Schedule based models are commonly used for simulation of detailed time-dependent transit assignment. Frequency based models are commonly used for planning purposes, yielding the average distribution of passengers over time and enables the handle of large-scale networks.

Network-based transport modelling is a challenging task. From the perspective of physical infrastructure, URT network is composed of the tracks (links), nodes(stations) and the relationship among the above facilities, which is a static many-origin-to-many-destination network supporting the trains operation. Usually the transit network is developed with a particular shape, e.g., radial, rectangular, grid and triangular. The grid and radial are the most common and basic geometrical shape of the URT network structure, based on which can form the shape of circle + grid and circle + radial by adding the circle line. The formulations of transit networks for dynamic models can be classified into two groups, i.e., frequency/headway based formulation and schedule/timetable based formulation. The frequency-based (also termed headway-based or line-based) approach considers services in terms of runs (lines) with the line headway (or its inverse) and the service frequency, regardless of the explicit run schedule times. This approach could be adaptable for services with high frequency, very low punctuality and not much user/passenger information, e.g., the public transport or the urban rail transit in early times without advanced intelligent transport techniques. In contrast, the schedule-based (also termed timetable-based or run-based) approach (also called dynamic approach) focuses on services in terms of runs explicitly considering vehicle arrival/departure time, which allows us to take into account the evolution in time both of supply and demand, as well as run loads and level of service attributes [26]. The schedule-based models are often adopted for low frequencies. A static frequency-based transit network is represented within space domain only (i.e. without time dimension), which does not reveal the peaked nature of capacity (i.e. crowdedness) problem. While the central idea of this approach for dealing with the capacity constraints lies in the introduction of a ''fail-to-board'' probability as passengers cannot board the first service arriving due to overcrowding [27]. Studies on transit users' route choice in the context of transit assignment can be categorised into three groups: static transit assignment; within-day dynamic transit assignment; and emerging approaches.

The physics-based statistic features of transport network include node-centrality property, small-world property, scale-free property, distribution of weight of intensity, network community structure, and static/dynamic robustness. In the frequency—based models, usually the topological directed graph (e.g., L-Space or P-Space) or the connection matrix is adopted to construct the physical network of URT system. The node degree distributions of the URT network obey the power law distribution, which has typical characteristics of small-world properties and scale-free degree distributions. Because of the difficulties in collecting traffic data, most of the previous studies have focused on the physical topology of subway systems, whereas few of them have considered the characteristics of traffic flows through the network. In view of the schedule-based model, representation of the transit network is a run-based spatio-temporal graph that can show individually serial runs as scheduled in timetables. Moreover, every move of the entities (i.e. transit vehicles and passengers) is marked with a timestamp, such that those entities can be located, described, and differentiated from each other in both temporal and spatial dimension, e.g., the stochastic time-dependent network, also called time–variation network, i.e., the travel times on the arcs are functions of time, and the travel time is represented by probability distributions rather than simple scalars. A time-expanded network is built on a two-dimension graph with one time axis and one space axis. While a diachronic network is built in a three-dimension graph with two space axes and one time axis. To simulate high-frequency services (e.g. URT), frequency-based (FB) models are usually preferred, because of the facts that most of these passengers do not choose a particular run and do not time their arrival at the stop to coincide with train arrivals [28]. Schmöcker *et al.* [27] proposed the first frequency-based dynamic transit assignment model for overcrowded high-frequency transit networks, where passengers might not be able to board vehicle and hence remain on platform. On the other hand. Some early studies focused on the transit network structure for system design in simplified radial, however, since the 1980s most approaches were either applied to realistic, irregular grid networks or the network structure was of no importance for the proposed model and therefore not specified at all [29].

In URT network, passenger load on the nodes/links is usually viewed as stochastic time-dependent, i.e. the passenger volume depends on the arrival rate at a node. Nuzzolo *et al.* [26] presented a schedule-based path choice model for high-frequency transit networks, considering the evolution in time of transit services, both within-day and day-to-day, as well as the day-to-day learning process of attributes by which users choose. However, in reality, the actual arrival/departure time of a run may not be consistent with the schedule but vary from time to time, especially for the high-frequency transit system. It is then considered as random with the published scheduling time being the mean.

The method presented in this paper is a dynamic frequency-based or semi schedule-based, which considers the dynamic effects by determination of the effective frequency/headway from both the demand side and supply side. The physical features of URT network include accessibility, extensibility, reliability, convenience, and advancement. The advancement is represented as the digital construction, informalization/automatic operation and maintenance, intelligent service towards passengers in the dimension of business operations. While in the perspective of technique equipment (e.g., the sensors and video monitors, the communication-based train control system) and the management level (e.g., Automatic vehicle location (AVL), automatic passenger counting (APC)), a typical feature of the URT system lies in the modern cyber-physics system, i.e., it is a kind of informed system, which lays the foundation for the intelligent operation. For simplification, in this study, an intermediate level of network representation for URT routes is used, i.e., section of routes where no train routes start, end or divert are represented as one link. In nature, the URT network belongs to the kind of time-spatial-state dependent cyber-physics system,with information systems offered to users at stops, e.g., tracking technology that provides accurate data on travel times, stopping times, and passenger counts, by which the passengers/operators can discern the operation status (e.g., the volume of on-board passengers). The optimization methods for network can be defined [30] within two classes, i.e., structural optimization and flow related optimization. Nevertheless, these two classes always mix together.

III. PROBLEM DESCRIPTION

According to the procedures of networked operation, it can be divided into the primary phase and the large passenger flow phase. In the primary phase, the total passenger volume is not very large, and the passenger intensity is not very high, with less transfer, but keeps increasing trend. In the large passenger flow phase, both the total passenger volume and the passenger intensity are very high, with more transfers. Fully considering the time-variation–dynamic interactions among the passenger demand, URT network infrastructure, and the passenger/train flow, we mainly target at the second phase to intelligently match the service supply with the passenger demand at the network level.

This study aims at the co-evolution process iteratively between passenger flow assignment (route choice) and train operation plan (transit service network design), within certain period over the URT network with information systems to users at stops, by considering both user equilibrium (boundedly rational user equilibria) and system optimal simultaneously. In details, the research goal is to optimize the operation of train services through adjustment of the headway times in accordance with the dynamic passenger demand, so as to achieve the maximum total satisfied passenger demand, to use the URT capacity efficiently, and to tune the travel reliability perception of passengers to reality. The nature of intelligent demand-supply matching lies in the dynamic synergy and coordination. The operation plan strategy includes: (i) maximize train load of passenger flow at rush hour, (ii) minimize waste of service capacity at off-peak hour, (iii) meet the system requirements of the timing, economy and capacity constraints as well as comfort.

The passenger route choice (traffic assignment) problem determines the flow pattern of the URT network. The non-trivial process that maps strategic passenger traffic flows to priority-complient train traffic flows is introduced. Here what the term 'priority-complient' [31] means is that in the context of URT vehicles with rigid capacities, the passengers sort the train lines in decreasing order of preference at each boarding node, and they board the first vehicle in the preference list whose residual capacity is nonzero. Few works focus on passenger demand together with train service supply along with passenger control strategy and train capacity. To formulate the model for intelligent passenger demand— service supply matching and develop an effective mechanism/framework, the following assumptions are made throughout this study.

Assumption 1: The passengers are frequent users, i.e., they know from previous experience how the URT system operates. The time-slice based transit demand OD matrix is generated in the many-origin-to-many-destination capacitated URT networks. It is assumed that passengers arrive at stations randomly according to a Poisson distribution. The passenger OD demand matrix is assumed symmetric, and only the unidirectional track line (e.g., upstream or downstream) is considered in the URT network.

Assumption 2: Passengers are boundedly rational in their decision-making process, i.e., they would only switch routes when the improved general travel cost exceeded some indifference bands. Passengers travel along the routes of the underlying network, where path cannot be modified en route (path and route are used interchangeably in this paper) once they get on the train.

Assumption 3: The formation of train fleet is homogeneous and fixed. The available fleet size is sufficient. The full-length train service is adopted, instead of the different types of lines such as short-turning, dead-heading and limited-stop lines. No stop skipping is allowed. The routes and frequency of the train lines are predetermined and adjusted accordingly and iteratively. The train running time on the rail track is deterministic. The operation patterns of the trains adopt the same headway time in the same network, i.e., the headway times among all of the trains within certain study period in the same URT network keep invariant.

Assumption 4: Following the system optimal perspective, it is assumed that agents are aware of other agents' decisions and collaborate to obtain the intelligent passenger demand — service supply matching. The learning and decision-making processes of agents are assumed to follow reinforcement learning principles for experience updating and choice techniques.

Assumption 5: Routes are created to serve desired destinations in the shortest possible time, meanwhile meeting maximum demand and minimizing passengers travel cost & train operation cost. The passenger's choice of an attractive line is made at a transit stop (station), whereby the passenger flows split among the attractive line.

What Assumption 3 means is that once the passengers get on the train, they could not change their travel path en route, which is determined by the feature of urban rail transit. This assumption is set from perspective of passengers. While Assumption 4 is set from the perspective of urban rail transit operators. Both the passenger agents and the train agents have their individual behaviors. Exactly, the seeming conflicts between Assumption 3 and Assumption 4 are the expressions of autonomous agent behavior, and they have been settled by the technique of distributed multi-agent system during simulation with NetLogo.

IV. MODELING DYNAMIC PASSENGER TRAFFIC DEMAND WITH AGENTS

A. ANALYSIS ON SPATIAL-TEMPORAL FEATURES OF PASSENER FLOW IN URT NETWORK

1) SPATIAL AND TEMPORAL PATTERN OF PASSENGER DISTRIBUTION

The spatial features of the passenger distribution can be seen as the function of urban structures and usage of urban land, which can be described from the spatial information (e.g. points of interests) to a certain extent. The passenger flow patterns are spatially heterogeneous due to the poly centricity of urban structures.

TABLE 1. Temporal pattern of URT passenger flow distribution and detailed characteristics.

Temporal pattern of URT passenger flow distribution	Detailed characteristics
Large passenger flow under passenger volume control during morning peak hours	High concentration of population, fully congested, most of the trip purposes are home-based work (HBW))
Evening peak hours under passenger volume control	High concentration of population, less congested, most of the trip purposes are non-home based (NHB).
Normal passenger flow without passenger volume control	Relatively low and stable concentration of population, uncongested, most of the trip purposes are home or non-home based other (NHB).
Pre-peak/after-peak period with extremely low passenger volume beyond the above three scenarios	Pre-peak morning period for the early bird trip-maker, after-peak night period for the late bird trip-maker

With more public transport agencies facing crowding problems, there is an increasing need to develop more structured conceptual and methodological approaches for public transport travel demand management, e.g. URT. Public travel demand is time-dependent and varies according to time of the day, day of the week, as well as time of the year. According to the number of passenger departures within a period and the temporal information (e.g. duration, time of day), the temporal pattern of URT passenger distribution can be segmented into four state scenarios within a day as shown in Table 1. Thus, the daily operation time of URT can be divided into such several large time intervals as early morning, am peak, midday, pm peak, evening and late evening. In uncongested passenger conditions at the off-peak hours it is unnecessary to consider transit capacity and passenger volume control. Otherwise, in the congested passenger conditions the transit capacity restrictions cannot be neglected at the peak hours.

2) PROPERTY OF SELF-ORGANIZATION

Public transport users tend to have less varied trip purposes than drivers, which are constrained by service schedules. The main technique feature of URT system is its dedicated and exclusive rail-based infrastructures, while its service form belongs to urban public passenger transport, in nature which is the subfield of public transportation. The trip of URT passenger possesses a property of self-organization and stochastics. However, its operation mode approximates to the national railroad, which means that the operators can make the train operation plan according to the principle of system optimum, so as to predecide the passenger route choice alternatives to a greater extent, i.e., URT passengers are constrained by service schedules. On the other hand, according to the various characteristics of traffic flow, it can be divided into the kinds of weak-controllable-autonomous traffic and strong-controllable-organized traffic. The technique feature of the URT passenger flow belongs to something in between.

3) GENERATION OF OD MATRIX

URT is a closed system, requiring users to tap their cards on entry and exit, so complete trip records are available. Passenger demand is represented as an Origin-Destination matrix. There are two kinds of methods for passenger OD reconstruction [32], one is the trip-based OD reconstruction by CDR data (Call Detail Records from mobile phone), the other is stay location-based OD reconstruction, including temporal-based clustering, distance-based clustering, frequency-based clustering, and time-distance clustering. There is considerable evidence that passenger arrivals appear to be Poisson for higher-frequency (lower-headway) service, with headways up to 10-15 minutes [25]. Thus, in the URT case, it is common to assume the Poisson passenger arrivals for generation of OD demand matrix.

B. AGENT-BASED MODELING PASSENGER BEHAVIOR IN URT NETWORK

1) ROUTE CHOICE BEHAVIOR

A complete journey in URT can be segmented into such successive parts as access to a stop (and buy a ticket if necessary), walking to a platform, waiting to board, travelling on-board, alighting (and/or interchange to another line if necessary) and egress. The behavioral assumption has a decisive impact on the problem formulation, expected waiting times and estimated distribution of passengers in the network [23]. There may be more than one route between the OD pairs in the transit network, which means users normally face common lines to reach their destination. The passenger agents are treated as the entities with different attributes and behavior, allowing passengers to choose their own boarding rules as they travel from origin to destination. Each passenger agent plans his transit trip by selecting one transit option. The main behavior of the passenger agent is the route choice with boundedly rational travel decision, so as to achieve the boundedly rational user equilibria [33]. The passenger route choice (traffic assignment) problem determines the flow pattern of the URT network. An O–D route is defined strictly as a sequence of track segments (or links) in the URT network. Each candidate route is a travel strategy. The route selection criteria for the passenger agents include (i) minimize general cost with boundedly rational travel decision according to following formula (1); (ii) the maximal number of transfer times is no more than 2. In the case of one or two transfer path, a passenger agent may choose the URT routes twice or three times: at the origin node and at the transfer node. Or else, the route choice is a pretrip behavior, i.e., the passengers maybe determine the whole route only once before they depart from the origin node. In this study the latter one is considered. Given the URT network, for each OD pair, the set of effective URT routes which can be used by the passengers is determined by calculating the general travel cost for each of the transfer path. In the public transit system, the common line problem deals with the choice of a line among a set of lines which share either all or part of the path of an O-D demand pair. A route section is the link between two consecutive transfer nodes [20]. Here the common tracks are defined in the URT network as follows: they are the tracks that more than one passenger paths share during the passenger travel.

The routes that the demand of each OD pair takes in the URT network compose of the path set. The path may be formed by either a segment of a single route (direct travel) or more than one single route (i.e. transfer is necessary.). Not all of the available lines at a platform would be considered by a wait-to-board passenger, but only a portion of (usually no more than 5 of) all lines that are available would be perceived as rational. Predicting the choice behavior within a passenger/train traffic assignment analysis is an essential part in determining the frequency/headway. The route choice in URT network is typically pre-trip, while particular train line choices are adaptive. Usually the passengers waiting on the platform board a vehicle on the First-Come-First-Serve (FCFS) basis. The mostly accepted assumption for boarding is that, in uncongested situations, all passengers would board the firstly arriving line out of attractive ones. Under crowding situations, passengers may fail to board, but keep waiting on the platform instead. Generally, the fail-to-board passengers can be classified into two groups of those (i) who fail to board due to limited standing space; and (ii) who are not willing to board because of unavailability of seats. For assumptions of the process that a passenger determines individually his/her perceived route choice set, we distinguish those wait-to-board, intend-to-board, decline-to-board and seat-sensitive passengers, to accommodate the fact of first refusal concerning, e.g., unwilling-to-stand and unwilling-to-sit cases. For the unsatisfied passenger demand, it is generally modeled by introducing a penalization weight in the objective function [15]. Given the limited size and its simple structure of a URT network, a brute-force-search algorithm could be more advantageous than other methods (e.g., link elimination, labelling and k-shortest-path) in generating route sets in shorter time [34]. In the literature there are multiple route choice models (mainly Multi Normal Logit (MNL) models or extensions of it) based on users' socioeconomic characteristics and route attributes. This study presents a more complete route choice analysis on URT networks, incorporating factors related to the different times involved (travel, waiting and walking times), trains and stations usage, transfer environment, level of service and the travelers' perceptions. One of the limitations of the MNL model is that it does not consider correlation between alternative routes due to overlapping. To address this issue, C-logit model is recommended [35].

2) CALCULATION OF PASSENGER TRAVEL COST

In most urban transit systems the trip fare is constant for any given OD pair, no matter which travel path is chosen. In other words, the trip fare cannot bias the results of the path choice process under most circumstances. Thus, the monetary cost can be neglected for route choice in URT network. The heterogeneity of passenger preference is taken into account. The calculation for general travel cost of route choice by passenger agent p is as formula (1), which is a flow-dependent cost function.

$$C_r^p \leq \pi(r) + \varepsilon(p) \tag{1}$$

where p denotes index of passenger agent p, $p \in P$, defined based on each OD pair. C_r^p denotes the travel cost of route r choice of passenger p, which departs from OD pair w, $w \in W$. W denotes set of URT network origin-destination (OD) pairs. w denotes an element of W. R_w denotes set of feasible routes associated with OD pair w, and r denotes an element of R_w, $r \in R_w$. $\pi(r)$ denotes the minimal route travel cost of OD pair w of passenger p on route r, which is calculated by formula (2). $\varepsilon(p)$ denotes the indifference band or tolerance value of passenger agent p.

Train/vehicle congestion and infrastructure/track network capacity in public transport assignment are not the same problems [36]. Considering both of them, the flow-dependent formula for calculation of minimum route travel time $\pi(r)$ for OD pair w of passenger agent p on route r is set as formula (2).

$$\pi(r) = \min_{r \in R_w}(\sum_{i=1}^{N} t_{i-1,i}^r \cdot (1 + Y_{i-1,i}^r) + \sum_{k=1}^{K} t_k^r + \frac{h_r}{2} + \sum_{m=1}^{M} (\alpha \cdot t_m^r) + \beta \cdot \varphi_r(v)) \tag{2}$$

The first term on the right side of formula (2) denotes the in-vehicle travel time on route r which departs from OD pair w considering on-board congestion; the second term denotes the dwell time for boarding/alighting on route r; the third term denotes the normal waiting time on route r; the fourth term denotes the walking time at transfer stations on route r; the fifth term denotes the additional waiting time (i.e., network congestion degree) on route r due to the URT network capacity problem (i.e. the trains traffic of URT passenger demand surpasses the infrastructure capacity supply), for which the passenger volume control strategy is necessary to be enforced.

In formula (2) Where $t_{i-1,i}^r$ denotes the in-vehicle running time from station i-1 to i on route r which departs from OD pair w. $Y_{i-1,i}^r$ denotes the in-vehicle congestion degree or discomfort, which calculates the additional time cost coefficient induced by congestion within the in-vehicle running time. t_k^r denotes the dwell time at stop station k. h_r denotes the planned dispatching headway time of consecutive trains on route r over the study period, i.e., the key decision variable or order parameter. M denotes the number of transfers on route r of OD pair w. t_m^r denotes the average walking time of passenger agent at transfer station m on route r. α ($\alpha > 1$) and β (0 or 1) are the parameters. $\varphi_r(v)$ denotes the additional waiting time due to URT network congestion, which also can be regarded as the network congestion degree. According to Bureau of Public Roads (BPR) function and De Cea [37], the calculation formula for $\varphi_r(v)$ can be defined as formula (3).

$$\varphi_r(v) = \frac{v^r + \widetilde{v}_r}{K_r} \tag{3}$$

where v^r denotes the URT passenger flow that can be assigned to route r; $\widetilde{v}_r$ represents the flows that compete with v^r for the same common capacity on the common track; K_r denotes the ideal capacity allocated to route r, and $K_r = \frac{\lambda T C_{apa}}{h_r}$, where T is the study period, C_{apa} is the maximal loading capacity of single train, usually $C_{apa} = 1400$ or 1000, and $0 < \lambda < 1$.

The calculation formula for $Y_{i-1,i}^r$ is defined as formula (4).

$$Y_{i-1,i}^r = \begin{cases} \dfrac{D_{i-1,i}^r - \frac{Z \cdot T}{h_r}}{\frac{Z \cdot T}{h_r}} A & \text{if } \dfrac{Z \cdot T}{h_r} < D_{i-1,i}^r < \dfrac{T \cdot C_{apa}}{h_r} \\ \dfrac{C_{apa} - Z}{Z} A + \dfrac{D_{i-1,i}^r - \frac{T \cdot C_{apa}}{h_r}}{\frac{T \cdot C_{apa}}{h_r}} B & \text{if } D_{i-1,i}^r > \dfrac{T \cdot C_{apa}}{h_r} \end{cases} \tag{4}$$

where $D_{i-1,i}^r$ denotes the onboard passenger volume between station i-1 and i (i.e. the track edge connecting the station i-1 and i) on route r, which is the context discernment of passenger agents in the CPS environment of URT network. Z the number of train seats. A denotes the cost for over seat capacity of the train. B the additional cost for over congestion.

According to Mahmassani [38], the formula for the term $\varepsilon(p)$ in formula (1) is defined as follows.

$$\varepsilon(p) = \max(\eta_p, \tau_p) \tag{5}$$

where η_p is a relative indifference threshold for passenger p. Results of laboratory experiments indicated that η_p is about 0.2 for typical urban commuters [39]. The quantity η_p governs passenger agents' response to the supplied information and their propensity to switch. τ_p is an absolute minimal travel time improvement needed for switching. Results of laboratory experiments indicate that τ_p is on average equal to one minute [39]. We set τ_p as a float random value in-between (0,1) in the URT network system.

For each OD pair $w \in W$ on the capacity constrained URT network, most of the trip demand q^w would be split into all effective routes as formula (6).

$$q^w \geq \sum_{r \in R} v_r^w \tag{6}$$

where v_r^w is the passenger flows assigned to the route r of OD pair w.

On the passenger demand side, the total travel cost C_{demand}^r incurred by the passenger demand on route agent r is as formula (7).

$$C_{demand}^r = C_r^p \cdot \sum_{w \in W} v_r^w \tag{7}$$

V. MODELING DYNAMIC SERVICE SUPPLY OF TRAIN TRAFFIC WITH AGENTS

A. DEFINITION AND ANALYSIS OF SERVICE SUPPLY OF TRAIN TRAFFIC

A run-based representation of services is employed in this study, which means the passenger choice behavior pertains not only to the route, but also to the specific train run of each line on the route. A URT network consists of a set of lines and stops/stations where passengers board and alight. A train service line/route is defined by its origin and destination terminals, as well as the sequence of stops that it serves in between, i.e., it corresponds to one scheduled service of the route. Each route has three attributes: travel time, frequency, and capacity. Its capacity is the frequency of that route line multiplied by the train capacity. The inverse of the frequency over a determined period is called the headway time, which corresponds to the time elapsing between consecutive line run departures or arrivals. The line frequencies for each time period (i.e., hour of the day) from the supply side should match the travel demand at best so as to avoid overcrowding and excessively large headways, and thereby reduce waiting time [40].

Usually the route options on the supply side are hyperpaths – a hyperpath being an oriented, acyclic sub-graph connecting the origin and destination nodes. For each route, the operational characteristics, e.g. frequencies or headways are typically determined on the supply side, through the calculations based on expected passenger volumes or by applying transit assignment techniques, considering the desired load factors, fleet size. The current advent of the research in Mobility-as-a-Service (MaaS) and big data analytics provides a rich opportunity to consider stochasticity in the supply side decision making [41], e.g., the statistic travel time. However, considering the high reliability of URT operation, the travel time of URT can be regarded as deterministic, especially at the off-peak hours. For the situations at peak hours, we introduce the network congestion degree $\varphi_r(v)$ as used in formula (3) for calculation the additional time cost due to train traffic congestion.

In this study, for service supply in the URT network system, what is mainly concerned is the service headway time between consecutive trains, which is taken as the key decision variable or order parameter. In general, it is believed that an even schedule with a constant headway between consecutive vehicles can reduce passenger total waiting time. The typical operation modes of URT network include: (i) Each line operates independently. This mode is adopted most commonly in

China, which is relatively simple and with less interferences among the lines, but not beneficial to the improvement of the service level. (ii) All lines adopt the same headway time within the whole URT network. This means that the trains run on all of the lines with the identical headway time in certain period, e.g., the operation mode of New York metro system. (iii) Periodic operation mode in the URT network. This is unnecessary to adopt the even headway time in the whole day, but on certain period (e.g., the off-peak period), it is helpful to build the regular timetable. From the perspective of statistics [42], the operational reliability can be assessed by the variation coefficient, i.e., standard deviation divides the mean value. Accordingly, in this study, the second operation mode is adopted, i.e., all lines adopt the same headway time within the entire URT network, in order to ensure the service reliability. In details, firstly the representative key OD pairs are selected, which are usually the ones with the longest travel distance, the most intermediate nodes and transfer possibilities. And then the alternative route sets can be determined. Thirdly, the headway time within certain value domains can be generated randomly and adjusted accordingly/iteratively in the URT network until the desired headway time is found. The concrete negotiation-based iterative mechanisms can be described in Section VI.

B. TRAIN TRAFFIC DYNAMICS FOR SETTING HEADWAY TIME

Frequency designates the number of trains that could run over a route between an OD pair, during a specific time interval. For simplicity, the model uses headways instead of frequencies, thus for a frequency f_r (vehicles per time range of T minutes), the equivalent headway h_r (minutes between consecutive departures) is $h_r = \frac{T}{f_r}$ and $h_r \in [h_{\min}, h_{\max}]$. $h_{\min}$ denotes the minimum headway time between consecutive trains, $h_{\max}$ denotes the maximum headway time between consecutive trains.

Since the URT system has the high level of right-of-way that can guarantee higher reliability and the link running times in URT network are constant to a large extent, it is assumed that the planned dispatching headway of train trips on a route r, i.e., the headway time between successive trips at the departure stop, h_r, will be maintained at any other stop of route r. On the other hand, the uniform (constant) headways not only can provide the most efficient operation (even for vehicle loading and schedule stability), but also are most attractive to passengers [43].

The train traffic dynamics for calculation of headway time of consecutive trains in URT network is illustrated as formula (8).

$$h_r = \begin{cases} t_{\ln ext} - t_{\text{lcritical}} & \text{if the boarding station is the transfer station} \\ t_{l+1} - t_l & \text{if the boarding station is the non-transfer station} \end{cases} \tag{8}$$

where $t_{\ln ext}$ denotes the arrival time of train trip Next, $t_{\text{lcritical}}$ denotes the arrival time of train trip Critical, t_{l+1} is the arrival time of the following train service $l + 1$ on route r for OD pair w, t_l is the arrival time of the proceeding train service l on route r for OD pair w.

At the transfer station, there is a transfer connection going from a train trip Feeder running on a feeder line l_{feeder} to a Receiving line $l_{receive}$ on route r of OD pair w. Let Critical be the train trip on $l_{receive}$ whose schedule time of arrival at the transfer point with $l_{critical}$ is closest to that of Feeder, and let Next be the train trip following Critical on $l_{receive}$. On the common track, the total service supply should sufficiently satisfy the associated passenger demand, which is also the decision mechanism of the track agent and expressed as formula (9).

$$\sum_{r \in R} D^r_{i-1,i} \leq \sum_{e_{i-1,i} \in r} \frac{C \cdot T}{h_r} \tag{9}$$

where R denotes the set of routes in URT network, $e_{i-1,i}$ denotes the common track connecting station i-1 and i, $D^r_{i-1,i}$ denotes the passenger volume between station i-1 and i (i.e. the track edge connecting station i-1 and i) on route r.

C. CALCULATION OF OPERATION COST FOR TRAIN AGENTS

Adopting the supply chain principles [44], Hadas and Shnaiderman [45] proposed the optimal cost-based frequency setting model to minimize the total cost incurred with decision variables of either frequency or vehicle capacity, considering two main cost elements, i.e., (i) empty-seat driven (unproductive cost) and (ii) overload and un-served demand (increased user cost). Here the decision mechanism of the train agent on the supply side is: (i) to minimize the unproductive cost (e.g., to maximize the minimum load factor at off-peak hour); (ii) to minimize the increased user cost (e.g., to minimize the maximum load factor at peak hour, to minimize the number of fail-to-board passengers at the first time of boarding). The analysis approach is analogous to Kogan and Shnaiderman [44], and the method proposed by Hadas and Shnaiderman [45] is adopted in this study. Let c^+ be the empty seat average cost per time unit, let c^- be the un-served passenger shortage cost per time unit, and let $t^l_{i-1,i}$ the running time between stop i-1 and stop i for train agent l.

If the real train load is smaller than the supply capacity, then overage cost of train agent l at station i is equal to formula (10).

$$c^+_{li} = t^l_{i-1,i} \cdot c^+ \cdot \max(C_{apa} - d^l_{i-1,i}, 0) \tag{10}$$

where $d^l_{i-1,i}$ is the load of train agent l between station i-1 to station i.

On the other hand, if the load is higher than the capacity, then the shortage cost of train agent l at station i is equal to formula (11).

$$c^-_{li} = t^l_{i-1,i} \cdot c^- \cdot \max(d^l_{i-1,i} - C_{apa}, 0) \tag{11}$$

The total cost of train agent l at station i is therefore the sum of c_{li}^{+} and c_{li}^{-} as formula (12).

$$c_{i-1,i}^{l} = t_{i-1,i}^{l} \cdot c^{+} \cdot \max(C_{apa} - d_{i-1,i}^{l}, 0) + t_{i-1,i}^{l} \cdot c^{-} \cdot \max(d_{i-1,i}^{l} - C_{apa}, 0) \quad (12)$$

The objective function for all trains of route agent at all stops on route r is as formula (13).

$$C_{supply}^{r} = \sum_{l=1}^{L} \sum_{i=1}^{N} c_{i-1,i}^{l} \quad (13)$$

where L denotes the number of trains running on route r for OD pair w, and $L = \frac{\lambda T}{h_r}$.

There are four kinds of different models for route set generation for road network, i.e., (i) all acyclic routes, (ii)k-shortest routes, (iii) essentially least-cost routes and (iv) most probable routes. For URT network in this research, the third model (essentially least-cost routes) is modified for candidate routes generation as follows, i.e., no more than 5 attractive routes with a general travel cost within a certain threshold from the least cost route. After the set of attractive paths is found, the probability for their selection is computed.

VI. INTELLIGENT MATCHING DEGREE BETWEEN PASSENGER DEMAND AND SERVICE SUPPLY

A. CALCULATIOIN OF MATCHING DEGREE

At the strategic level, frequency setting interacts with passenger route choice, leading to the spatial-temporal coupling characteristics between them in URT network system. A trip would fail to board if there is insufficient capacity on the line. In most cases passenger is sensitive to the quality of the transit service. Service supply and travel demand are changing over time and interplay by the force of passenger choice behaviors. Service drives demands. The passenger load depends on the service schedule and varies along the route. Changes in the supply impact the demand, and the vice versa. The match between the passenger demand and the service supply can be classified in terms of total volume, distribution structure, service quality, and the service supply adapting to the passenger demand. The dynamics of the interaction between traveler decisions and the URT supply are an endogenous source for deteriorating the transit system performance [35], e.g. the service reliability.

In accordance with the passenger load dynamics analyzed by [46], the formula for calculating the volume of passengers boarding train service l at station k on route r of OD pair w in URT network can be deduced as follows.

$$qe_{l}^{r}(k) = \min(qf_{l-1}^{r}(k-1) + h_r \cdot b_{l}^{r}(k) - o_{l}^{r}(k), C_{apa} - \rho_l(k) \cdot q_{l}^{r}(k-1)) \quad (14)$$

where $qf_{l-1}^{r}(k-1)$ denotes the volume of passengers that fail to board the proceeding train service $l-1$ at station $k-1$ due to the capacity constraints of the train service; $b_{l}^{r}(k)$ denotes the arrival rate of passengers that enter station k, after the proceeding train leaves and before the following train arrivals; $o_{l}^{r}(k)$ denotes the volume of passengers that are restricted outside of station k, so as to control the number of passengers that enter station k for train service l; $\rho_l(k)$ denotes the proportion of passengers alighting when train service l arrives at station k; $q_{l}^{r}(k-1)$ denotes the number of passengers on board of train service l when it leaves from the proceeding station $k-1$ on route r of OD pair w.

And the total volume of passengers $\sup_k$ that can board the train services at station k within the planning horizon T can be calculated as formula (15).

$$\sup_k = \sum_{r\in R_w} \sum_{l=1}^{T/h_r} qe_{l}^{r}(k) = \sum_{r\in R_w} \sum_{l=1}^{T/h_r} \min(qf_{l-1}^{r}(k-1) + h_r \cdot b_{l}^{r}(k) - o_{l}^{r}(k), C_{apa} - \rho_l(k) \cdot q_{l}^{r}(k-1)) \quad (15)$$

Thus, the matching degree mat_k between passenger demand and service supply at station k can be calculated as formula (16).

$$mat_k = \frac{\sup_k}{dem_k} \quad (16)$$

where dem_k denotes the total volume of travel demand at station k. While the mean matching degree mat_{mean} between passenger demand and service supply in the whole URT network can be calculated as formula (17).

$$mat_{\text{mean}} = \frac{1}{\mathrm{K}} \sum_{\mathrm{k}=1}^{\mathrm{K}} mat_k \quad (17)$$

The implications of the mean matching degree calculated from formula (17) can be interpreted as Table 2. When the value of the mean matching degree falls in between 0.86∼0.90, the service supply can best meet the passenger demand. In this value domain 0.86∼0.90, the operation strategies of the URT network can both provide the sound transport service, and save the supply cost economically. Outside of the above value areas, it indicates that either the transport service cannot satisfy the travel demand effectively, or the service supply surpasses the passenger demand wastefully.

B. DISTRIBUTED SYSTEM ARCHITECTURE FOR AGENT-BASED MATCHING MECHANISM

Six types of heterogeneous distributed agents are built in this study, i.e., passenger agent, train agent, route agent, track agent, station agent, network agent. The environment is the URT landscape on which agents interact and can be geometric, network-based, or drawn from operational data. Following the system optimal perspective, it is assumed that agents are aware of other agents' decisions and collaborate to obtain the intelligent match between passenger demand and service supply, e.g., passenger agents have all the information they need to make decisions, e.g., they have the mental map of the URT network. Each passenger agent plans his transit trip by selecting one URT transit option, according to each passenger type, exploration rate. Moreover, from the perspective of the agent cognition, the passenger agent and train

TABLE 2. Implications of the mean matching degree.

Value domains of mean matching degree	0.86~0.90	0.91~0.95 or 0.7~0.85	0.96~1.00 or 0.54~0.69	1.01~1.10 or 0.38~0.53	Greater than 1.10 or 0~0.37
implications of mean matching degree	best	better	average	bad	worse

agent can be abscribed to the kind of utility-based adaptive agent [47], which can make different decisions if given the same set of inputs by modifying their actions or strategies, based on the utility or cost. Network agent supervises the conditions in the entire network. It builds the center piece of the agent system that holds all the single parts together and ensures that complete simulations/optimizations with multiple iterations can be run. To a large extent, the network agent behaves as the Observer in NetLogo [47], [48]. It also means that the network agent contributes most of the monitoring and supportive decisions (which can be replaced by human), while the other agents take most of the action and operational responsibilities. The route agent provides the effective routes as alternatives for OD pairs. The route choice mechanism for passenger agent observes formula (1). In the station agent, the passenger's choice of an attractive route is made at a transit stop (station), whereby the passenger flows split among the attractive lines. Each agent is endowed with the interest in making decisions so that the URT system performs well. No doubt any agent in the architecture system is considered as an intelligent agent, which means they have reactive, proactive and interactive properties. In these regards, the learning activities of agents include updating their mechanism for making predictions about network conditions and the strategies for making choices and decisions. The distributed multi-agent system architecture is developed as Fig. 1.

C. NEGOTIATION-BASED ITERATIVE/ADJUSTING MECHANISMS FOR BALANCING MATCH

Here the passenger demand — service supply matching problem are treated as multi-agent negotiation in a distributed system. To simulate the supply-demand interaction in a closed-loop manner, the iterative/adjusting negotiations proceed based on marginal cost calculations to guide optimal policies for balancing between passenger demand and service supply. The generalized costs of the URT trip, e.g., the passengers' travel cost, the trains' unproductive cost and increased user cost, as modelled in Section IV and Section V, are what the agents' learning and adaptation is about. The condition or criteria for negotiation is the matching degree between passenger demand and service supply, which is an explicit representation of interactions between demand and supply in URT network systems. In accordance with the principle of machine learning, which uses the previous experience to improve their performance, the MAS architecture and the negotiation-based mechanism are employed to learn the reasonable headway time values for all agents in the URT network system, by combining agent-based modeling with machine learning [49]. Analogous to that work of [21], the passenger route choice behavior in the URT network is integrated into the multi-agent based transport assignment framework. The negotiation-based iterative/adjusting mechanisms for balancing match between passenger demand and capacity supply is designed as Fig. 2.

VII. NUMERICAL EXAMPLE FOR SIMULATION WITH NETLOGO

A. EXPERIMENTAL DESIGN

In a complex URT network system with many station nodes and track sections, trips between each origin-destination (OD) pairs can be made by using two or more alternative routes. The train lines go across each other at the transfer stations. Passengers have to change from one track line to another at the transfer stations. The number of OD pairs in the real URT network is very large, which approximates square relation with the number of stations. In order to reduce the URT network model scale, the zonal ODs can be divided by merging some ODs under the precondition of meeting the application demand. Just as in the work of [50], the representation of the railway infrastructure is quite abstract, omitting any consideration of its actual functioning details. In this numerical simulation example, our models and frameworks are tested on a compressed hypothetical URT network. The transit network is assumed to be given as a 6-node URT network with 3-track lines, including the red line, the green line and the blue line, as illustrated in Fig. 3. The link length of the network configuration is showed in Table 3. In the hypothetical URT network, the node 0 (with capacity 3200 passenger/h) is a central transfer station for three lines, node 4(with capacity 3600 passenger/h) is a larger transfer station for two lines, node 1(with capacity 3200 passenger/h) and node 3(with capacity 3200 passenger/h) are smaller transfer stations, node 2 (with capacity 2500 passenger/h) is an intermediate station, and node 5(with capacity 2000 passenger/h) is an end station.

Data availability is a big challenge when initiating new scenarios. The demand distribution and variation of passenger arrival rates over stations and periods can be estimated with APCs (automatic passenger counters). Since collecting the travel demand data is a very complex and expensive task. During the simulation, the OD passengers are generated by a trip demand distribution following a Poisson arrival process in an

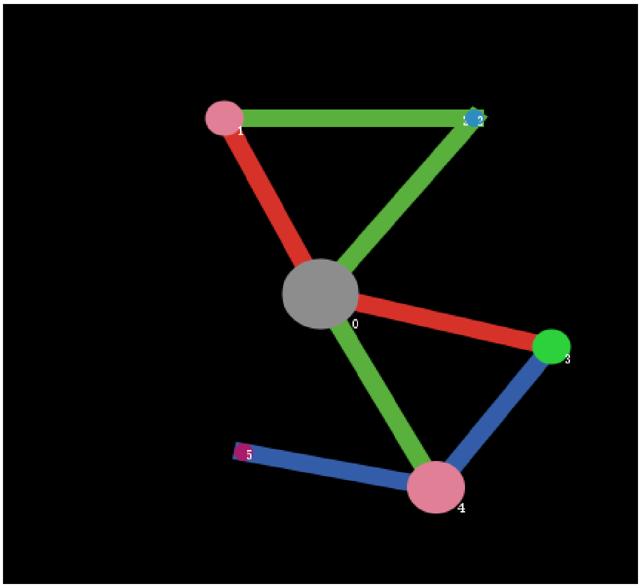

FIGURE 3. 6-node URT network with 3-track lines.

hourly-based schedule for the URT network system. Analogous to [10], the creation of passenger entities is driven by a time-dependent Poisson process in accordance with origin-destination matrices. Each newly generated passenger agent would select a path in the effective routes set. Especially, in node 0 the arrival rate is set as 3 persons per minute, while in other nodes the arrival rate is set as 2 persons per minute. The technically lowest possible headway is 2 minutes, and the highest one is set as 15 minutes. Using measures of the distances between all reasonable transfer options and the findings of the mean walking speed (i.e., 1.34 meters per second with a deviation of $\pm 19\%$ deviation), the mean transfer times can be calculated. And it varies between 60 seconds and 240 seconds among travelers. The dwell time at stations is set between 30 seconds and 60 seconds. The train speed is set as 9.72m/s. The maximal loading capacity of a single train is set as 1000 passengers. The length of period for planning is one hour. From the society-economy perspective, the empty seat average cost per time unit is estimated as 10 RMB, and the un-served passenger shortage cost per time unit 8 RMB.

TABLE 3. Link length of URT network configuration unit: meter.

	0	1	2	3	4	5
0		1200	1000	700	650	
1	1200		850			
2	1000	850				
3	700				900	
4	650			900		1100
5					1100	

B. SIMULATION RESULTS AND DISCUSSION

NetLogo [48] is a programmable modelling environment that can simulate the phenomena of nature and human society. It is most adaptable to model the complex system that evolves with the time, by which the modeler can instruct thousands

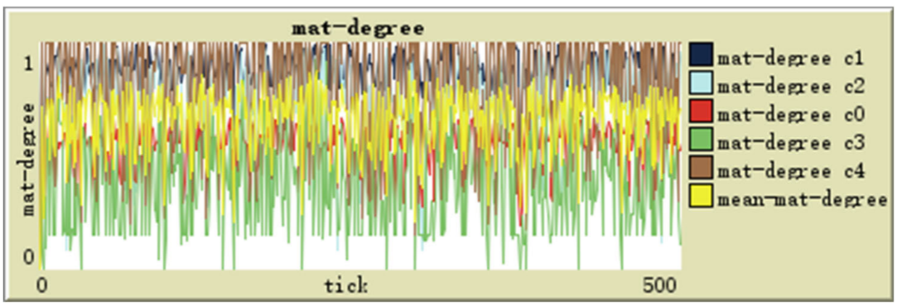

FIGURE 4. Mean matching degree profile for the whole URT network and matching degree profile for each of station nodes.

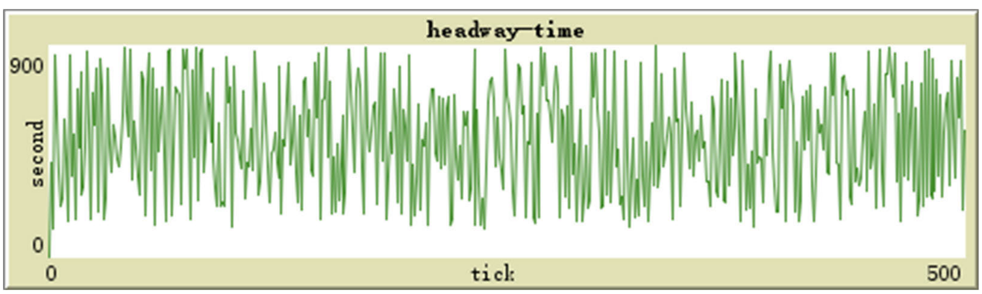

FIGURE 5. Headway time profile of train traffic in the URT network.

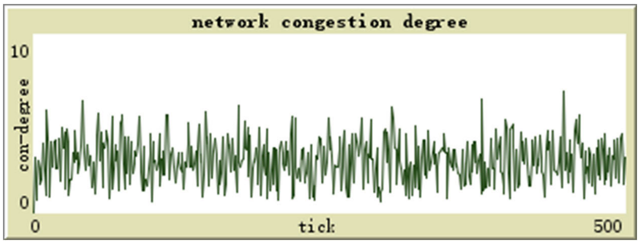

FIGURE 6. **Network congestion degree profile.**

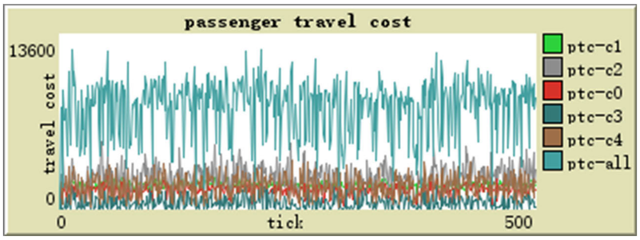

FIGURE 7. **Passenger travel cost profile.**

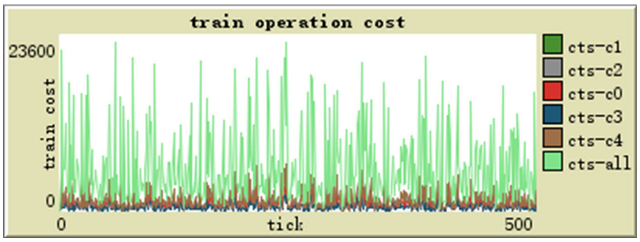

FIGURE 8. **Train operation cost profile.**

of independent agents to run. This makes it possible to explore the connections between the individual behavior of the micro level and the macro patterns, while the macro patterns emerge from the interactions of the individuals. By using the simulation platform NetLogo, the URT network (Fig. 3) is setup as the simulation environment, then the passenger OD demand matrix is generated following the random-Poisson arrival rates and three alternative routes, i.e., r1:1-0(transfer)-4(transfer)-5, r2:1-2-0-4(transfer)-5, r3: 1-2-0(transfer)-3(transfer)-4-5. After 500 ticks of simulation runs within 20 min, the profiles of the mean matching degree for the whole URT network and the matching degree for each of the station nodes (Fig. 4) can be achieved, also the headway time for trains operation in the URT network (Fig. 5), the network congestion degree (Fig. 6), the passenger travel cost(Fig. 7), the train operation cost(Fig. 8), and passengers distribution on routes (Fig. 9), correspondingly, where the series of alphabets c, i.e., c0, c1, c2, c3, c4, represent the station nodes of the URT network. Among these series of simulation results, the desirable mean matching degree for the whole URT network is achieved as 0.897433, at the cost of 11565.28 minutes for passengers travel and 3067.92 RMB for trains over-seat & empty-seat operation respectively. Particularly, at this desirable mean matching degree point, the desired headway time for train traffic in the network is 688 seconds, and the corresponding network congestion degree is 4.339.

Among these three alternative routes, the longest route r3 attracts the most volumes of the passenger flow, which indicates that it is reasonable to regard it as the critical route, and guarantees that the transit capacity is fully exploited. When the desirable mean matching degree from 500 ticks of simulation runs are obtained, the corresponding matching degree of the nodes from c0 to c4 is 0.0.595539, 0.97807, 0.9856, 0.927956,1, respectively. According to Table 2, all of the matching degrees for nodes fall within the domains of average class, which shows that the nodes, especially the central transfer node, incline to be the bottleneck in the network. In the URT network system, congestion occurs and propagates when the train service supply capacity cannot satisfy the passenger flow demand. From the perspective of the whole URT network, the value of the desirable mean matching degree (i.e., 0.897433) falls in the best value domains (i.e.,0.86∼0.90), which emphasizes the importance of representing the supply side and the demand side simultaneously, and justifies both the distributed architecture system and negotiation-based iterative/adjusting mechanisms proposed in this study. As the maximum mean matching degree results from 500 simulation runs is less than 1, according to formula (16) and (17), it can also judge that the URT network supply capacity is fully exploited in this way.

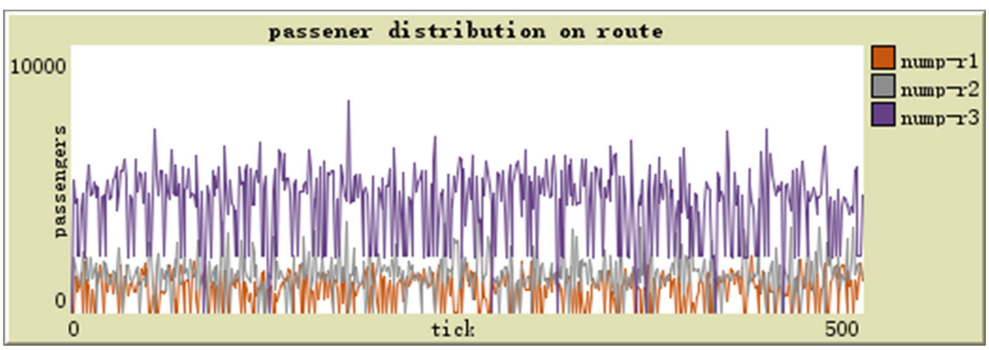

FIGURE 9. **Passenger distribution on routes.**

The study period in this paper is only one hour. However, dividing and then combining all the working hours in a day together, the within-day dynamic profile of service supply — passenger demand matching degree can be achieved, by using the methodology proposed in this paper in a divide-and-conquer way.

VIII. CONCLUSION

This research built a system to optimize the matching degree between passenger demand and service supply in the URT network. Agent was used as a tool to model the dynamic features of the passenger traffic and train traffic. The intelligent passenger demand — service supply matching methodology was proposed, including calculation of the matching degree, distributed multi-agent system architecture, and negotiation-based iterative/adjusting mechanisms for balancing match. The numerical example has demonstrated the effectiveness of the models and mechanisms through the distributed simulations on the NetLogo platform, and the optimum mean matching degree for the whole URT network can be obtained from 500 simulation ticks, i.e., 0.897433. The results show that it is reasonable to set the headway time as the key decision variable or the order parameter for optimization of passenger demand — service supply matching degree in the URT network. With the help of NetLogo, some feasible suggestions to optimize passenger demand— service supply matching degree are provided for URT network. These findings are meaningful for polices on both development of efficient URT network capacity usage strategies and provision of high level of service for passengers. Compared with the conventional methods, the proposed method has the following advantages:

1.Our models incorporate key factors jointly in urban rail transit network systems from both an operation-oriented and a passenger-oriented perspective. By adopting the multi-agent distributed simulation philosophy, all related goals/constraints from both the demand side and the supply side have been accounted for directly or indirectly, without the increment of the complexity for settling the problem.

2.By distinguishing between the train/vehicle congestion and the infrastructure/track network congestion, the formula for calculating the passenger general travel cost was updated systematically.

3.By using the NetLogo platform as the distributed simulations tool, as many as possible service supply options, i.e., the headway times, can be considered to provide a required service policy on the given URT infrastructure network.

4.Optimum headway times can be recommended according to the intelligent passenger demand — service supply matching degree in the URT network.

5. The efficiency of the proposed method is desirable, which can complete 500 ticks of simulation runs within 20 min on PC.

Further studies could consider the robust line planning through the elasticity of frequencies/headway time, and to take the passenger route choice behavior into account for the train traffic organization process. It is supposed to be applied in more URT networks to expand the application area of the method. In reality, the passenger OD flow demand may be asymmetric, but not absolutely. In certain ideal situation, it may be symmetric, i.e., the total passenger volume is symmetric within certain period. For simulation simplification, it is assumed to be symmetric. In theory, whether it is symmetric or not does not affect the matching mechanism in nature. In future research for real application, we will set free assumption 2 and apply the proposed matching methods to the asymmetric passenger OD flow demand (e.g., big data of passenger flow) situation for being closer to the reality. Due to limitation by available data to certain extent, this study only focused on the passenger route choice behavior in general traffic condition (of course, the congested condition is not excluded), by using the general travel cost function, i.e., formula (1) — formula (7). However, analogous to BPR function, the network congestion degree (i.e. the URT trains traffic of passenger demand surpasses the infrastructure capacity supply) was introduced besides in-vehicle train congestion and station congestion, which is more precise and practical compared with the prior general travel cost calculation. Exactly, the passenger behavior is more complex, e.g., traveling backward (TB) behavior [51]. The more complex behavior of passenger agents for the real case study with big data and artificial intelligence (AI) technique will be conducted in the future study. The difference between the numerical example for simulation with NetLogo and the real-world case study lies in the passenger flow data and URT network scale. The questions about the passenger flow data have been explained just then. Exactly, the principles for the application of the proposed models and frameworks to the real-word case study have been demonstrated in Section VII as far as the network scale is concerned.

REFERENCES

[1] Z. Huang, R. Xu, W. D. Fan, F. Zhou, and W. Liu, "Service-oriented load balancing approach to alleviating peak-hour congestion in a metro network based on multi-path accessibility," *Sustainability*, vol. 11, no. 1293, pp. 1–16, 2019.

[2] E. Hassannayebi, A. Sajedinejad, and S. Mardani, "Urban rail transit planning using a two-stage simulation-based optimization approach," *Simul. Model. Pract. Theory*, vol. 49, pp. 151–166, Dec. 2014.

[3] Z.-P. Du and A. Nicholson, "Degradable transportation systems: Sensitivity and reliability analysis," *Transp. Res. B, Methodol.*, vol. 31, no. 3, pp. 225–237, Jun. 1997.

[4] Y. Hamdouch, W. Y. Szeto, and Y. Jiang, "A new schedule-based transit assignment model with travel strategies and supply uncertainties," *Transp. Res. B, Methodol.*, vol. 67, pp. 35–67, Sep. 2014.

[5] E. Quaglietta, "A microscopic simulation model for supporting the design of railway systems: Development and applications," Ph.D. dissertation, Dep. Syst. Eng., Univ. Naples Federico II, Naples, Italy, 2011.

[6] A. Schöbel, "Line planning in public transportation: Models and methods," *OR Spectr.*, vol. 34, no. 3, pp. 491–510, Jul. 2012.

[7] M. Mehrabipour, L. Hajibabai, and A. Hajbabaie, "A decomposition scheme for parallelization of system optimal dynamic traffic assignment on urban networks with multiple origins and destinations," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 10, pp. 915–931, Oct. 2019.

[8] R. Guo and Z. Huang, "Mass rapid transit ridership forecast based on direct ridership models: A case study in Wuhan, China," *J. Adv. Transp.*, vol. 2020, Mar. 2020, Art. no. 7538508.

[9] X. Xu, Y. Lu, Y. Wang, J. Li, and H. Zhang, "Improving service quality of metro systems—A case study in the Beijing metro," *IEEE Access*, vol. 8, pp. 12573–12591, 2020.
[10] D. Schmaranzer, R. Braune, and K. F. Doerner, "Population-based simulation optimization for urban mass rapid transit networks," *Flexible Services Manuf. J.*, vol. 32, no. 4, pp. 767–805, Dec. 2020.
[11] M. Owais and M. K. Osman, "Complete hierarchical multi-objective genetic algorithm for transit network design problem," *Expert Syst. Appl.*, vol. 114, pp. 143–154, Dec. 2018.
[12] Y. Wang, B. De Schutter, T. van den Boom, B. Ning, and T. Tang, "Real-time scheduling for single lines in urban rail transit systems," in *Proc. IEEE Int. Conf. Intell. Rail Transp.*, Beijing, China, Aug. 2013, pp. 585–590.
[13] J. H. Banks, "Optimal headways for multiroute transit systems," *J. Adv. Transp.*, vol. 24, no. 2, pp. 127–155, Mar. 1990.
[14] T. Albrecht, "Automated timetable design for demand-oriented service on suburban railways," *Public Transp.*, vol. 1, no. 1, pp. 5–20, May 2009.
[15] K. Gkiotsalitis and O. Cats, "Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways," *Transp. Res. C, Emerg. Technol.*, vol. 88, pp. 187–207, Mar. 2018.
[16] A. Ceder, "Operational objective functions in designing public transport routes," *J. Adv. Transp.*, vol. 35, no. 2, pp. 125–144, Mar. 2001.
[17] A. Ceder, S. Hassold, and B. Dano, "Approaching even-load and even-headway transit timetables using different bus sizes," *Public Transp.*, vol. 5, no. 3, pp. 193–217, Oct. 2013.
[18] M. Goerigk and M. Schmidt, "Line planning with user-optimal route choice," *Eur. J. Oper. Res.*, vol. 259, no. 2, pp. 424–436, Jun. 2017.
[19] O. Cats, "Multi-agent transit operations and assignment model," *Procedia Comput. Sci.*, vol. 19, pp. 809–814, Jan. 2013.
[20] G. Zhang, H. Zhang, L. Li, and C. Dai, "Agent-based simulation and optimization of urban transit system," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 589–596, Apr. 2014.
[21] J. Narayan, O. Cats, N. van Oort, and S. Hoogendoorn, "Integrated route choice and assignment model for fixed and flexible public transport systems," *Transp. Res. C, Emerg. Technol.*, vol. 115, pp. 1–20, Jun. 2020.
[22] B. Chen and H. H. Cheng, "A review of the applications of agent technology in traffic and transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 485–497, Jun. 2010.
[23] N. Oliker and S. Bekhor, "A frequency based transit assignment model that considers online information," *Transp. Res. C, Emerg. Technol.*, vol. 88, pp. 17–30, Mar. 2018.
[24] X. Di, H. X. Liu, J.-S. Pang, and X. Ban, "Boundedly rational user equilibria (BRUE): Mathematical formulation and solution sets," *Transp. Res. B, Methodol.*, vol. 57, pp. 300–313, Nov. 2013.
[25] G. Desaulniers and M. D. Hickman, "Public transit," *Handbooks Oper. Res. Manage. Sci.*, vol. 14, pp. 69–127, Jan. 2007.
[26] A. Nuzzolo, F. Russo, and U. Crisalli, "A doubly dynamic schedule-based assignment model for transit networks," *Transp. Sci.*, vol. 35, no. 3, pp. 268–285, Aug. 2001.
[27] J.-D. Schmöcker, M. G. H. Bell, and F. Kurauchi, "A quasi-dynamic capacity constrained frequency-based transit assignment model," *Transp. Res. B, Methodol.*, vol. 42, no. 10, pp. 925–945, Dec. 2008.
[28] V. Trozzi, G. Gentile, M. G. H. Bell, and I. Kaparias, "Dynamic user equilibrium in public transport networks with passenger congestion and hyperpaths," *Transp. Res. B, Methodol.*, vol. 57, pp. 266–285, Nov. 2013.
[29] K. Kepaptsoglou and M. Karlaftis, "Transit route network design problem," *J. Transp. Eng.*, vol. 135, no. 8, pp. 491–505, 2009.
[30] R. Ding, N. Ujang, H. B. Hamid, M. S. A. Manan, R. Li, S. S. M. Albadareen, A. Nochian, and J. Wu, "Application of complex networks theory in urban traffic network researches," *Netw. Spatial Econ.*, vol. 19, no. 4, pp. 1281–1317, Dec. 2019.
[31] Y. Hamdouch, P. Marcotte, and S. Nguyen, "Capacitated transit assignment with loading priorities," *Math. Program.*, vol. 101, no. 1, pp. 205–230, Aug. 2004.
[32] C. Anda, A. Erath, and P. J. Fourie, "Transport modelling in the age of big data," *Int. J. Urban Sci.*, vol. 21, no. 1, pp. 19–42, Aug. 2017.
[33] J. Liu and X. Zhou, "Capacitated transit service network design with boundedly rational agents," *Transp. Res. B, Methodol.*, vol. 93, pp. 225–250, Nov. 2016.
[34] L. Sun, Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen, "An integrated Bayesian approach for passenger flow assignment in metro networks," *Transp. Res. C, Emerg. Technol.*, vol. 52, pp. 116–131, Mar. 2015.
[35] J. Zhang, "A review on route choice behavior and volume control of passengers in urban rail transit network," *IOP Conf., Mater. Sci. Eng.*, vol. 677, Dec. 2019, Art. no. 042047.
[36] J. Schmocker, "Dynamic capacity constrained transit assignment," Ph.D. dissertation, Dep. Civil Environ. Eng., London Univ., London, U.K., 2006.
[37] J. de Cea and E. Fernández, "Transit assignment for congested public transport systems: An equilibrium model," *Transp. Sci.*, vol. 27, no. 2, pp. 133–147, May 1993.
[38] H. S. Mahmassani, "Dynamic network traffic assignment and simulation methodology for advanced system management applications," *Netw. Spatial Econ.*, vol. 1, nos. 3–4, pp. 267–292, Sep. 2001.
[39] H. S. Mahmassani and Y.-H. Liu, "Dynamics of commuting decision behaviour under advanced traveller information systems," *Transp. Res. C, Emerg. Technol.*, vol. 7, nos. 2–3, pp. 91–107, Apr. 1999.
[40] V. Guihaire and J.-K. Hao, "Transit network design and scheduling: A global review," *Transp. Res. A, Policy Pract.*, vol. 42, no. 10, pp. 1251–1273, Dec. 2008.
[41] H. Gehlot and S. V. Ukkusuri, "User equilibrium with a policy-based link transmission model for stochastic time-dependent traffic networks," *Transportmetrica B, Transp. Dyn.*, vol. 7, no. 1, pp. 1013–1043, Dec. 2019.
[42] Transportation Research Board, *Transit Capacity and Quality of Service Manual*, 2nd ed. Washington, DC, USA: National Research Council, 2003.
[43] Y. Shafahi and A. Khani, "A practical model for transfer optimization in a transit network: Model formulations and solutions," *Transp. Res. A, Policy Pract.*, vol. 44, no. 6, pp. 377–389, Jul. 2010.
[44] K. Kogan and M. Shnaiderman, "Continuous-time replenishment under intermittent observability," *IEEE Trans. Autom. Control*, vol. 55, no. 6, pp. 1460–1465, Jun. 2010.
[45] Y. Hadas and M. Shnaiderman, "Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time," *Transp. Res. B, Methodol.*, vol. 46, no. 8, pp. 1068–1084, Sep. 2012.
[46] R. Liu, S. Li, and L. Yang, "Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy," *Omega*, vol. 90, Jan. 2020, Art. no. 101990.
[47] U. Wilensky and W. Rand, *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems With NetLogo*. Cambridge, MA, USA: MIT Press, 2015.
[48] U. Wilensky, "NetLogo," Center Connected Learn. Comput. Model., Northwestern Univ., Evanston, IL, USA, Tech. Rep. 1419AQ, 1999. [Online]. Available: http://ccl.northwestern.edu/netlogo/
[49] R. V. Vohra and M. P. Wellman, "Foundations of multi-agent learning: Introduction to the special issue," *Artif. Intell.*, vol. 171, no. 7, pp. 363–364, May 2007.
[50] H. Proença and E. Oliveira, "An adaptive multi-agent system for railway traffic control," *GESTS Int. Trans.*, vol. 9, no. 1, pp. 1–12, 2005.
[51] C. Yu, H. Li, X. Xu, and J. Liu, "Data-driven approach for solving the route choice problem with traveling backward behavior in congested metro systems," *Transp. Res. E, Logistics Transp. Rev.*, vol. 142, pp. 1–28, Oct. 2020.

JIAMIN ZHANG received the Ph.D. degree from Beijing Jiaotong University.

She is currently a Lecturer with the College of Transportation, Shandong University of Science and Technology, Qingdao, China. Her research interests include planning and management of railway transportation, urban rail transit, public transport, capacity planning, behavioral modeling, systems engineering, supply chain management, intelligent transportation systems, and smart transport.