Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems
M. Shafik (Ed.)

doi:10.3233/ATDE240030

A Collaborative Optimization Method for Train Scheduling and Passenger Flow Assignment Based on Multi-Agent Reinforcement Learning

Xinyi NING [a], Wei DONG [b,1], Xinya SUN [a], and Yindong JI [a]
[a] *Department of Automation, Tsinghua University, Beijing, China*
[b] *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China*

Abstract. This paper presents an online optimization method for metro network train scheduling and passenger flow assignment based on multi-agent reinforcement learning, aiming at minimizing traction energy consumption and average passenger waiting time. The problem is modeled as a multi-agent Markov decision process using a multi-agent actor-critic framework for network train scheduling and a deep deterministic policy gradient framework for passenger flow assignment. All agents interact with the same metro simulation environment, which generates train timetables and passenger flow assignments that meet complex constraints. Results of the case study on anonymized data of Chongqing Metro show that the proposed method outperforms baseline scenarios and is able to adjust train schedules and passenger flow assignments in real-time when passenger flow distribution fluctuates, demonstrating its effectiveness and robustness.

Keywords. Metro network system, Train scheduling, Passenger flow assignment, Multi-agent deep reinforcement learning, Online collaborative optimization

1. Introduction

Metro systems play an important role in alleviating traffic congestion and improving the quality of life for people. However, the enormous energy consumption contributes to significant operating costs. At the same time, train schedules often fail to match dynamic passenger demand. During peak hours, overwhelming passenger flow frequently surpasses train capacity, whereas during off-peak hours, the low number of passengers may cause resource wastage. Therefore, it is vital to achieve a match between passenger demand and train scheduling to improve the energy efficiency of the metro system while maintaining high-quality passenger service.

There have been a number of researches regarding the online optimization of train timetables or passenger flow. Many studies employ integer programming or mixed integer programming to solve real-time train scheduling problem [1,2,3]. Distinct from integer programming or heuristic methods, reinforcement learning yields not a specific

[1] Corresponding Author: Wei Dong, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, 100086, China; E-mail: weidong@mail.tsinghua.edu.cn

solution, but a trained model, which can rapidly generate a new solution when environment shifts, making it suitable for online optimization. Many scholars have developed reinforcement learning for train scheduling with a focus on minimizing energy consumption [4] or passenger waiting time [5]. Yang et al. [6] applied deep deterministic policy gradient algorithm (DDPG) to timetable rescheduling problem under disturbances with the goal of energy conservation. Ying et al. [7] proposed a proximal policy optimization method based on deep reinforcement learning to achieve integrated optimization of metro service scheduling and train composition on a single line. They later extended the problem to multiple lines in a metro network and employed a multi-agent deep deterministic policy gradient algorithm to it [8]. Aiming at real-time passenger route guidance, Jia [9] implemented DDPG algorithm under the condition that guidance information is provided to every passenger, and mitigated the congestion problem in partial network.

However, there is room for further optimization when considering the entire system, rather than optimizing the train schedule or passenger flow distribution unilaterally. Train flow, passenger flow, and energy flow form the heart of the dynamic operation of the metro system. They do not exist independently but instead interact with each other. Achieving integrated optimization of the train flow and passenger flow is the key to improving the overall efficiency of the metro system. Some studies propose train scheduling with collaborative passenger flow control on oversaturated metro lines, limiting the number of passengers from entering platforms to prevent congestion [10,11,12,13]. Shang [14] developed a train operation and passenger distribution space-time network to optimize train skip-stopping pattern, schedules, and passenger flow control strategies. Liu et al. [15] proposed a mixed integer linear programming model for train scheduling, train connection, and passenger flow control issues.

Other literature considers offering appropriate guidance for passengers during route selections and proposes the joint optimization of passenger flow assignment and train scheduling [16]. Zhang et al. [17] proposed a global safety evaluation method for regional rail transit systems and explored the collaborative optimization of passenger flow assignment and train scheduling from minimizing global risk. Zhao [18] considered the same problem from the perspective of energy consumption and passenger waiting time using CCGA combined with NSGA–II. However, most of the methods are suitable only for offline optimization. When disturbances occur in train operations or passenger flow, they are unable to adjust them in real-time. Therefore, further research is needed for online collaborative optimization of train scheduling and passenger flow assignment.

This paper introduces a collaborative online optimization method for train scheduling and passenger flow assignment based on multi-agent reinforcement learning. A multi-agent actor-critic framework is used for train scheduling and a deep deterministic policy gradient framework is adopted for passenger flow assignment. At each stage, the schedule control agents decide the dispatching time of the service, the travel times between stations and the dwell times at each station. The agent for passenger flow assignment provides the distribution ratio of different passenger routes. The goal is to minimize traction energy consumption and the average passenger waiting time. All agents interact with the same metro simulation environment, which generates train timetables and passenger flow assignments that meet complex constraints. The approach allows for real-time adjustments to timetables and passenger flow assignments in response to dynamic passenger flow. Case studies show that the proposed method outperforms baseline scenarios, demonstrating its effectiveness and robustness.

The rest of the paper is organized as follows. Section 2 provides a description of the metro system model. Section 3 presents the problem as a multi-agent Markov decision process and introduces the deep reinforcement learning framework. Section 4 tests the effectiveness of the approach under anonymized data from real-world scenarios of Chongqing metro. Section 5 presents conclusions and future work.

2. Metro System Model

2.1. Assumptions

In order to simplify the metro system, we adopt the following assumptions:

1. All passengers are assumed to follow the assigned route, thereby implementing passenger flow guidance strategy [18].
2. Passengers are assumed to arrive directly at the platform to wait for the train, with the walking time at the stations disregarded.
3. The process of passenger embarkation and disembarkation is assumed to be instantaneous and not affected by the dwelling time of the trains [18].
4. All trains are assumed to run according to the timetable without encountering congestion or disturbances [8].

Parameters and symbols used in the metro system model are shown in Table 1.

2.2. Metro Network Model

For the metro network, $\mathcal{L} = \{1,2,\ldots,L\}$ denotes the set of service lines, where L is the total number of lines, and $l \in \mathcal{L}$ symbolizes each individual line. Every line has two directions which share the same physical stations but possess different parameters and states when calculating passenger and train flow. To make a distinction, we use $\hat{\mathcal{L}} = \{\pm 1, \pm 2, \ldots, \pm L\}$ to represent the set of lines in different directions, where $\hat{l} \in \hat{\mathcal{L}}$ is the line index, $\hat{l} = 1,2,\ldots,L$ corresponds to the upward direction of the lines, and $\hat{l} = -1,-2,\ldots,-L$ represents their downward direction.

We further define $\mathcal{R}_{\hat{l}} = \{1_{\hat{l}}, 2_{\hat{l}}, \ldots, \mathrm{R}_{\hat{l}}\}$ as the set of stations on line l, where $\mathrm{R}_{\hat{l}}$ is the total number of stations, and $r_{\hat{l}} \in \mathcal{R}_{\hat{l}}$ stands for each station. The origin station in the upward direction on line l is 1_l, and the terminal station is R_1. For the downward direction, it starts at 1_{-l} and terminates at R_{-1}.

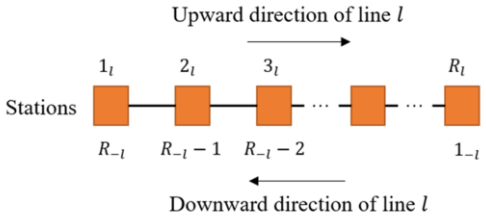


Figure 1. Index of stations in the upward and downward directions on line l

Table 1. Parameters and symbols used in the metro system model

Notation	Definition
$\hat{\mathcal{L}}$	Set of service lines that indicates different directions
$\hat{l}$	Index of service lines that indicates different directions
l	Index of service lines that do not indicate different directions
$\mathrm{R}_{\hat{l}}$	Number of stations on line $\hat{l}$
$\mathcal{R}_{\hat{l}}$	Set of stations on line $\hat{l}$
$r_{\hat{l}}$	Index of stations on line $\hat{l}$
k_l	Index of train service on line l
$t_{k_l,r_{\hat{l}}}^{depart}$	Departure time of service k_l at station r_l
$t_{k_l,r_{\hat{l}}}^{arrive}$	Arrival time of service k_l at station $r_{\hat{l}}$
$t_{k_l,r_{\hat{l}}}^{dwell}$	Dwelling time of service k_l at station $r_{\hat{l}}$
$t_{k_l,r_{\hat{l}}}^{run}$	Running time of service k_l between station $r_{\hat{l}}$ and station $r_{\hat{l}}+1$
T_{begin}	Metro operation start time
T_{end}	Metro operation end time
$E^j(t)$	The traction energy in the power supply zone j from the start of the simulation to time t
ϕ	The sampling frequency of the OD matrix
$w_{i,r_{\hat{l}}}$	The passenger waiting time at station $r_{\hat{l}}$ from the departure of the (i-1)-th train until the arrival of the i-th train
$C_{i,r_{\hat{l}}}$	The number of passengers on board when the i-th train on line $\hat{l}$ leaves station $r_{\hat{l}}$
J_d	Set of stations in section d
w_d^j	The average waiting time at each station of section d during $\left(T_{begin}+(j-1)\phi, T_{begin}+j\phi\right]$
p_d^j	The average load rate at each station of section d during $\left(T_{begin}+(j-1)\phi, T_{begin}+j\phi\right]$

2.3. Train Operation Model

We now present the discrete event model for train operations. Given the train service k_l on line l, the departure time of such service at station $r_{\hat{l}}$ can be derived as

$$t_{k_l,r_{\hat{l}}}^{depart} = t_{k_l,r_{\hat{l}}-1}^{depart} + t_{k_l,r_{\hat{l}}}^{dwell} + t_{k_l,r_{\hat{l}}-1}^{run}, r_{\hat{l}} \in \mathcal{R}_{\hat{l}}, r_{\hat{l}} \neq 1_{\hat{l}} \tag{1}$$

where $t_{k_l,r_{\hat{l}}}^{run}$ is the running time between station $r_{\hat{l}}$ and station $r_{\hat{l}}+1$ and $t_{k_l,1_{\hat{l}}}^{depart}$ is the dwelling time at station $r_{\hat{l}}$. Similarly, the arrival time of service k_l at station $r_{\hat{l}}$ can be determined as

$$t_{k_l,r_{\hat{l}}}^{arrive} = t_{k_l,r_{\hat{l}}-1}^{depart} + t_{k_l,r_{\hat{l}}-1}^{run}, r_{\hat{l}} \in \mathcal{R}_{\hat{l}}, r_{\hat{l}} \neq 1_{\hat{l}} \tag{2}$$

We adopt the optimal speed curve of maximum acceleration, coasting and maximum braking [19] for train operations between stations, where it first reaches a predetermined speed with maximum traction force, then switches to coasting before applying maximum braking force until it stops. The running resistance is estimated by the Davis Formula [18]. Constant coefficients are used to calculate the motor and line losses. Then, the mechanical power of the train can be determined according to its dynamics, and the electric power $P_T^j(t)$ and $P_R^j(t)$, produced by trains in the traction and braking states respectively within power supply zone j at time t, can be derived. Let the proportion of regenerative energy directly used by other trains be λ_r, the total traction energy $E^j(t)$ in the power supply zone j from the start of the simulation to time t can be established as

$$E^j(t) = \int_0^t max\left(P_T^j(u) + \lambda_r P_R^j(u), 0\right)du \quad (3)$$

2.4. Passenger Flow Model

Passenger flow can be modeled by OD demand, which describes the number of passengers from departure stations to destination stations. Let ϕ be the the sampling frequency, T_{begin} be the metro operation start time, and B be the total duration, then the OD demand of line l in the time period $(\tau - \phi, \tau]$ can be represented by the matrix $OD_l(\tau)$, where $\tau \in \{T_{begin} + \phi, \dots, T_{begin} + B\phi\}$. From the OD matrix, we can calculate passenger entry rate $od_{r_{\hat{l}}, r'_{\hat{l}}}(t)$ from station $r_{\hat{l}}$ to station $r'_{\hat{l}}$ on line l at time t (unit: person/unit time). The number of passengers arriving at station $r_{\hat{l}}$ with a destination of station $r'_{\hat{l}}$ is the time integral of $od_{r_{\hat{l}}, r'_{\hat{l}}}(t)$. If all passengers waiting at station $r_{\hat{l}}$ can board the train without causing overloading, then they all board. Otherwise, they board according to a certain ratio, so that the train is just fully loaded.

Let $n_{r_{\hat{l}}}(t)$ be the number of passengers waiting at station $r_{\hat{l}}$ at time t, the waiting time $w_{i,r_{\hat{l}}}$ for passengers at station $r_{\hat{l}}$ from the moment the (i-1)-th train departs until the arrival of the i-th train can be calculated as

$$w_{i,r_{\hat{l}}} = \int_{t_{i-1,r_{\hat{l}}}^{depart}}^{t_{i,r_{\hat{l}}}^{depart}} n_{r_{\hat{l}}}(t)dt \quad (4)$$

In order to extract the characteristics of passenger flow distribution, we divide the network into different sections using transfer stations as segmentation points. The two ends of each section are the two closest transfer stations in the same direction on the same line. Let D be the total number of sections, J_d be the set of stations in section d, R_d be the total number of stations in section d, then the average waiting time w_d^j and the average load rate p_d^j at each station of section d during $\left(T_{begin} + (j-1)\phi, T_{begin} + j\phi\right]$ can be obtained as

$$w_{\mathrm{d}}^j = \frac{\sum_{r_{\hat{l}} \in J_d} \int_{T_{begin}+(j-1)\phi}^{T_{begin}+j\phi} n_{r_{\hat{l}}}(t)\mathrm{d}t}{R_d} \quad (5)$$

$$p_d^j = \frac{\sum_i C_{i,r_{\hat{l}}}}{R_d C_{max} Q_d^j}, t_{i,r_{\hat{l}}}^{depart} \in \left(T_{begin} + (j-1)\phi, T_{begin} + j\phi\right], r_{\hat{l}} \in J_d \quad (6)$$

where Q_d^j represents the number of trains passing through section d during $\left(T_{begin} + (j-1)\phi, T_{begin} + j\phi\right]$, $C_{i,r_{\hat{l}}}$ is the number of passengers on board when the i-th train on line $\hat{l}$ leaves station $r_{\hat{l}}$, and C_{max} is the maximum number of passengers on a train.

For passenger transfers, if a passenger traveling from station `a' to station `m' chooses the route `a-b-c-g-h-l-m' (see Figure 2), we divide it into three different routes `a-b-c', `c-g-h', and `h-l-m', generate three passengers each choosing the three routes separately, and add them to the OD matrix.

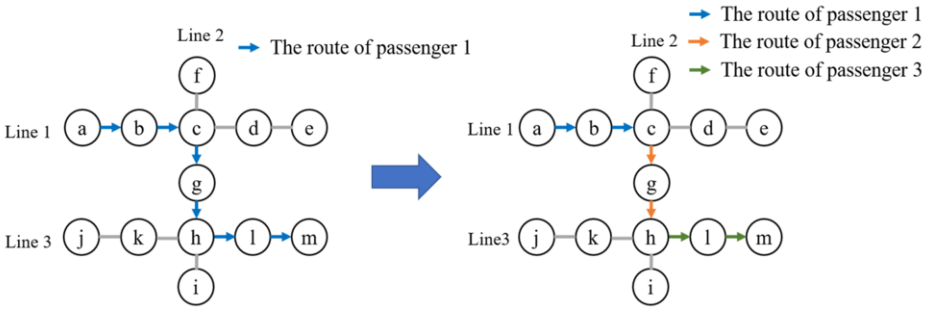


Figure 2. Diagram of passenger transfer model

3. Metro System Model

In this section, we model train scheduling and passenger flow assignment problems as Markov decision processes. We develop multiple agents each responsible for train scheduling in one direction on one line, along with another agent specifically for passenger flow assignment. All agents interact with the same metro simulation environment, which is built according to Section 2. The list of associated symbols is shown in Table 2.

Table 2. Symbols used in the reinforcement learning model

Notation	Definition
$\boldsymbol{s}_{n_{\hat{l}}}$	State feature set of train scheduling on line $\hat{l}$ at stage n
$\boldsymbol{a}_{n_{\hat{l}}}$	Action set of train scheduling on line $\hat{l}$ at stage n
$\boldsymbol{s}_{n_{psg}}$	State feature set of passenger flow assignment at stage n
$\boldsymbol{a}_{n_{\mathrm{psg}}}$	Action set of passenger flow assignment at stage n
$b_{n_{\hat{l}},r_{\hat{l}}}$	The number of people boarding the n-th train of line $\hat{l}$ at station $r_{\hat{l}}$
$x_{n,ij}^{m}$	The proportion of passengers traveling from station i to station j assigned to the m-th route during $\left(T_{begin}+n\phi, T_{begin}+(n+1)\phi\right]$
$\omega_{\mathrm{n,d}}$	The attraction factor of section d to the passenger flow during $\left(T_{begin}+n\phi, T_{begin}+(n+1)\phi\right]$
D_{ij}^{m}	The set of sections in the m-th route of the OD pair from station i to station j
$r_{n_{\hat{l}}}$	Reward of train scheduling on line $\hat{l}$ at stage n
$r_{n_{\mathrm{psg}}}$	Reward of passenger flow assignment at stage n
t^{turn}	Minimum turnaround time for a train at the terminal station
h_{min}, h_{max}	Minimum and maximum departure intervals of trains
u_{min}, u_{max}	Minimum and maximum headways over all stations
g_{min}, g_{max}	Minimum and maximum dwelling time of trains
$d_{r_{\hat{l}},min}, d_{r_{\hat{l}},max}$	Minimum and maximum running time of trains between station $r_{\hat{l}}$ and station $r_{\hat{l}}+1$
C_{max}	Maximum number of passengers on a train
$Q_{\hat{l}}$	Maximum number of trains that can be dispatched on line $\hat{l}$

3.1. State Sets and Transitions

We model train scheduling as a multi-agent Markov process, where each agent generates the timetable of all trains on one direction of a single line. Stage n describes the journey

of a train running from the origin station to the terminal station. For each line \(\hat{l}\), we define the state feature set of stage n as

$$\boldsymbol{s}_{n_{\hat{l}}} = \{[w_{n_{\hat{l}},r_{\hat{l}}}, t^{depart}_{n_{-\hat{l}}-1,1_{-\hat{l}}}, t^{run}_{n_{-\hat{l}}-1,r_{-\hat{l}}}, b_{n_{\hat{l}},r_{\hat{l}}}, t^{depart}_{n_{\hat{l}},r_{\hat{l}}}, t^{arrive}_{n_{\hat{l}},r_{\hat{l}}}] | r_{\hat{l}} \in \mathcal{R}_{\hat{l}}\} \tag{7}$$

where $w_{n_{\hat{l}},r_{\hat{l}}}$ is the passenger waiting time at station $r_{\hat{l}}$ from the departure of the (n-1)-th train until the arrival of n-th train, $t^{depart}_{n_{-\hat{l}}-1,1_{-\hat{l}}}$ and $t^{run}_{n_{-\hat{l}}-1,r_{-\hat{l}}}$ are the departure and running times of the (n-1)-th train in the opposite direction of line $\hat{l}$, and $b_{n_{\hat{l}},r_{\hat{l}}}$ is the number of people boarding the n-th train of line $\hat{l}$ at station $r_{\hat{l}}$.

The train scheduling simulation is established with a variable step size and $P_{n_{\hat{l}}}$ is defined in $\boldsymbol{s}_{n_{\hat{l}}+1} \sim P_{n_{\hat{l}}}(\boldsymbol{s}_{n_{\hat{l}}}, \boldsymbol{a}_{n_{\hat{l}}})$ as the state transition function. The initial state is set to $\boldsymbol{s}_{0_{\hat{l}}} = \boldsymbol{0}$. The agent gives action $\boldsymbol{a}_{n_{\hat{l}}}$ to the environment, and the environment updates the next state $\boldsymbol{s}_{n_{\hat{l}}+1}$. The episode terminates when the arrival time of the last train at the terminal is greater than metro operation end time T_{end}.

For passenger flow assignment, we also model it as a Markov decision process. Stage n describes the passenger flow in $\left(T_{begin} + (n-1)\phi, T_{begin} + n\phi\right]$, where n=1,2,..,B. Given the total number of sections D, each stage n can be characterized by the average waiting time w^n_d and the average load rate p^n_d calculated in Eq.(5) and Eq.(6).

$$\boldsymbol{s}_{n_{psg}} = \{[n, w^n_d, p^n_d] | d \in [1, D]\} \tag{8}$$

We use the fixed step size of ϕ in the passenger flow assignment simulation and define P_{psg} in $\boldsymbol{s}_{n_{psg}+1} \sim P_{psg}(\boldsymbol{s}_{n_{psg}}, \boldsymbol{a}_{n_{\mathrm{psg}}})$ as the state transition function. The initial state is set to $\boldsymbol{s}_{0_{psg}} = \boldsymbol{0}$ and the episode terminates when n equals B.

3.2. Action Sets

The action set $\boldsymbol{a}_{n_{\hat{l}}}$ of the schedule control agent of line $\hat{l}$ includes the departure interval $h_{n_{\hat{l}}+1}$ between the n-th train and the (n+1)-th train, the running times and dwelling times of the (n+1)-th train at each station.

$$\boldsymbol{a}_{n_{\hat{l}}} = \{[h_{n_{\hat{l}}+1}, t^{run}_{n_{\hat{l}}+1,r_{\hat{l}}}, t^{dwell}_{n_{\hat{l}}+1,r_{\hat{l}}}] | r_{\hat{l}} \in \mathcal{R}_{\hat{l}}, r_{\hat{l}} \neq \mathrm{R}_{\hat{l}}\} \tag{9}$$

The action set $\boldsymbol{a}_{n_{psg}}$ of the agent for passenger flow assignment is denoted as

$$\boldsymbol{a}_{n_{psg}} = \{\omega_{n,d} | d \in [1, D]\} \tag{10}$$

where $\omega_{n,d}$ is the attraction factor of section d to the passenger flow during $\left(T_{begin} + n\phi, T_{begin} + (n+1)\phi\right]$. We further convert it into the proportion of each OD demand choosing each route. We denote the number of routes of the OD pair from station i to station j as M_{ij}, the set of stations in the m-th route as D^m_{ij}, and the proportion of passengers assigned to the m-th route in stage n as $x^m_{n,ij}$. To satisfy $\sum_{m=1}^{M_{ij}} x^m_{ij} = 1$, we have

$$x_{n,ij}^{m} = \frac{\sum_{d \in D_{ij}^{m}} \omega_{n,d}}{\sum_{p=1}^{M_{ij}} \sum_{d \in D_{ij}^{p}} \omega_{n,d}} \tag{11}$$

3.3. Reward Functions

The optimization goal for train scheduling is to minimize total traction energy consumption and the average passenger waiting time. We design the reward $r_{n_{\hat{l}}}^{E}$ related to the energy consumption on line and the reward $r_{n_{\hat{l}}}^{w}$ associated with passenger waiting time. Given the coefficients b_1 and b_2, the reward function $r_{n_{\hat{l}}}$ for line $\hat{l}$ can be written down as

$$r_{n_{\hat{l}}} = b_1 r_{n_{\hat{l}}}^{E} + b_2 r_{n_{\hat{l}}}^{w} \tag{12}$$

For the reward $r_{n_{\hat{l}}}^{w}$, we define $g_{n_{\hat{l}}}$ as the total number of passengers waiting on the platform when the n-th train arrives at each station, $w_{n_{\hat{l}}}$ as the sum of $w_{n_{\hat{l}},r_{\hat{l}}}$ at each station, w_0 as a specified waiting time benchmark, b_3 and b_4 as the weighting factors, then the reward can be represented as

$$\Delta w_{n_{\hat{l}}} = w_{n_{\hat{l}}} - w_0 g_{n_{\hat{l}}} \tag{13}$$

$$r_{n_{\hat{l}}}^{w} = \operatorname{sgn}\left(\Delta w_{n_{\hat{l}}}\right) g_{n_{\hat{l}}} \, exp\left(b_3 \left|\Delta w_{n_{\hat{l}}}\right| - 1\right) + b_4 g_{n_{\hat{l}}} \tag{14}$$

For the reward $r_{n_{\hat{l}}}^{E}$, we define $E_{n_{\hat{l}}}$ as the total traction energy from the start of the simulation to time $t_{n_{\hat{l}},R_{\hat{l}}}^{arrive}$, $E_{base}(t_{n_{\hat{l}},R_{\hat{l}}}^{arrive})$ as the known traction energy of the baseline scenario without optimization up to time $t_{n_{\hat{l}},R_{\hat{l}}}^{arrive}$, then the reward can be derived as

$$r_{n_{\hat{l}}}^{E} = 1 - \frac{E_{n_{\hat{l}}}}{E_{base}\left(t_{n_{\hat{l}},R_{\hat{l}}}^{arrive}\right)} \tag{15}$$

The objective for passenger flow assignment is to minimize total passenger waiting time. We consider the reward function $r_{n_{\mathrm{psg}}}$ for passenger flow assignment at stage n as

$$r_{n_{\mathrm{psg}}} = \mathrm{b}_5 \sum_{i=n-\delta}^{n} \sum_{d=1}^{D} w_{base,d}^{\mathrm{i}} - \sum_{i=n-\delta}^{n} \sum_{d=1}^{D} w_{d}^{\mathrm{i}} \tag{16}$$

where w_d^{i} is the average waiting time of each station in section d during $\left(T_{begin} + (i-1)\phi, T_{begin} + i\phi\right]$, $w_{base,d}^{\mathrm{i}}$ is the corresponding known waiting time of the baseline scenario, and b_5 is a coefficient. The reward $r_{n_{\mathrm{psg}}}$ reflects the gap in passenger waiting time between the optimized passenger flow and the original flow during $\left(T_{begin} + (n-\delta-1)\phi, T_{begin} + n\phi\right]$.

Finally, we aim to minimize the cumulative rewards of train scheduling and passenger flow assignment over all stages.

3.4. Constraints

There are a series of constraints set for the optimization model. We first have the boundaries for train departure intervals as

$$h_{min} \leq t^{depart}_{k'_l,1_{\hat{l}}} - t^{depart}_{k_l,1_{\hat{l}}} \leq h_{max}, \forall k'_l > k_l \tag{17}$$

where h_{min} and h_{max} are the minimum and maximum departure intervals respectively. We also set a limit Q_l for the number of trains that can be dispatched on line l and C_{max} for the maximum number of passengers on a train. Furthermore, there is a minimum turnaround time t^{turn} for a train (denoted as q_l) after it arrives at the terminal station.

$$t^{depart}_{q_l,k'_l,1_{-\hat{l}}} \geq t^{arrive}_{q_l,k_l,R_{\hat{l}}} + t^{turn}, \forall k'_l > k_l \tag{18}$$

We introduce u_{min} and u_{max} as the minimum and maximum headway over all stations, which is the time interval between the departure of the previous train and the arrival of the next train at the same station.

$$u_{min} \leq t^{arrive}_{k'_l,r_{\hat{l}}} - t^{depart}_{k_l,r_{\hat{l}}} \leq u_{max}, \forall k'_l > k_l \tag{19}$$

The dwelling time of trains has to satisfy the minimum and maximum bounds of g_{min} and g_{max}. Similarly, there are boundaries for the running time of trains between station $r_{\hat{l}}$and station $r_{\hat{l}} + 1$, noted as $d_{r_{\hat{l}},min}$ and $d_{r_{\hat{l}},max}$.

$$g_{min} \leq t^{dwell}_{k_l,r_{\hat{l}}} \leq g_{max} \tag{20}$$

$$d_{r_{\hat{l}},min} \leq t^{run}_{k_l,r_{\hat{l}}} \leq d_{r_{\hat{l}},max} \tag{21}$$

In terms of passenger flow assignment, we believe the longest feasible path for each OD should be at most three stops or one transfer station longer than the shortest path to avoid assigning passengers to excessively long paths. In addition, because the distribution ratio $x^m_{n,ij}$ needs to be non-negative, if $\sum_{p=1}^{M_{ij}} \sum_{d \in D^p_{ij}} \omega_{n,d} = 0$ in Eq. (11), all passenger flows will be allocated to the shortest path.

3.5. Multi-agent Reinforcement Learning Framework

Due to the dimension explosion problem in train scheduling and passenger flow assignment, it is difficult to find an optimal solution. Here, we propose a multi-agent reinforcement learning framework to solve the train scheduling problem, and a deep deterministic policy gradient framework for passenger flow assignment.

Regarding train scheduling, we use a multi-agent framework based on the actor-critic algorithm (MAA2C). We adopt two actor networks for each line responsible for the scheduling of trains in the upward and downward direction respectively. The two actors work cooperatively, sharing the same reward and critic network. During training,

the decentralized actors receive local states from the environment and derive a set of actions with the goal of finding the optimal policy that maximizes future rewards. The centralized critic is used to estimate the state value function from a global perspective. During execution, only the actors are required to generate actions, while the critic is no longer needed. This framework with centralized training and decentralized execution not only makes it possible to maintain the association and synergy between the two agents but also speed up the execution process [20].

We further extend the single-line MAA2C network to multiple lines, thus forming the framework for network train scheduling (See Figure 3). Given that the trains running on different lines are not shared, train schedules on one line have negligible impact on the energy consumption or passenger waiting times on other lines. Therefore, we consider scheduling on different lines as independent tasks, with the MAA2C network for each line trained separately. Although the networks of different lines do not share the same parameters, states or actions, they all interact with the same environment. The training process of the networks is the same as the A2C algorithm in [21].

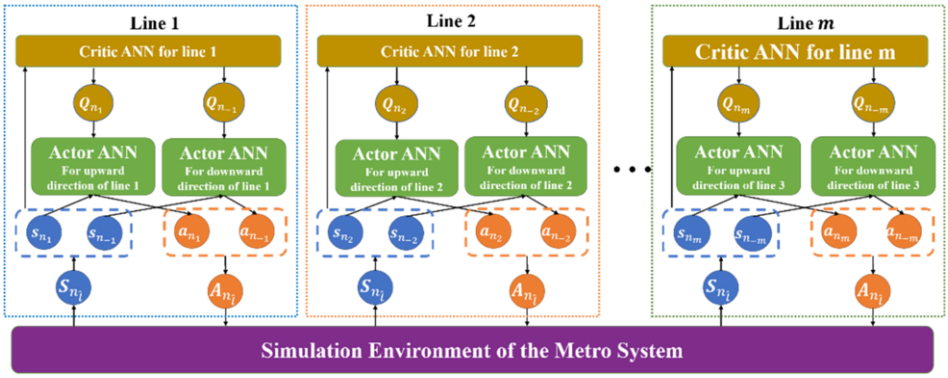


Figure 3. The MAA2C framework for network train scheduling

For passenger flow assignment, we adopt a deep deterministic policy gradient algorithm (DDPG), the training process of which is the same as the DDPG algorithm in [22]. The DDPG network for passenger flow assignment and the MAA2C networks for train scheduling operate independently, and do not share parameters, states or actions. However, they all interact with the same environment, deriving their corresponding states from the same passenger flow distribution and timetables, so as to achieve the collaborative optimization of train scheduling and passenger flow assignment.

4. Numerical Experiments

4.1. Experiment Settings

The reinforcement learning framework is tested with anonymized data from real-world scenarios of Chongqing Metro Line 1, Line 2, and Line 3. The metro network topology is presented in Figure 4 with a total of 84 stations and 20 sections. The sampling frequency of the passenger flow data is 15 minutes, and the operating hours are from 6:00 am to 23:00 pm [17, 18].

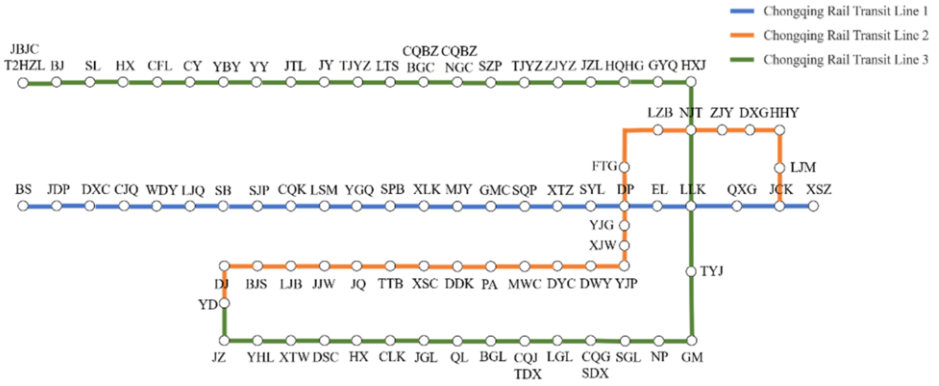


Figure 4. The metro network topology of Chongqing Metro Line 1, 2, 3

In the experiment, we set the maximum passenger capacity C_{max} to 1440. For each line, the maximum number of trains that can be dispatched Q_l is set at 50. The minimum and maximum departure intervals h_{min} and h_{max} are set as 170s and 480s respectively. The minimum and maximum headways u_{min} and u_{max} over all stations are set as 120s and 720s. The minimum train dwelling time g_{min} is set to be 20s, and the maximum g_{max} is set to be 35s. The cruising speed between stations ranges from a minimum of 20m/s to a maximum of 25m/s, from which the minimum and maximum running times $d_{r_i,min}$ and $d_{r_i,max}$ can be derived based on the station distances. The minimum turnaround time t^{turn} is set as 218s.

All experiments in this study are carried out on a Windows machine with 32.0GB of memory, an Intel Xeon E5-2678 v3 CPU with 12 cores and 24 logical processors, and a NVIDIA GeForce RTX3090 GPU with 10496 CUDA cores. All algorithms and programs are coded in Python 3.8 and Tensorflow 2.4.0.

4.2. Results

Figure 5 illustrates the reward curve of the training process with a total of 400 episodes performed. The vertical axis depicts the cumulative reward of all agents per episode, and the horizontal axis represents the number of episodes. Within one episode, each schedule control agent conducts about 120-170 steps. In each step, it generates the arrival and departure time for the next train at all stations. The agent for passenger flow assignment conducts 68 steps in one episode, generating the attraction factors of each section within the next 15 minutes. Despite the potential of a drop in the middle of the curve leading to a sub-optimal solution, the agents can navigate their way out of it and eventually identify a better outcome. After about 150 episodes, the reward begins to stabilize, although minor fluctuations may occur due to random exploration.

Figure 6 presents the curves of passenger demand and the corresponding departure intervals for Line 3 in both upward and downward directions. During the peak hours around 8:00 am and 18:00 pm, passenger demand spikes to its maximum, and the departure interval derived by the agent decreases in response to this, effectively reducing passenger waiting time. In addition, the running times are shortened and the train speed increases during these periods, further reducing passenger waiting times. In off-peak periods, the agent increases the departure interval to minimize traction energy consumption of the metro system.

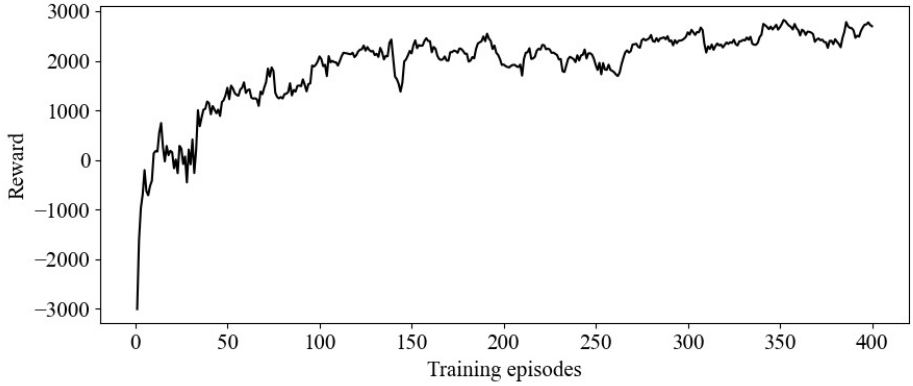
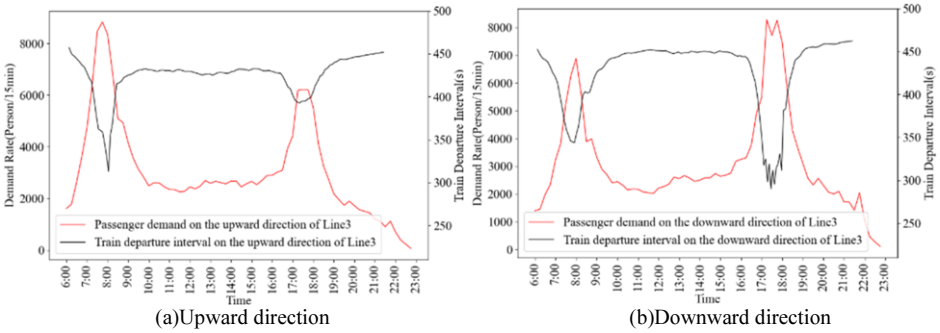


Figure 5. The reward curve of the training process



(a)Upward direction (b)Downward direction

Figure 6. Passenger demand and the corresponding departure intervals derived by the agent on Line 3

To further compare the effects of our optimization, we design five scenarios as follows:

1. Baseline: Trains adopt a fixed timetable, and all passengers follow the shortest path without any guidance. The operating speed of trains follows a curve of maximum acceleration up to 21.75m/s, at which point it enters a coasting state, before finally decelerating at the maximum rate. Train departure interval is fixed at 418s and the dwelling time at each station is set to 30s.
2. PG: The passenger flow assignment agent trained under baseline train schedules is applied. The departure interval is fixed at 418s, the dwelling time at 30s, and the cruising speed between stations is set to 21.75m/s.
3. TS: The schedule control agents trained in joint optimization are applied without passenger flow assignment. All passengers follow the shortest path.
4. TS+PG: The schedule control agents trained in joint optimization and the passenger flow assignment agent trained under baseline train schedules are applied. In other words, train scheduling and passenger flow assignment are optimized separately, then simply superimposed.
5. TSPG: The schedule control agents and the passenger flow assignment agent both trained in joint optimization are applied.

Table 3 shows the traction energy consumption and the average passenger waiting time of the five scenarios. Compared with baseline timetable, those after the trial-and-error learning of the agents show improvements in both of these indicators. With the

implementation of passenger flow assignment, the average passenger waiting time in scenario TS+PG decreased by 1.5% compared to scenario TS, which shows the effectiveness of passenger flow assignment. Both indicators of scenario TS+PG are higher than those of scenario TSPG, indicating that the simple superposition of the two separate optimizations may lead to worse solutions if they are not well-matched with each other. Therefore, it is necessary to optimize train scheduling and passenger flow assignment jointly to achieve better results.

Table 3. Test results of different scenarios

Plans	Full-day traction energy consumption/kWh				Passenger average waiting time/min
	Line 1	Line 2	Line 3	All	
Baseline	$65028^{\downarrow 0\%}$	$67132^{\downarrow 0\%}$	$104551^{\downarrow 0\%}$	$236711^{\downarrow 0\%}$	$5.73^{\downarrow 0\%}$
PG	$65122^{\uparrow 0.15\%}$	$67023^{\downarrow 0.16\%}$	$104678^{\uparrow 0.12\%}$	$236823^{\uparrow 0.05\%}$	$5.68^{\downarrow 0.87\%}$
TS	$57900^{\downarrow 10.96\%}$	$61057^{\downarrow 9.05\%}$	$104850^{\downarrow 0.29\%}$	$221967^{\downarrow 6.23\%}$	$5.21^{\downarrow 9.08\%}$
TS+PG	$58210^{\downarrow 10.48\%}$	$60457^{\downarrow 9.94\%}$	$103993^{\downarrow 0.53\%}$	$222660^{\downarrow 5.94\%}$	$5.13^{\downarrow 10.47\%}$
TSPG	$58729^{\downarrow 9.69\%}$	$59831^{\downarrow 10.88\%}$	$103432^{\downarrow 1.07\%}$	$221992^{\downarrow 6.22\%}$	$5.11^{\downarrow 10.82\%}$

The proposed collaborative optimization approach in this study, while generally requiring longer computing time for training compared with unilateral optimizations, takes into account the interaction between trains and passenger flow during training, ultimately yielding better optimization results. This demonstrates the effectiveness of the proposed multi-agent reinforcement learning framework in joint optimization. Additionally, it shows that collaborative optimization is not merely a simple aggregation of individual optimizations of separate parts. The interaction between trains and passenger flow can influence the optimization effect. There is indeed room for optimization in the metro system as a whole.

4.3. Robustness and Real-Time Analysis

One of the advantages of reinforcement learning is that they can generate optimal policies, rather than solutions. The passenger flow changes every day, and it would be time-consuming to retrain the model each time [23]. If the trained agent is able to respond to the current demand changes without needing to retrain, its robustness can be verified, as well as its practical value in engineering applications. Here, we generate a surge in passenger flow to test the robustness. We assume a tenfold increase in the passenger flow at GM Station on Line 3 from 10:30 am to 12:00 am, while the passenger flow during other time periods and at other stations remains unchanged. Figure 7 shows the curves of the shifted passenger demand and the corresponding departure intervals derived by the agents trained on the original passenger flow. In response to the changes in passenger demand, the departure interval of the shifted passenger flow decreases to a certain extent after 10:30, and gradually returns to the original interval by 12:00. This suggests that the reinforcement learning networks are capable of capturing the complex state transitions and related interactions of train and passenger flows during training, generating reliable state estimation and train schedules in response to demand changes.

Finally, we evaluate the real-time performance of the proposed approach. For generating a set of actions based on the current state, it takes an average of 0.06s for a train scheduling agent and an average of 0.07s for the passenger flow assignment agent. The reinforcement learning agent can adjust the timetable of the next train in real-time, thus is capable of online optimization.

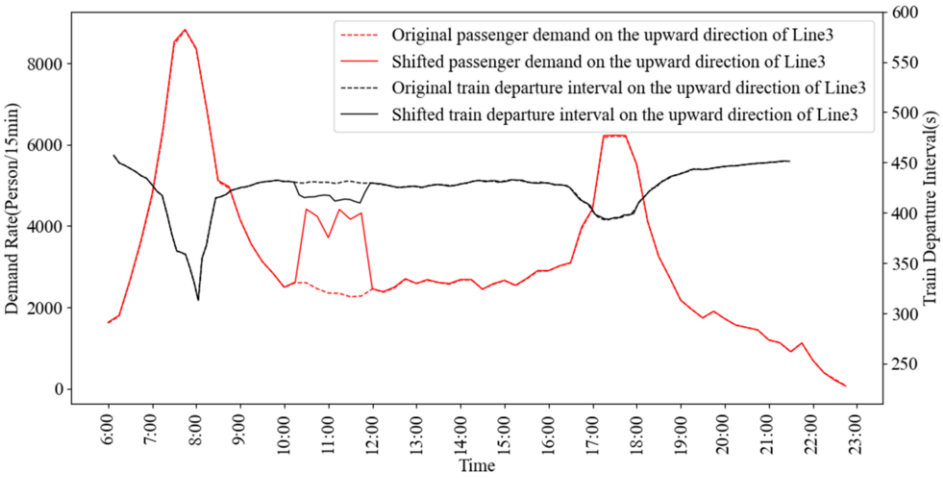


Figure 7. Shifted passenger demand and the corresponding departure intervals derived by the agents

5. Conclusion

This study introduces a multi-agent deep reinforcement learning framework for collaborative online optimization of networked train scheduling and passenger flow assignment. Our approach is able to adjust train schedules and passenger flow assignment strategies in real-time when passenger flow fluctuates, thereby effectively reducing traction energy consumption and the average passenger waiting time. The train and passenger flow simulation is based on a discrete event system and the optimization problem is modeled as a multi-agent Markov decision process, using a multi-agent actor-critic algorithm for train scheduling and a deep deterministic policy gradient algorithm for passenger flow assignment. The simulation environment and the agents are tested under anonymized data from real-world scenarios of Chongqing Metro Lines 1, 2, and 3. The results show that our agents can outperform baseline scenarios, demonstrating its effectiveness, robustness and real-time performance.

For future research, we hope to expand our experiment to the entire network of Chongqing Metro to test the versatility of our approach. Beyond train scheduling and passenger flow assignment, there is potential for further optimization within the urban rail transit system such as train operation and train composition. How to incorporate more factors and combine them for an integrated optimization is a direction worth exploring.

Acknowledgement

This work was supported by the National Key Research and Development Program of China under Grant 2022YFB4300502, the Research and Development Project of Qingdao National Innovation Center of High Speed Train under Grant CX/KJ-2020-0006 and the Research and Development Project of CRSC Research \& Design Institute Group Co., Ltd.

References

[1] Wang Y, Ning B, Tang T, Van Den Boom TJ, De Schutter B. Efficient real-time train scheduling for urban rail transit systems using iterative convex programming. IEEE Transactions on Intelligent Transportation Systems. 2015;16(6):3337-52.

[2] Hou Z, Dong H, Gao S, Nicholson G, Chen L, Roberts C. Energy-saving metro train timetable rescheduling model considering ATO profiles and dynamic passenger flow. IEEE Transactions on Intelligent Transportation Systems. 2019;20(7):2774-85.

[3] Zhang H, Li S, Yang L. Real-time optimal train regulation design for metro lines with energy-saving. Computers \& Industrial Engineering. 2019;127:1282-96.

[4] Liao J, Yang G, Zhang S, Zhang F, Gong C. A deep reinforcement learning approach for the energy-aimed train timetable rescheduling problem under disturbances. IEEE Transactions on Transportation Electrification. 2021;7(4):3096-109.

[5] Ying C-s, Chow AH, Chin K-S. An actor-critic deep reinforcement learning approach for metro train scheduling with rolling stock circulation under stochastic demand. Transportation Research Part B: Methodological. 2020;140:210-35.

[6] Yang G, Zhang F, Gong C, Zhang S. Application of a deep deterministic policy gradient algorithm for energy-aimed timetable rescheduling problem. Energies. 2019;12(18):3461.

[7] Ying C-S, Chow AH, Wang Y-H, Chin K-S. Adaptive metro service schedule and train composition with a proximal policy optimization approach based on deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems. 2021;23(7):6895-906.

[8] Ying C-s, Chow AH, Nguyen HT, Chin K-S. Multi-agent deep reinforcement learning for adaptive coordinated metro service operations with flexible train composition. Transportation Research Part B: Methodological. 2022;161:36-59.

[9] Jia F. Optimization of passenger route dynamic guidance strategy in urban rail transit network. PhD [dissertation]. Beijing: Beijing Jiaotong University; 2021.

[10] Li S, Dessouky MM, Yang L, Gao Z. Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines. Transportation Research Part B: Methodological. 2017;99:113-37.

[11] Shi J, Yang L, Yang J, Gao Z. Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. Transportation Research Part B: Methodological. 2018;110:26-59.

[12] Gong C, Mao B, Wang M, Zhang T. Equity-oriented train timetabling with collaborative passenger flow control: a spatial rebalance of service on an oversaturated urban rail transit line. Journal of Advanced Transportation. 2020;2020:1-17.

[13] Xue H, Jia L, Li J, Guo J. Jointly optimized demand-oriented train timetable and passenger flow control strategy for a congested subway line under a short-turning operation pattern. Physica A: Statistical Mechanics and its Applications. 2022;593:126957.

[14] Shang P. Research on the passenger flow state estimation and operation organization optimization in an urban rail transit system. PhD [dissertation]. Beijing: Tsinghua University; 2019.

[15] Liu R, Li S, Yang L. Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy. Omega. 2020;90:101990.

[16] Shang P. Timetable optimization in urban railway transit network under dynamic passenger demand [master's thesis]. Beijing: Tsinghua University; 2016.

[17] Zhang M, Dong W, Sun X, Ji Y, editors. A method for enhancing global safety of regional rail transit based on coordinative optimization of passenger flow assignment and train scheduling. Journal of Physics: Conference Series; 2020: IOP Publishing.

[18] Zhao D. Key technology research on collaborative optimization and simulation platform of new urban rail transit system[master's thesis]. Beijing: Tsinghua University; 2022.

[19] Howlett PG, Milroy I, Pudney P. Energy-efficient train control. Control Engineering Practice. 1994;2(2):193-200.

[20] Gupta JK, Egorov M, Kochenderfer M, editors. Cooperative multi-agent control using deep reinforcement learning. Autonomous Agents and Multiagent Systems: AAMAS 2017 Workshops, Best Papers, São Paulo, Brazil, May 8-12, 2017, Revised Selected Papers 16; 2017: Springer.

[21] Sutton RS, Barto AG. Reinforcement learning: An introduction: MIT press; 2018.

[22] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. arXiv preprint arXiv:150902971. 2015.

[23] Wu T, Dong W, Ye H, editors. A Deep Reinforcement Learning Approach for Optimal Scheduling of Heavy-haul Railway. The 22nd World Congress of the International Federation of Automatic Control (IFAC World Congress); 2023; Japan.